SAE 1997 TRANSACTIONS

JOURNAL OF PASSENGER CARS

SECTION 6 - PART 1

SAE The Engineering Society For Advancing Mobility Land Sea Air and Space
INTERNATIONAL

SAE 1997 TRANSACTIONS

JOURNAL OF PASSENGER CARS

Section 6 - Volume 106
Part 1

Ronald K. Leonard/President
Neil A. Schilke/Treasurer
Karl Goering/Assistant Treasurer
Max E. Rumbaugh, Jr./Executive Vice President & Secretary

PUBLISHED BY: SOCIETY OF AUTOMOTIVE ENGINEERS, INC.
400 Commonwealth Dr., Warrendale, PA 15096-0001
Phone (724)776-4841 Fax (724)776-5760



GLOBAL MOBILITY DATABASE
All SAE papers, standards, and selected books are abstracted and indexed in the Global Mobility Database

The appearance of the ISSN code at the bottom of this page indicates SAE's consent that copies of the articles may be made for personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay a $1.00 per page copy fee through the Copyright Clearance Center, Inc., Operations Center, 222 Rosewood Drive, Danvers, MA 01923 for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

SAE routinely stocks printed papers for a period of three years following date of publication. Direct your orders to SAE Customer Service Department.

To obtain quantity reprint rates, permission to reprint a technical paper or permission to use copyrighted SAE publications in other works, contact the SAE Publications Group.

ISBN 0-7680-0301-6 (Section 6)
ISBN 0-7680-0302-4 (Set)
ISSN 0096-736X
Copyright 1998 Society of Automotive Engineers, Inc.
Printed in U.S.A.

Vol. 106 1997

SAE TRANSACTIONS

JOURNAL OF PASSENGER CARS

A publication of the
Society of Automotive Engineers, Inc.

This Transactions volume contains the best 383 technical papers of all those presented in 1997. Each has been reviewed according to the established review procedures of the Engineering Meetings Board of SAE. These are the best quality SAE Passenger Cars papers presented in 1997 and have been judged by technical experts to be worthy of preserving in the permanent technical literature.

The reviewing procedures require a technical review of all SAE papers by at least four well-qualified individuals who are technically competent within the scope of the papers' subject/discipline and who are experienced in reviewing papers.

Each paper is evaluated in four areas – long term reference value; presentation of technically new and innovative materials or constructive and in-depth review; professional integrity; and clarity of presentation. Based on the results of this review process, papers are selected for publication.

Engineering Meetings Board
James M. Novak, Chairperson
David Eames, Robert L. Ireland, Douglas C. Johnson, Patrick T. Keenan,
William R. Leppard, Lawrence Looby, Josef Mack, Stanley E. Rayford, Takanori Sonoda,
V. Sumantran, K. Clark White, Norman R. White, Ronald E. York

Land and Sea Group
V. Sumantran, Chairperson
James E. Smith, Vice Chairperson
John Connor, Ronald L. Cutler, Frank S. Gerry, Richard T. Gutowski, Iftekhar A. Ibrahim, Gary Lambert, William A. Leasure Jr., Josef Mack, Philip J. Mazziotti, Robert F. Parker, Prakash T. Sathe, Kenneth J. Stoss, John P. Sunne

Air & Space Group
Douglas C. Johnson, Chairperson
Michael E. Nelson, Vice Chairperson
Bradley N. Andelian, Jerry E. Beam, Ambrose A. Hauser, Robert F. Jaekel, Carl E. Meece, Benjamin J. Neumann, Kamran Rokhsaz, W. Glen Smith, Benton E. Visser,
Calvin F. Watson, Jr.

Technical Committee of the Engineering Meetings Board
Donald C. Siegla, Chairperson
Tanvir Ahmad, Duane L. Abata, Alex C. Alkidas, Charles A. Amann, Carl L. Anderson, George C. Davis, James R. Eagan, John H. Johnson, Radhey L. Kushwaha, William A. Leasure, Jr., Phillip S. Myers, David L. Newcom, Donald J. Patterson, Gino Sovran, Simon C. Tung

Transactions Journals
John H. Johnson, Editor-in-Chief
James E. Smith, Technical Editor, Journal of Engines
Ronald G. Hurley, Technical Editor, Journal of Fuels & Lubricants
Gary R. Lambert, Technical Editor, Journal of Materials & Manufacturing
Philip J. Mazzotti, Technical Editor, Journal of Commercial Vehicles
V. Sumantran, Technical Editor, Journal of Passenger Cars
Kirk L. Yerkes, Technical Editor, Journal of Aerospace

Publications Committee of the Board of Directors
Duane Miller, Chairperson
Thomas E. Austin, Karl Goering, Dave Lancaster, Jim L. Rau, Roy Radakovich

ACTIVITY AND READERS COMMITTEE CHAIRPERSONS

AIR & SPACE GROUP

AEROSPACE VEHICLE & PROPULSION ACTIVITY – ROBERT JAEKEL
Commercial Transport Aircraft – Herbert L. Mailander
General Aviation Aircraft – Weldon Garrelts
V/STOL & Military – David H. Axelson
Aircraft Propulsion – Nick Itsines

AEROSPACE MANUFACTURING MAINTENANCE ACTIVITY – BRAD ANDELIAN
Aerospace Automated Fastening – Mahboob Alam
Manufacturing Technology –
Materials/Processes –
Aviation Maintenance –

AEROSPACE SYSTEM ACTIVITY – GLEN SMITH
Aerospace Power Systems – Jack E. Garner
Control & Guidance Systems – Dick Quinlivan
Environmental Systems – H. Jack Fivel
Aerospace Medical– Donald Reid

AEROSPACE TECHNOLOGIES ACTIVITY – KARMAN ROKHSAZ
Human Behavioral Technology –
Aerodynamics –
Structures –
Simulation Technology –
Information Sciences –

LAND & SEA GROUP

AUTOMOBILE BODY ACTIVITY – V. SUMANTRAN
Passenger Protection – David Viano
Body Research & Engineering – Robert Dubensky
Human Factors –
Vehicle Configuration – V. Sumantran

FARM, CONSTRUCTION & INDUSTRIAL MACHINERY – KEN STOSS
Construction – Kenneth Wolfgram
Farm Tractors & Equipment – John Leen
FCIM Transmissions & Drivelines – Jon Peppel
Advanced Concepts – George H. Morgan
Fluid Power – David Newcom

FUELS & LUBRICANTS ACTIVITY – JOHN SUNNE
Lubricants – Beth Schwab
Fuels – Tom Austin
Combustions – Brian Sangeorzan
Emissions – J. Robert Mondt

MANUFACTURING ENGINEERING ACTIVITY – PRAKASH SATHE
Manufacturing Engineering – Adnan Aswad

MATERIALS ENGINEERING ACTIVITY – GARY LAMBERT
(Ferrous Materials)
Coated Sheet and Other Flat Rolled Steels – Raju Venkataraman
Other Ferrous Products - Richard B. Bertolo
(Non-Ferrous Materials)
Wrought, Cast & Forged Non-Ferrous Products – Martin Myers
Polymers & Composites – Jack Gardner
Waste Materials Treatment
Powder Metals – Jean Lynn

PASSENGER CAR ACTIVITY – JOE MACK
Noise & Vibration – Paul Donovan
Brakes - David Sheridan
Steering, Chassis & Suspension – Frederick Winsor
Tires – John Andrus
Trans/Drivelines & Axles – Martin G. Gabriel
Engines – Donald C. Siegla
Safety – Howard S. Willson
Testing & Research
Electrical & Electronics Systems – Mark Dobies
Electric Car – Bradford Bates
Climate Control – James Baker
Computer

MAINTENANCE TECHNOLOGY ACTIVITY – PHILIP J. MAZZIOTTI
Automobile & Light Truck Maintenance – John M. Allderige
Diagnostic & Service Tool – John M. Allderige
Rebuild, Remanufacture & Recycling – Philip J. Mazziotti

POWERPLANT ACTIVITY – JAMES A. SMITH
Diesel Engine – Otto Uyehara
Small Powerplants – George G. Lassanske
Advanced Powerplants – Frank Stodolsky

COMMERCIAL VEHICLE ENGINEERING & OPERATING ACTIVITY – WILLIAM A. LEASURE, JR.
Commercial Vehicle Operations & Maintenance – Nolan MacLaren
Total Vehicle –
Chassis & Suspension Systems –
Braking & Steering Systems – Karl Schipper
Light Truck & Van – Katherine Cockey
Powertrain & Drivetrain Systems – Russell Holmes
Electrical & Electronic Systems – Ramin Younessi
Bus Engineering - Peter G. York
Military Committee – Roger R. Gay
Intelligent Transportation Systems – Douglas McKelvey

VEHICLUAR HEAT EXCHANGE & HEAT TRANSFER SYSTEMS ACTIVITY – RICHARD GUTOWSKI
Vehicular Heat Exchange & Heat Transfer Systems - Alex Alkidas

TABLE OF CONTENTS

PART 1

Part 2

970091

Objective Ride and Handling Goals for the 1997 Chevrolet Corvette

Joseph P. Ryan, Steven P. Fuja, and Henry A. Schmid
General Motors Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The process of gathering information, analysis, and selection of ride and handling goals for the 1997 Chevrolet Corvette is described. The goals consist of measurable objective metrics and standardized subjective evaluation. The input elements of the process are: the Voice of the Customer, Engineering Direction and Competitive Assessment. As values are developed for the each metric, synthesis and analysis methods are used to confirm the full set of metrics are consistent and non-exclusive. The targets selected guided the chassis design and development of the 1997 Chevrolet Corvette to insure outstanding customer satisfaction in handling and ride.

INTRODUCTION

OVERVIEW - The fifth generation Chevrolet Corvette introduced in 1997 (C5) is a completely new vehicle, featuring significant design changes in all areas compared to the fourth generation Chevrolet Corvette of 1984-96 (C4).

Four papers have been written that summarize the Systems Engineering approach to the suspension design of the 1997 Chevrolet Corvette. This paper reviews the customer research, Engineering Direction, human factors and analytical evaluation that established the objective Ride and Handling targets for the vehicle. The second paper [1] describes the "flow down" of the vehicle level performance targets to specific subsystem attributes and parameters for the suspension, steering, and tires. The third paper [2] reviews the execution of the front and rear suspension in production hardware that complies with the specifications established in the previous two steps. The fourth paper [3] summarizes the final development tuning of the suspension, steering and tires. Careful execution of the procedures and techniques discussed in this series of papers was critical in producing the highly optimized suspension of the 1997 Chevrolet Corvette and assurance of high customer satisfaction for ride and handling performance.

PAPER ORGANIZATION - The material presented is organized into six major topics: Vehicle Level Metrics, Voice of the Customer, Engineering Direction, Mule Vehicles, Competitive Assessment and Targets and Synthesis/Analysis. An Appendix contains charts of supporting Competitive Assessment data for each metric discussed.

VEHICLE LEVEL METRICS

The objective metrics established at the beginning of a new vehicle program provide crucial guidance to the design process. Without numerical values assigned to measurable (and simulation compatible) objective performance metrics, efficient synthesis of subsequent design specifications would be impossible. Metrics were chosen in response to the Voice Of the Customer or Engineering Direction, and were formalized in the Vehicle Technical Specification (VTS), the document that states vehicle performance requirements and guides all subsequent design and development steps.

By rigorously following the Systems Engineering process established, there is very high confidence the first "out of the box" vehicles will exhibit overall performance close to the final desired package, and require a minimum of tuning or modification to the fundamental suspension. The goal of the Ride and Handling engineer is to meet all of the performance targets, but above all, produce a balanced, pleasing vehicle.

As the vehicle architecture matured, many suspension configurations were investigated for compatibility with the targets. All suspension designs pursued beyond a preliminary assessment were required to demonstrate the potential to meet all the Ride and Handling metrics.

During design and development, frequent assessments were made of vehicle progress by running laboratory and road tests. Laboratory tests verified suspension and tire performance parameters. Full

vehicle road tests were conducted to verify compliance with the Ride and Handling metrics.

HANDLING

There are hundreds of metrics that may be derived from instrumented vehicle tests. The key to efficient design synthesis is to reduce the list to the minimum set necessary to establish the fundamental handling character of the vehicle.

The Ride and Handling metrics selected for guiding C5 suspension, steering and tire design are shown in Table I. Four categories are listed: Linear Range handling (directional control below 0.3 g lateral acceleration), Non-linear Range handling (vehicle ultimate cornering and limit stability), On-Center handling (straight-ahead steering performance below 0.15 g) and basic Ride. General Motors test methods and metrics for handling have been previously described for Directional Control by Kunkel and Leffert [4], and On-Center handling by Norman [5].

Table I. C5 Ride and Handling Metrics.

LINEAR HANDLING	Understeer Gradient	deg/g
	Steering Sensitivity	g/100 deg
	(Steering Sensitivity) Linear Range	g
	Lateral Acceleration Response Time	s
	Roll Gradient	deg/g
	Roll Damping	--
	Yaw Velocity Damping	--
	Pitch Gradient - Acceleration & Braking	deg/g
NON-LINEAR HANDLING	Maximum Lateral Acceleration	g
	Dropped Throttle Stability	g
	Single Bump Roadholding	deg
ON-CENTER	Minimum Steering Sensitivity	g/100 deg
	Steering Sensitivity Ratio	--
	Steering Hysteresis	deg
	Lateral Acceleration at 0 Torque	g
	Steering Torque Gradient Ratio	--
	Steering Work Sensitivity	g^2/100 Nm
	Static Steering Effort	N
	Curb-to-Curb Turning Circle	m
RIDE	Front Ride Frequency	Hz
	Ride Frequency Ratio (Rear/Front)	--
	Impact Harshness	*
	Vertical Damping	*

* Subjective rating plus experimental metrics.

A "cross-over" test that provides both handing and ride information is a single-bump roadholding test, which has been previously described by various sources. Typically, the testing procedure establishes the vehicle in a constant radius turn and then encounters a single, fixed profile bump. The derived metric is usually based on yaw velocity deviation caused by the bump. The scope of the vehicle dynamics being exercised is limited, but simplicity and repeatability make this test very useful. Chassis contributors to roadholding may be evaluated, in addition to control of the unsprung mass, which is also fundamental ride consideration.

Measuring continuous roadholding in objective terms is more difficult. Test repeatability is hampered by the ability to provide a consistent road input to the vehicle (path, profile, coefficient of friction, etc.). Experimental tests and metrics are currently under development, but subjective evaluation is still the primary means for assessing continuous roadholding.

RIDE

Objective definition of ride is not as mature as handling. For C5, objective metrics from road test and simulation combined with subjective evaluation of vehicles formed the basis for defining ride. New test and analysis techniques that produce objective metrics are continuously being developed, evaluated, and adopted when appropriate.

Metrics from road tests generally fall into two areas: vehicle response to single events and continuous response over sections of standard Proving Ground roads. Single event tests are easily reduced to mechanical terms such as maximum displacements, settling times, damping levels, etc. Extended road response tests are typically reduced to a single cumulative parameter such as Power Spectral Density or an integrated area under a curve.

Subjective evaluations are used for vehicle attributes that do not have a reliable objective metric assigned, and are most beneficial in tabulating relative rankings of vehicles or tuning configurations by the same driver under similar conditions. The inherent limitations (such as driver-to-driver variability) and inability to simulate subjective ratings illustrate the difficulties of relying on subjective methods for efficient design synthesis.

Control - Control describes rigid body motions of the sprung mass (1-2 Hz) and the unsprung masses (10-15 Hz). Guidelines for rigid body ride have been in existence since the 1930's based on pioneering work by Olley [6]. Concepts such as ride frequency, oscillation centers, sprung mass damping and wheel control were explored. Objective description of Control includes natural frequencies, damping, sprung and unsprung mass motions, etc. Vehicle parameters that benefit rough road handling also tend to improve Control attributes of ride (i.e. head toss) and useful ride information may be gathered from the roadholding test described above.

Isolation - Higher frequency responses (above 15 Hz) that impact the driver as vibration or noise are also more difficult to reduce to simple metrics. Typical Isolation issues are: impacts, rough road shake and

coarse road isolation. The beneficiary of improved Isolation are the passengers, and a meaningful Isolation metric must comprehend the human element. Human Factors studies are continually being developed to quantify the perception of isolation, which can vary significantly from one individual to another. Development of Isolation metrics are maturing (as a result of increasingly complex computer simulations, new test procedures, sophisticated data analysis techniques and ongoing Human Factors studies) but are not in wide use.

VOICE OF THE CUSTOMER

The C5 Engineering Team and Chevrolet researched customer desires and expectations through clinics held at various locations across the country. In focus groups, C4 owners ranked "well built" and "performance" as their two most critical desires for the new Corvette. Well-built was ranked slightly ahead of performance, which in turn was further defined as primarily handling. Owners of other high-end sports cars also expressed a strong penchant for handling performance, but not at the sacrifice of ride quality.

In general, the positive comments of owners relating to C4 chassis performance included: high handling limits, predictable control at handling limits, and the suspension's ability to take the punishment of hard driving (including race track and autocross sessions). Dislikes included: harsh ride and poor straight-ahead tracking.

Customer comments used for C5 Ride and Handling metric development are shown in Table II as the "input" side of a sample QFD (Quality Function Deployment) table. The "output" side lists the C5 ride and handling metrics. In practice, the QFD chart is further detailed by identifying known relationships between input and output items, and rating each as strong, medium or weak.

Another source of customer input are J. D. Powers surveys. Data from C4 owners indicated extremely high levels of satisfaction for handling, and somewhat lower opinions for overall ride. These overall impressions of the C4 contributed to the early direction set for the C5 to provide best-in-class ride but not at the expense of handling.

ENGINEERING DIRECTION

OBJECTIVE METRICS - The Chief Engineer and the Ride and Handling Development Engineers defined the targets for specific metrics considered important to achieve the desired C5 performance and pleasability envelope.

TABLE II. QFD - Voice of the Customer to Objective Metrics.

METRIC / VOICE OF THE CUSTOMER	Steering Wheel Torque at 0 g	Curb to Curb Turning Circle	Steering Sensitivity Ratio	Static Steering Effort	Off Center Steering Torque Gradient Ra	Strg wheel Torque at 0 g	Lateral Acceleration Response Time	Understeer Gradient	Steering Sensitivity	Roll Gradient	Maximum Lateral Acceleration	Yaw Velocity Damping	Pitch Damping	Pitch Gradient - Acceleration	Pitch Gradient - Braking	Single Bump Disturbance	Impact Isolation	Vertical Damping	Front Ride Frequency	Ride Frequency Ratio
HANDLING																				
Performance, braking and handling allow driver to safely control car																				
Car feels heavy and sturdy																				
Car's handling allows for fast cornering																				
Car has good handling and stability at highway speeds																				
Car handles well over poor roads																				
Car handles well during all weather conditions																				
Car's handling provides a total feel of the road																				
Car maneuvers in and out of traffic with ease																				
Car responds well to steering wheel movement																				
Car is easy to park																				
Car comes to a smooth stop																				
Car has smooth ride over bumpy roads																				
RIDE																				
Car has smooth ride over bumpy roads																				
Car does not shake or vibrate over bumpy roads																				
Road and wind noises are absent from the car																				
Vehicle noise is absent form the car																				
Car has no squeaks and rattles																				

Values for "tuning" type metrics such as Steering Sensitivity and Roll Gradient were recommended. Maximum Lateral Acceleration was given special emphasis in part due to wide public acceptance of this metric as a key benchmark

Chassis Engineering was the liaison to other corporation resources to provide necessary support for the evolution of C5 Ride and Handling. The cumulative corporate knowledge base for many areas (customer usage, durability, manufacturability, liability, Human Factors, performance, technology, etc.) were integrated and applied to the C5 design.

Throughout design and development, the Ride and Handling Engineering Team periodically reviewed progress against performance targets in the VTS. When necessary, design or tuning changes were implemented to insure compliance of the C5 to the specified performance targets.

VEHICLE ARCHITECTURE

Several different vehicle concepts and layouts were considered for the C5. These included: mid-engine, integral powertrain and chassis structure, all-wheel-drive, and others. As the vehicle architecture matured, several overall vehicle features adopted for the C5 were strongly supported by the Ride and Handling Engineering Team. New features of the C5 that benefit Ride and Handling include: longer wheelbase, reduced overhang, wider tracks, rear transaxle, mid-car tandem fuel tanks, highly optimized body structure, structural front and rear crossmembers, and Extended Mobility Tires (EMT) tires.

Wheelbase - The 208 mm longer wheelbase helps in many areas of vehicle dynamics, including quicker response times for lateral acceleration, yaw velocity and sideslip; reduced sideslip; reduced phase lag between yaw velocity and lateral acceleration; reduced pitch under braking and acceleration; and flat ride tuning.

Track - The front track is 103 mm wider, and the rear track is 62 mm wider. Wider tracks improve roll stiffness efficiency and promote longer suspension links for more robust geometry.

Weight Distribution - The heritage of the Corvette strongly influenced the decision to keep the engine in the front, yet careful packaging produced a 50/50 weight distribution. The 50/50 weight distribution enables reduced front cornering compliance and understeer and enhances Maximum Lateral Acceleration by taking weight off the front tires.

Reduced Moments of Inertia - Moving the transmission to the rear of the vehicle reduced weight on the front axle and was a major contributor to the 50/50 weight distribution, which located the center of gravity at the midpoint of the wheelbase. (Another benefit of moving the transmission was to improve passenger compartment room, driver ergonomics, and place the driver further ahead of the rear axle, contributing to improved ride quality.) The C5 differential design locates the pinion gear on the rear side of the ring gear, which reduced the overall length of the transaxle assembly and allows less rear body overhang. The spare tire, which would have been located behind the rear axle, was eliminated by adopting EMT tires. The fuel tank was moved (from behind the rear axle on the C4) to tandem mid-car fuel tanks, located each side of the center tunnel, ahead of the rear axle.

As a result of the overall packaging of the transaxle, fuel tanks, and elimination of the spare tire, the yaw and pitch moments of inertia were reduced approximately 7%. Reduced yaw inertia provides quicker response times for lateral acceleration, yaw velocity and sideslip; and eliminates sideslip reversal during steer initiation from straight ahead, improving the driver's perception of a crisp turn-in.

Tires - C5 tire sizes are P245/45ZR17 front and P275/40ZR18 rear (Standard tires on the 1996 Corvette were P255/40ZR17 and P285/40ZR17, respectively). The narrower tread of the C5 tires help reduce vehicle sensitivity to lateral disturbances from uneven or cambered roads (i.e. truck grooves or troughs).

The required rear tire cornering stiffness is over two times higher than the front to achieve the target front and rear cornering compliances, which are absolutely critical for establishing the fundamental handling character of the vehicle. To achieve the required cornering stiffnesses with equal weight on each tire, different sizes are mandatory. There must be no possibility of intermixing front and rear tires, and optimization of cornering, ride, traction, and rolling resistance are more readily achieved in unique size tires for the front and rear.

C5 tires were required to provide high mileage and good value to the customer in addition to excellent vehicle performance. The Ride and Handling Engineering Team specifically avoided the use of soft tread compound (low mileage) tires, which is an easy way to increase cornering stiffness and compensate for a poor chassis design.

Structure - The C5 body structure features hydroformed side rails and large structural tunnel. The new body structure is 450% stiffer torsionally and has greatly improved local stiffnesses [7]. Suspension interfaces benefit from increased vertical and lateral stiffness of the body attachment points and hard mounted structural crossmembers.

Springs, shock absorbers and bushings are more efficient and tunable when the series compliances of the attaching structure are minimized.

MULE VEHICLES

Four mule vehicles were used in early development work. The first was built in 1990, a C4 with a Short-Long Arm (SLA) rear suspension. The second was also a modified C4 with SLA suspensions in front and rear that featured the longer C5 wheelbase and bell-crank actuated, horizontal coil-over-shock springing. When the overall C5 vehicle architecture was established, two "pre-Alpha" vehicles were built from the ground up, incorporating all of the mainstream design features.

Evaluation of specific components and suspension configurations provided valuable feedback for setting the ride and handling targets. Areas investigated were: ride frequencies, roll rates, suspension anti's, camber geometry, roll steer, spring configurations, fore-aft compliance, tires, fundamental damping, braking and acceleration steer and steering system tuning. Invaluable feedback was collected for fundamental ride, roll, isolation, pitch under braking and acceleration, steering isolation, cornering stability under braking and acceleration and roadholding.

COMPETITIVE ASSESSMENT AND TARGETS

COMPARISON VEHICLES - From an extensive database of instrumented vehicle tests maintained by the NAO Vehicle Dynamics Laboratory, a group of 45 peer vehicles and 22 production and development C4 Corvettes were selected. Appendix I contains scatter plots for selected Ride and Handling metrics. Measured values for the comparison vehicles plus the C5 target value or range are shown. Targets are identified either as a range, minimum or maximum.

LINEAR RANGE HANDLING

Understeer Gradient - The target range was determined by the Handling Synthesis method described below, to achieve the targeted Lateral Acceleration Response Time and Yaw Velocity Damping. The target range falls near the median value for comparison vehicles.

Steering Sensitivity - The target range was recommended by Engineering Direction based on many years of experience with the C4, to provide an appropriate balance between responsiveness and ease of highway driving. Human Factors studies [10,11] were also consulted to optimize driver-vehicle performance in path following, which eases the stress of the everyday driver and helps the precision of the autocrosser and racer.

Steering Sensitivity Linear Range - The target value was selected to place the C5 at best-in-class. A high linear range enhances the driver's perception of linear, predictable handling as lateral acceleration increases on a highway on-ramp or a decreasing radius turn.

Lateral Acceleration Response Time (LART) - The target value was selected to place the C5 at best-in-class. The driver perceives LART as a primary cue of vehicle responsiveness. Human Factors studies [8,9,10,11] recommend a short LART is important for driver controllability, path following, and obstacle avoidance.

Roll Gradient - The target range was recommended by Engineering Direction to achieve a desired level of overall vehicle motion on the road. This range is "softer" than a typical C4, in an effort to improve ride quality and rough road handling.

Allowing more roll enables smaller stabilizer bars, and closes the gap between one wheel and two wheel ride rate, enhancing the ability of a passive damper to control wheel motions over a wider range of roads. (Two wheel ride rate refers to parallel jounce where only the ride spring(s) resists vertical motion. One wheel ride rate refers to one side only jounce where the ride spring plus the stabilizer bar resist vertical motion.)

Roll Damping - The target value was chosen as a complement to the roll gradient. In general, more roll may be allowed if it is well damped, to minimize head-toss and avoid abrupt transients in tire lateral load transfer during cornering.

Yaw Velocity Damping (YVD) - The target value was chosen to place the C5 at best-in-class. Human Factors studies [8,9,10,11] recommend increased YVD is important for stability in severe maneuvers, path following and obstacle avoidance.

Pitch Gradient, Acceleration and Braking - The targets were also chosen to soften the body motions of the C5. Most chassis parameters that allow more body pitch under braking and acceleration also are enablers for better ride. Lower ride rates may be used in front and rear to enhance ride. Reduced anti-dive and less side-view swing arm maybe used in the front suspension, increasing rearward wheel motion in jounce, and improved impact harshness. The longer wheelbase of the C5 helps offset the effect of reduced ride rate and front anti-dive.

NONLINEAR RANGE HANDLING

Maximum Lateral Acceleration - The target was selected by Engineering Direction to place the C5 among the best-in-class. A high cornering limit not only maximizes autocross and racetrack performance, but

also provides a larger maneuvering envelope to the everyday driver in emergency situations.

Dropped Throttle Stability - The target was recommended by Engineering Direction based on C4 experience, and also maximizes the overall stable envelope of the vehicle to aid the driver in emergency situations.

Single Bump Roadholding - Targets were chosen for path deviation caused by a single bump in three modes: the contribution of the front axle, rear axle, and net (front - rear) to assess rough roadholding, ride, and uneven road disturbance rejection.

ON-CENTER HANDLING

Minimum Steering Sensitivity - "Crispness". The target was chosen to place the C5 at best-in-class for front engine vehicles, and was verified by Steering System Synthesis methods as realistic. (Comparison vehicles measuring higher are rear-engine and lower front mass or have unusually low steering gear ratios. The mass effect was also observed by Norman in [5]).

Steering Sensitivity Ratio - "Gain Linearity". The target was chosen to place the C5 at best-in-class for front engine power steering vehicles. (Comparison vehicles measuring higher are rear-engine and manual steering). A vehicle with a low value would be subjectively described as having a steering "dead zone" on-center.

Steering Hysteresis - "Friction". The target was chosen as an appropriate goal verified by Steering System Synthesis methods. High levels of friction adversely affect returnability and straight-tracking ability. Low levels of friction may decrease yaw velocity damping in a free-control situation.

Lateral Acceleration at 0 Steering Wheel Torque "Returnability". The target range was chosen as an appropriate goal verified by Steering System Synthesis methods. This metric is an indicator of the vehicle's ability to return to center, which similarly can be too great or too little.

Steering Torque Gradient Ratio - "Effort Linearity". The nominal target was chosen to place the C5 at best-in-class for power steering vehicles. (Comparison vehicles measuring higher are manual steering).

Steering Work Sensitivity - "Gain and Effort Balance". The target range was chosen as being representative of the appropriate balance of steering gain and effort for a vehicle in this class. Vehicles with lower values tend to be "heavy and sluggish", and vehicles with higher values tend to be "light and darty".

Static Steering Effort - The target was chosen to provide a comfortable parking effort for all drivers. The standard Magnasteer™ II variable effort steering system enables low parking effort without over-boosting the steering at higher speeds.

RIDE

Front Ride Frequency - The target range of 1.0-1.2 Hz for the undamped Ride Frequency was recommended by Engineering Direction from C4 experience.

Ride Frequency Ratio - The target range of 1.13-1.2 Hz for the undamped Rear/Front Ride Frequency ratio was recommended by Engineering Direction from C4 experience, plus Synthesis/Analysis studies anticipating the effect of the longer C5 wheelbase.

Impact Harshness and Vertical Damping - Target subjective ratings were chosen based on Engineering Direction. Experimental objective metrics were also considered as they were developed and verified.

BALANCING PERFORMANCE

It is certainly possible to build a car that exceeds all others in terms of a specific metric. The challenge of good chassis design is to strike a balance between all areas of performance to produce a car that will exceed the expectations of both the novice and expert driver. The C5 was targeted to be best-in-class in all areas, with justifiable expectations that all metrics were achievable.

SYNTHESIS AND ANALYSIS

All metrics were scrutinized by various analytical methods to verify each was achievable. Also, the full set of metrics were continuously monitored to insure preliminary chassis design parameters did not exclude any metric at the expense of meeting others.

HANDLING

Analysis - There are a wide variety of handling simulations (developed at the GM NAO Vehicle Dynamics Laboratory) that can reproduce all the standard road tests. "Raw" data is generated, then fed into the same data reduction software used for road test data. Simulated vehicles may then be evaluated in the same objective metric terms as actual vehicles.

Synthesis - An extension of the group of handling simulations is the family of Chassis Synthesis tools. These tools facilitate the synthesis of intermediate chassis parameters given specific vehicle performance metrics to be met.

FUNDAMENTAL HANDLING SYNTHESIS

The first level of chassis parameter synthesis is to identify the fundamental vehicle characteristics that yield the steady-state and dynamic handling behavior described by the target metrics.

Understeer Gradient is a mathematical description of the fundamental stability of the vehicle during cornering. In simple terms, the sign and magnitude of the numerical value describes the slip angle of the center of gravity of the vehicle during cornering, which may be subjectively described as the attitude of the vehicle in a corner. Also, for an understeer vehicle on a fixed radius path, as speed increases, additional steering input is required to maintain path. (A neutral steer vehicle requires no change in steering input, and an oversteer vehicle requires decreasing steering input.)

Normalizing the vehicle (center of gravity) slip angle to lateral acceleration defines the Understeer Gradient in deg/g. The magnitude varies as lateral acceleration increases, due to chassis tuning and the nonlinear characteristics of the tires. The amount of variation in Understeer Gradient as lateral acceleration increases affects the subjective impression of handling linearity. The C5 was targeted to have no significant change in understeer gradient, until a very high lateral acceleration, where tire saturation becomes dominant, and terminal understeer results.

To further define vehicle behavior, the individual axle contributions to Understeer Gradient may be expressed as individual Front and Rear Cornering Compliances, as described by Bundorf and Leffert [12]. Similar to Understeer Gradient, the Cornering Compliance is the slip angle of the vehicle at each axle. Mathematically, Understeer Gradient is Front minus Rear Cornering Compliance. The Cornering Compliance is also a lumped parameter representation of each axle's tire, mass and suspension effects that contribute to understeer.

There are an infinite number of Cornering Compliance combinations that yield the same linear range Understeer Gradient (steady state behavior). The individual magnitudes of the Front and Rear Cornering Compliances have profound effect on the dynamic behavior of the vehicle. For C5, the two primary indicators used for dynamic performance were Lateral Acceleration Response Time (LART) and Yaw Velocity Damping (YVD).

In Figure 1, LART versus YVD is shown for the group of comparison vehicles, plus the C5 target point. When these two modes of dynamic behavior are reviewed together, the target performance for the C5 is clearly breaking new ground for responsiveness and stability, where no other vehicles have demonstrated capability.

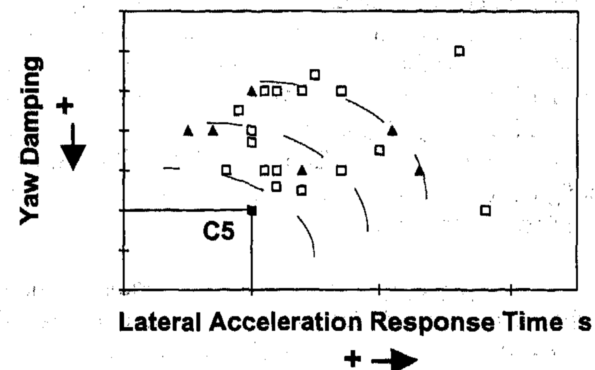


FIGURE 1. Lateral Acceleration Response Time and Simultaneous Yaw Velocity Damping, C5 and Comparison Vehicles.

A useful synthesis technique determines the Understeer Gradient and Front and Rear Cornering Compliances required to simultaneously achieve the target LART and YVD. The graphical representation of this concept is known as a Handling Map and was first described by Mackovjac and Lang [13]. A Handling Map used for the C5 is shown in Figure 2.

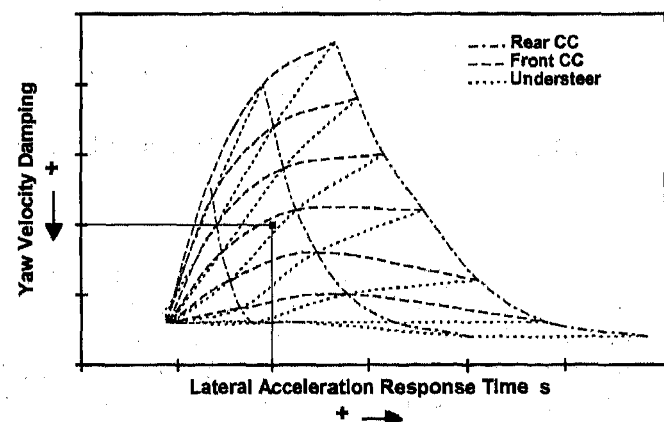


FIGURE 2. C5 Handling Map.

In this example, the target LART and YVD values were located on the horizontal and vertical axes to locate the simultaneous point on the chart. The location of this point on the parametric surface identifies the required Front and Rear Cornering Compliances and Understeer Gradient. All subsequent chassis design was strictly controlled to assure the target front and rear Cornering Compliances are maintained. With these values established, analysis can proceed more efficiently, by then turning attention to the individual contributors that make up the Front and Rear Cornering Compliances, as described in [1].

Parameter Variation - In addition to analyzing the mainstream chassis design, evaluations were made of alternatives and variations. Checks were made to verify the sensitivity of LART and Yaw Damping to various chassis parameters. If a specific suspension architecture exhibited a tendency to allow wide variations in a particular suspension characteristic (i.e. Roll Steer), and the Handling Map indicated there was unacceptable degradation in LART, YVD or Understeer Gradient, that configuration would be rated poorly for robustness in handling performance.

Case Study - An example parameter evaluation is shown in Figure 3. On the Handling Map the baseline chassis tuning is shown, plus eight other points representing the variation in vehicle response due to Front and Rear Roll Steer variations of 0.08 deg/deg.

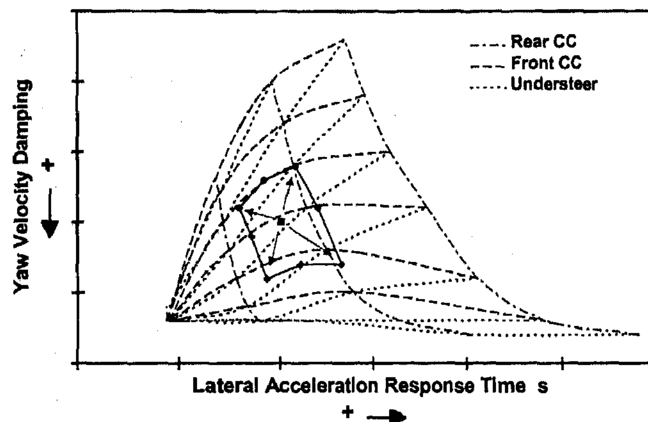


FIGURE 3. Roll Steer Variation Effects on Dynamic Behavior.

For proposed suspension "X", it was found that the assembly operation would allow geometric variation of point "Y" of +/- 1.0 mm which in turn produced a roll steer variation of +/- 0.08 deg/deg. Within the resulting performance envelope, the LART can vary as much as 0.05 s, YVD 8 percent, and Understeer Gradient 1.2 deg/g. Variations of this magnitude would be easily detected by the driver and result in vehicles with handling performance outside the acceptable range for C5. In conclusion, this configuration be rated poorly for robustness in handling performance.

RIDE - Lumped parameter simulations checked the viability of ride improvements. Tire stiffness, vertical damping and various suspension parameters were assessed for ride quality.

CONCLUSION

Using customer research, Engineering Direction, and competitive assessment, objective Ride and Handling goals were established for the 1997 Chevrolet Corvette. This was the first critical step in the Systems Engineering process used for suspension design. All subsequent design and development [1,2,3] was focused to produce a final vehicle that met all of the Ride and Handling targets.

Dynamic handling behavior was carefully assessed and specified in terms of Lateral Acceleration Response Time and Yaw Velocity Damping. From these targets, the Front and Rear Cornering Compliances were synthesized, establishing the basis for finalizing the kinematic and compliant properties of the suspension.

ACKNOWLEDGMENTS

The Authors wish to acknowledge the following individuals for their invaluable assistance and guidance in this phase of C5 engineering: Dave McLellan, Dave Hill, John Heinricy, Dave Wickman, Jim Lloyd, Al Bodner and Scott Allman of General Motors Corporation.

REFERENCES

[1] "Synthesis of Chassis Parameters for Ride and Handling for the 1997 Chevrolet Corvette," Fuja, S.P., Ryan, J.P., Schmid, H.A., SAE 970097.

[2] "Design Synthesis of Suspension Architecture for the 1997 Chevrolet Corvette," Schmid, H.A., Fuja, S.P., Ryan, J.P., SAE 970092.

[3] "Chassis Development of the 1997 Chevrolet Corvette," Neal, M. W., Dona, M. A., SAE 970098.

[4] "Objective Directional Response Testing," Kunkel, D. T., Leffert, R. L., SAE 885008.

[5] "Objective Evaluation of On-Center Handling Performance," Norman, K. D., SAE 840069.

[6] "Independent Wheel Suspension - Its Whys and Wherefores," Olley, M., SAE Transactions, March, 1934.

[7] "The 1997 Chevrolet Corvette Structure Architecture Synthesis," Longo, S., Moss, E., Deutschel, B., SAE 970089.

[8] "The Use of a Variable Response Vehicle in Handling Research," Bundorf, R. T., General Motors Research Laboratories, Research Publication GMR-455, 1965.

[9] "Vehicle Directional Control Motion Interpretation - Subjective and Objective," Bundorf, R. T., General Motors Research Laboratories, Technical Memorandum S-2254, July 15, 1966.

[10] "Driver Performance in a Driving Simulator with Varying Steady-State Vehicle Characteristics," Repa, B.S., et.al., General Motors Research Laboratories, Research Report EM-308, February 4, 1977.

[11] "Driver-Vehicle Performance in Simulated and Actual Driving," Repa, B.S., et.al., General Motors Research Laboratories, Research Report EM-325, June 2, 1977.

[12] "The Cornering Compliance Concept for Description of Vehicle Directional Control Properties", Bundorf, R.T., Leffert, R.L., SAE 760713.

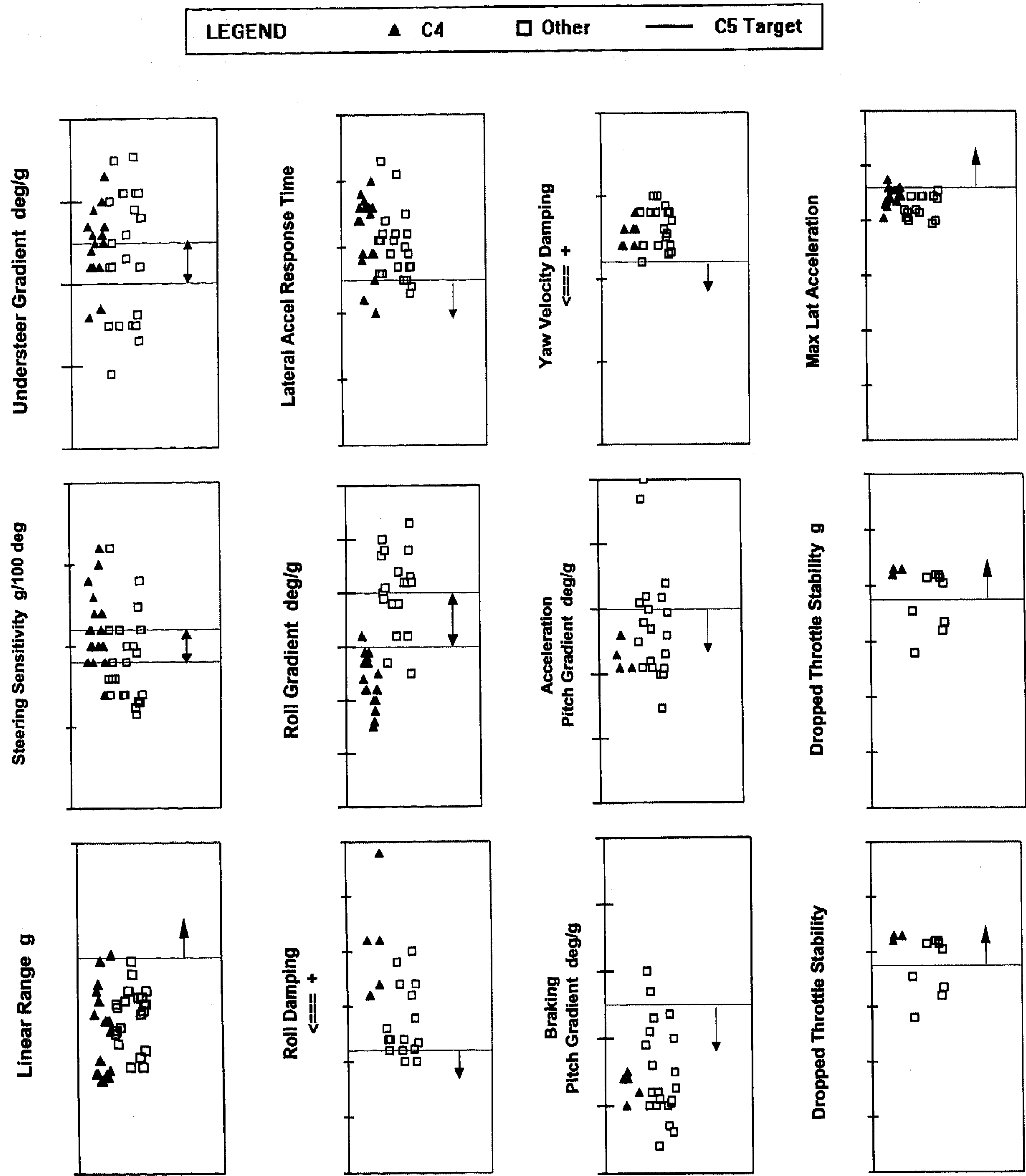
[13] "Vehicle Dynamics Synthesis Techniques for the Integration of Chassis Systems in Total Vehicle Design," Mackovjac, J, Lang, J, SAE 922104.

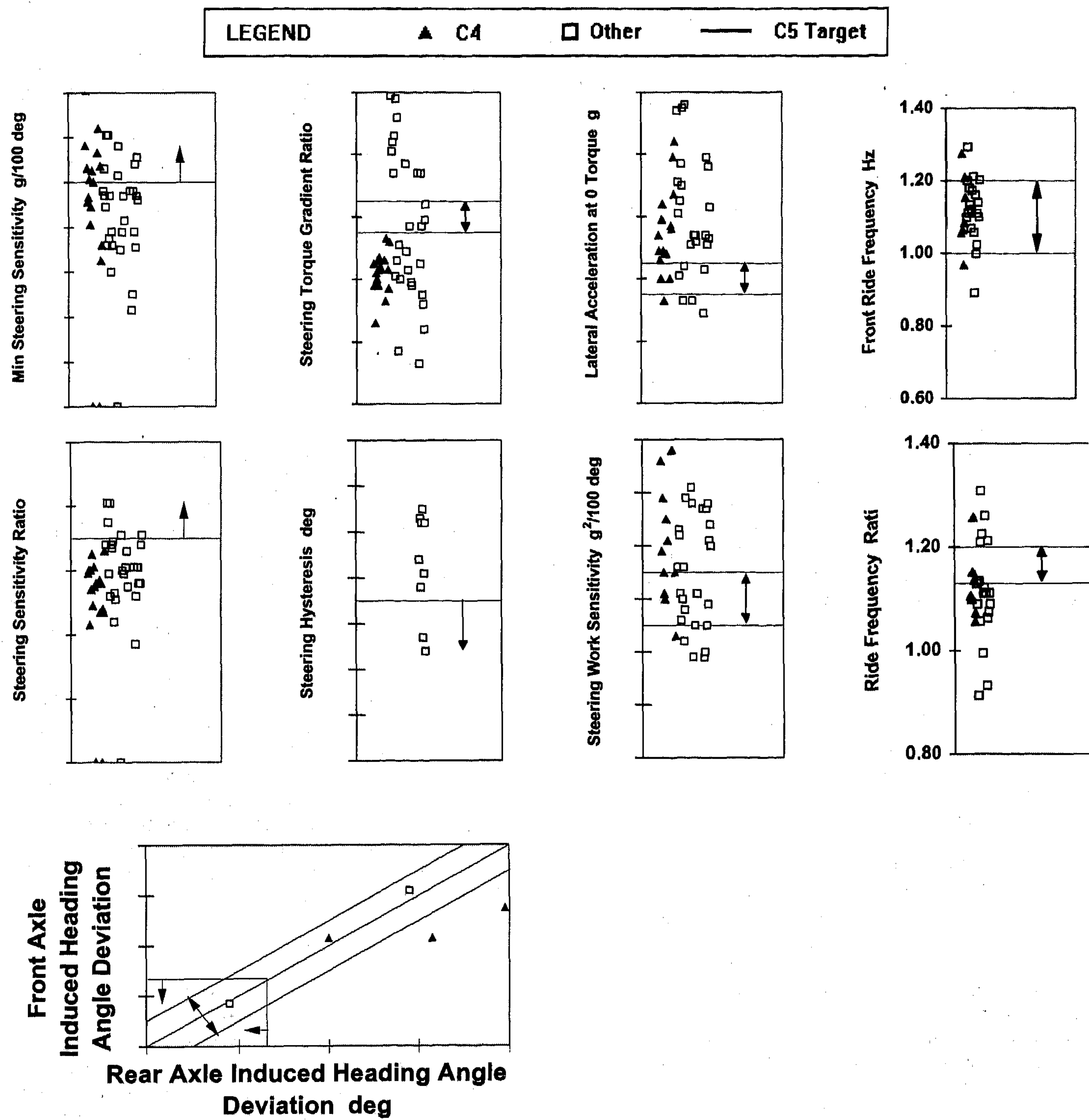
ABOUT THE AUTHOR

Joe Ryan was educated at General Motors Institute (BME 79) and the University of Michigan (MSME 85). He has worked for General Motors exclusively, in the NAO Vehicle Dynamics Laboratory, CPC Advanced Vehicle Engineering - Chassis, and most recently as Rear Suspension Design Release Engineer for the 1997 Corvette. He has been directly involved with Corvette since 1988, and C5 since inception. He also has designed, built, and raced formula and production based cars in club racing.

APPENDIX

Handling Metrics - Comparison Vehicle Measured Values and 1997 Chevrolet Corvette Targets





LEGEND
C4
Other
C5 Target
Min Steering Sensitivity g/100 deg
Steering Torque Gradient Ratio
Lateral Acceleration at 0 Torque g
Front Ride Frequency Hz
1.40
1.20
1.00
0.80
0.60
Steering Sensitivity Ratio
Steering Hysteresis deg
Steering Work Sensitivity g^2/100 deg
Ride Frequency Rati
1.40
1.20
1.00
0.80
Front Axle Induced Heading Angle Deviation
Rear Axle Induced Heading Angle Deviation deg

970092

Design Synthesis of Suspension Architecture for the 1997 Chevrolet Corvette

Henry A. Schmid, Steven P. Fuja, and Joseph P. Ryan
General Motors Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper describes the hardware execution of the front and rear suspensions of the all new 1997 Chevrolet Corvette. Topics covered include: alternative design trade-off, mass optimization, alignment and trim, structural interfaces, shared components, component design and a review of the overall design of the front and rear suspensions. Two case studies are detailed for the front upper and rear lower control arms. The systems engineering process used for suspension design is described throughout the paper.

INTRODUCTION

VEHICLE TERMINOLOGY - The fifth generation Chevrolet Corvette introduced in 1997 (C5) is a completely new vehicle, featuring significant design changes in all areas compared to the fourth generation Chevrolet Corvette of 1984-96 (C4).

DESIGN DOCUMENTATION - The first paper in the series [1] describes the initial step of setting vehicle level goals for ride and handling. The second paper [2] describes how the vehicle level targets are "rolled down" to specific subsystem attributes for the suspension, tires, and steering system. This paper describes the execution of the front and rear suspension in production hardware, in compliance with the specifications established in the previous steps. The last paper [3] describes the final development tuning of the suspension, tires, and steering systems.

Together, the four papers describe the Systems Engineering approach to C5 suspension design. Customer preferences, engineering input and evaluation selected the C5's Ride and Handling targets, which are rolled-down to suspension specifications, and finally, design and development of the production hardware. Careful execution of the engineering process has produced a 1997 Chevrolet Corvette with best-in-class ride and handling, assuring a high degree of satisfaction for every C5 customer.

ARCHITECTURAL TRADE-OFFS

In addition to engineering review of the C4 suspensions, a series of trade-off studies were conducted of new alternative designs. Ranking criteria included: ability to meet performance attributes described in [2], packaging, mass, cost, serviceability, manufacturability, assembly, and durability. Each design proposal was ranked in each category, weighting factors applied, and a summary ranking of each design produced. Weighting factors were higher for attributes that enabled best-in-class handling, improved ride, and low mass.

The required suspension Static Design Factors (SDFs) (i.e. roll steer, aligning torque compliance steer) [2] were formalized in a Subsystem Technical Specification (SSTS) which guided hardware design. As the overall vehicle architecture evolved and new engineering, manufacturing, and assembly strategies were applied, criteria and weighting factors were revised accordingly. Relevant comments on design trade-off are included in each section below.

MASS OPTIMIZATION

During phase one of the C5 suspension program the Engineering Team established a target mass reduction versus C4 for the front and rear suspension subsystems. Since the C4 already used premium low mass materials, further mass efficiency in the C5 was particularly challenging. Mass reduction initiatives implemented on the C5 were: aluminum bar pins in the front upper control arm bushings, aluminum inner metals in the rear upper control arm bushings, bushings pressed directly into control arms without outer metals, glass reinforced resin stabilizer bar links, thin wall stabilizer bars, mass optimized 36 mm (standard) shock absorbers, FEA modeling targeted at mass optimization and extensive loads measurement for accurate structural optimization. All fasteners, bushings and ball joints were reviewed and subject to resizing for minimum mass.

WHEEL ALIGNMENT

OVERVIEW - Wheel alignment is an important consideration for any vehicle. Successful optimization of ride and handling relies on precise and consistent control of tire orientation. For a sports car with high cornering stiffness tires, relatively high lateral forces are developed at low camber and slip angles, which may be fed back to the driver as steering leads and pulls. Wheel alignment is critical for good on-center steering feel, minimized tire wear, steering isolation, vehicle insensitivity to uneven roads and ultimate handling performance.

GOALS - The overall vision of an alignment system consistent with the Corvette's sport car mission led to emphasis of four specific goals. First, 100% of shipped vehicles will conform to the four wheel alignment specification, verified by end-of-line inspection. Second, cross-caster and cross-camber will be significantly reduced over C4 levels. (Customer surveys indicated a significant opportunity to improve customer satisfaction by reducing steering wheel leads and pulls.) Third, implement a robust alignment mechanism that is compatible with assembly line operation and will maintain service alignment. Fourth, allow easy adjustment for caster, camber and toe, and additional negative camber capability for autocrossing and racing.

ALIGNMENT SYSTEM REQUIREMENTS- Early variation simulation modeling quantified body and chassis dimensional build capability and the resulting variation in suspension alignment. When added to other vehicle data potential steering leads or pulls were predicted. Pull predictions were evaluated against a complaint steering wheel torque threshold developed from Human Factors studies. Component dimensional control, cross-camber, cross-caster, and trim height tolerances were developed to avoid customer complaint vehicles.

SUSPENSION IMPLEMENTATION - Cams on the lower control arms were chosen primarily to facilitate assembly process strategy, as detailed below. The C4 front suspension used shims on the upper control arm, which suffers from some inherent disadvantages: higher cycle time, open loop pattern adjustment, and fairly coarse resolution for meeting nominal specifications. Also, an adjustment mechanism in the upper control arm shortens the cross-car length, degrading kinematics and robustness.

The cam bolts react against machined vertical slots in the crossmember. The front suspension uses two cam bolts at the fore and aft bushings for caster and camber adjustment. The rear suspension uses one cam bolt at the forward bushing for camber adjustment.

Cam bolts provide continuous adjustment which is required for the simultaneous automated adjustment system at each wheel. Additional considerations were plant cycle time, ease of adjustment and access from below the vehicle. Additional negative camber adjustment (beyond production and normal service specifications) was provided to accommodate Corvette drivers involved in production class auto-crossing and road racing.

ASSEMBLY PROCESS - Three alignment process strategies were considered: net build, pattern build and custom set. The process strategy selected was a hybrid pattern set with limited custom set. The pattern set process was automated into the crossmember to control arm build station, where the lower control arm to crossmember position is held, and cam bolt rotation is unconstrained prior to tightening. The limited custom set process provides cycle time for an end-of-line adjustment of one side only to minimize cross-camber and cross-caster. At least one side of the vehicle must arrive at the alignment station within specification for camber and caster. (For the C4 front suspension, each side was pattern set with shims to a nominal camber and caster. Normal assembly processing did not provide for adjustment of cross-camber and cross-caster.)

Figure 1 shows process first time quality of front camber and caster for a group of pilot phase vehicles. The data shown is the minimum of the left or right side deviation from nominal. 100% of vehicles exhibit at least one side of the vehicle within specification for camber and caster. The end-of-line adjustment can focus on minimizing cross-camber and cross-caster by automatic adjustment of <u>one side only</u>. The C5 alignment strategy resulted in a 50% reduction in the cross-camber and cross-caster specification compared to C4. Toe continues to be 100% set at the end-of-line on all four wheels.

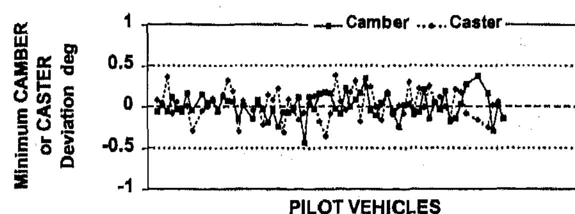


Figure 1. Minimum Front Camber and Caster Deviation, as Built, Pilot Phase.

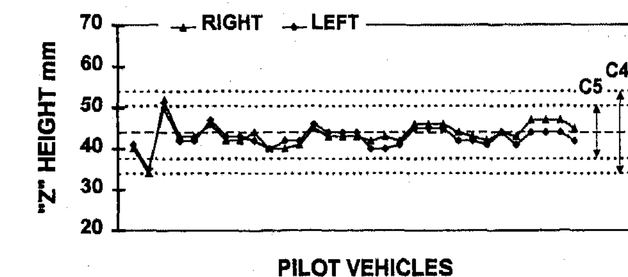
TRIM HEIGHT

OVERVIEW - Trim height control is particularly important for a sports car with somewhat restricted ride travel. Inherent features of a sports car that limit ride travel include: low profile styling, large tires, dense component packaging and minimal changes in vehicle loading. Effectively managing limited ride travel demands minimized trim height variation: available ride

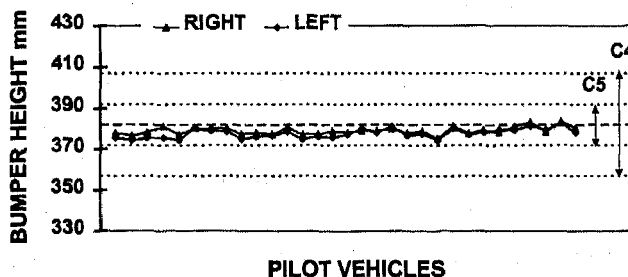
travel may be better utilized to avoid early jounce and rebound bumper engagement. Tire to wheel opening appearance is more consistent. Handling and ride optimization and robustness are improved due to less variation in suspension SDFs. Bumper design also benefits: C5 uses a unique bumper impact system, enabled by accurate control of bumper height - a direct beneficiary of accurate trim height control.

The improved trim height control of C5 allowed a 36% reduction in the trim height tolerance and a 60% reduction in the bumper height tolerance compared to C4. Figures 2a and 2b show front "Z" height and bumper height for a group of pilot phase vehicles. Also shown are the C4 and C5 tolerance bands for each.

STRATEGY - To meet these stringent requirements, variation simulation modeling indicated that a suspension height adjustment mechanism would be required. Contributors to trim variation include: mass variation due to option content, spring rate and load variation, dimensional variation and load offset from shock absorber gas charge. C5 suspension trim adjustment is implemented with threaded adjustable links that connect the spring to the lower control arm, as shown in Figures 3a and 3b.



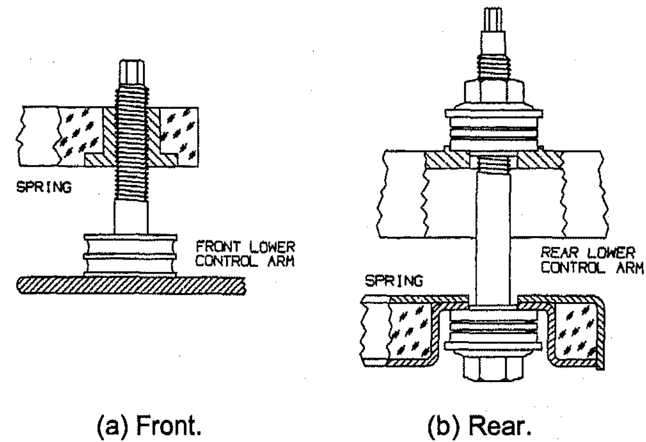
(a) "Z" Height.



(b) Bumper Height.

Figure 2. Front Suspension and Bumper Height, as Built, Pilot Phase,

ASSEMBLY PROCESS - Assembly plant process tooling engineers devised an automated spring link adjustment procedure into the crossmember to spring sub-assembly station. Vehicle mass due to option content is determined, shock absorber preload (due to gas pressure for optional units) is identified and individual spring load is measured. The length of each spring link is then automatically adjusted to insure each vehicle is at the proper trim height.



(a) Front. (b) Rear.

Figure 3. Suspension Trim Height Adjustment Mechanism

INTERFACE REQUIREMENTS

GLOBAL STIFFNESS -The C5 exhibits a very high overall structural stiffness for an open roof production car, enabling significant improvements in ride, noise, squeaks and rattles. Static torsional stiffness is 450% higher than C4. Comprehensive structural modeling and managing of modal requirements were key contributors to the significant improvement in C5 ride quality and isolation performance [4].

LOCAL STIFFNESS - Improved local body stiffness at suspension interface points benefit handling, ride and noise. Insuring that local body stiffnesses are significantly higher than the corresponding suspension bushing rates improves the flexibility and robustness of bushing rate tuning. The efficiency of the spring and shock absorbers improve, due to reduced series compliance. Isolation and noise benefit from high impedance of the natural load paths into the body: bushings become more efficient for energy absorption and isolation. Early analytical work identified major noise paths and local stiffness targets were established with an appropriate input mobility mismatch across the interface for each isolator.

ENGINEERING PROCESS - Successful development of the C5 body structure illustrates the importance of cross-discipline cooperation and setting appropriate interface requirements. Targets for each local stiffness were assigned, tracked, and managed through the course of C5 development. Each interface was developed with extensive structural modeling, verification tests were conducted on early prototypes and component were redesigned as required.

A more subtle benefit of properly selecting and verifying interface stiffness targets is accuracy and effectiveness of the preliminary ride and handling synthesis and analysis. Performance predictability is greatly enhanced if appropriate stiffness ratios between suspension bushings and structural attachments are maintained. As a result, early C5 development vehicles demonstrated very good ride and handling, and required relatively minimal development work to optimize performance and pleasability.

COMMON COMPONENTS

The low volume of the C5 offered a unique challenge to design the front and rear suspension with shared components. Higher volumes reduce piece cost and deproliferation of components reduces investment. The components shared by the C5 front and rear suspensions are:

Knuckles (left front = right rear)
Wheel bearings
Ball joints
Stabilizer bar links
Inner spring insulators
Spring mount clamps

Upper and lower ball joints are a low friction, lubed-for-life design. Frictional characteristics in both ride and steer modes were specified and monitored because of their effect on ride comfort and steering performance. Stabilizer bar links are dual ball joint to minimize system compliance and provide a linear roll response. The body of the link is glass reinforced resin for low mass and design simplicity.

ALUMINUM COMPONENTS

All upper and lower control arms and knuckles are aluminum for low unsprung mass. Manufacturing processes were selected for strength, shape requirements and cost. Castings have historically been least expensive and permit a large degree of flexibility in component shape. Forgings are typically more expensive, provide the highest strength and are best suited for simpler shapes. Alternative manufacturing process trade-off studies were completed for each structural suspension component.

Forging was chosen for the front upper control arm using AA6061-T6 aluminum, due to critical strength and shape requirements. A hybrid cast preform-forging process using A356-T6 aluminum was chosen for the lower control arms, rear upper control arm and knuckles for shape flexibility to minimize mass. This process provides improved surface material properties over castings for improved fatigue strength.

FRONT SUSPENSION

OVERVIEW - Continuing the Corvette tradition, the front suspension is a Short-Long Arm (SLA) configuration, as shown in Figure 4. Similar to the C4, control arms and knuckle are aluminum, and a composite transverse leaf spring and monotube shock absorber are used. New for C5, the lower control arms and steering rack attach to a hard mounted cast aluminum crossmember [5]. The upper control arm attaches directly to the hydroformed frame rail using aluminum bar-pin type bushings. The steering rack features Magnasteer™ II (a variable effort steering system developed by Delphi Saginaw Division of General Motors), is located forward of wheel center, and has cast aluminum outer tire rod ends. The stabilizer bar is packaged forward of wheel center, is mounted to the crossmember with low friction insulators and is attached to the lower control arm with dual ball joint links. Standard shock absorbers are high pressure monotube with digressive valving.

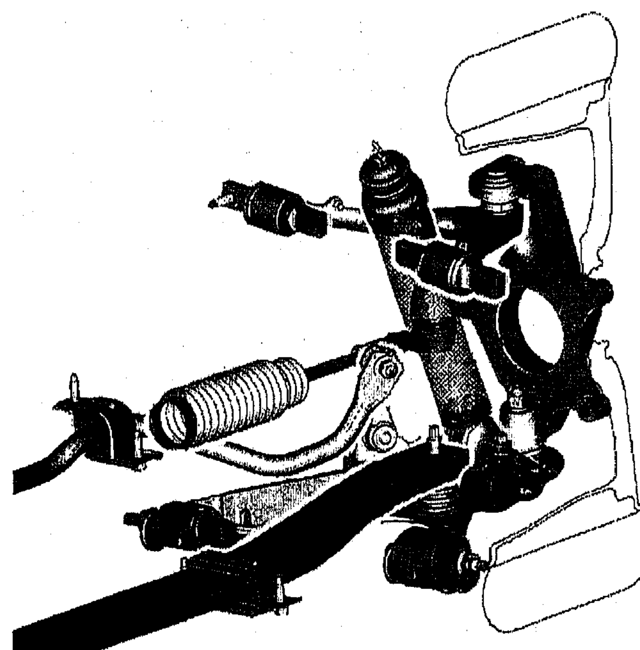


Figure 4. 1997 Chevrolet Corvette Front Suspension.

ALTERNATIVE SELECTION

Lower Control Arm - A strong direction from the Voice of the Customer was the desire for improved ride quality and impact isolation [1]. Extensive design activity focused on reducing fore/aft suspension stiffness to provide improved impact isolation. Comparison cars that use elaborate mechanisms to promote low fore-aft stiffness were benchmarked for performance attributes, packaging efficiency, mass, overall robustness and complexity. These more exotic concepts were evaluated against a conventional aluminum lower control arm. Mass, packaging efficiency, robustness, overall simplicity

and practicality led to the selection of a semi-L shaped aluminum lower control arm that carries the ride spring, shock absorber, and stabilizer bar attachments. Separation of bushing function (stiff handling and soft ride) was maintained to a high degree to decouple suspension fore/aft compliance and lateral stiffness. Subsequent ride and handling development [3] found that bushing tuning was highly flexible, and suitable fore-aft compliance could be achieved without degrading lateral stiffness.

Kingpin Geometry - Two basic kingpin geometry's were considered, differentiated by the upper ball joint and upper control arm location. The configuration chosen placed the upper ball joint inside the wheel and as far outboard as possible to provide a relatively short spindle length [2]. A high arm SLA was also considered. The two concepts were evaluated for characteristics such as steering feel, alignment robustness, fore/aft stiffness and hood height. The high arm SLA was considered unacceptable for reasons that included knuckle compliance and roll center geometry.

A synergy that evolved out of the alternative selection process was applied to the upper ball joint location. The typical placement of the upper ball joint body in the knuckle in a high arm SLA was adopted in the production design. This allows the longest kingpin length possible. Kingpin geometry, spindle length and scrub radius were developed to optimize steering feel and improve mid-range handling, both important customer desires [1,2].

Ride Spring - The first C5 suspension concepts were conventional coil over shock absorber for both front and rear suspension. One proposal utilized a horizontal coil over shock actuated by a bell crank upper control arm, and was evaluated in a mule vehicle.

A composite transverse leaf spring was selected for efficient packaging and low overall mass. Shock towers may be lower and smaller for hood height and less intrusion in the engine compartment. The shock absorber may be placed further outboard on the lower control arm than is possible with a coil over shock unit. The improved shock lever arm ratio provides efficient damping authority and reduced jounce bumper loads (the jounce bumper is integral with the shock absorber).

The gull wing shape of the spring was developed to fit the envelope between the ground line and engine oil pan at the center, and rise up at the tip to rest on the top surface of the lower control arm. An adjustable height insulator at the tip allows custom set of trim height, which is critical to the overall trim height strategy. The insulator is a low shear rate rubber-steel sandwich that reduces the spring contribution to suspension fore-aft rate for improved impact isolation.

Brakes - Opposed piston calipers were considered at one time, but were not pursued because of adverse packaging implications (longer spindle length and excessive positive scrub radius). Floating (single side piston) calipers were chosen to permit better optimization of the kingpin axis, minimized spindle length, and minimized scrub radius.

CASE STUDY: FRONT UPPER CONTROL ARM

The front upper control arm is tightly packaged between the frame rail, road wheel and shock absorber. The C5's longer wheelbase and tighter turn diameter necessitated larger steer angles at full lock, further reducing packaging space. In addition to tight packaging, the front upper control arm is also subjected to high loads under severe events such as a large pothole or panic brake stop. Due to the orientation of the upper ball joint, high bending loads are also present, unlike the C4 front upper control arm.

Extensive shape optimization was conducted with automated finite element analysis to arrive at the final shape. As with any optimization problem, system constraints must be accurately defined to produce a meaningful answer (load environment, package restrictions, and handling-related stiffnesses). The maximum stress level, cross-section envelope and relatively simple V-shape, led to selecting forging as the manufacturing process. The final design that meets all the packaging, stress, and mass criteria is shown in Figure 5.

The upper control arm is attached to the frame rail by means of aluminum bar-pins integral with the pivot bushings. The bar pin design allows a low profile attachment to the hydroformed frame rail, maximizing the cross car length of the upper control arm, and improving front view geometry.

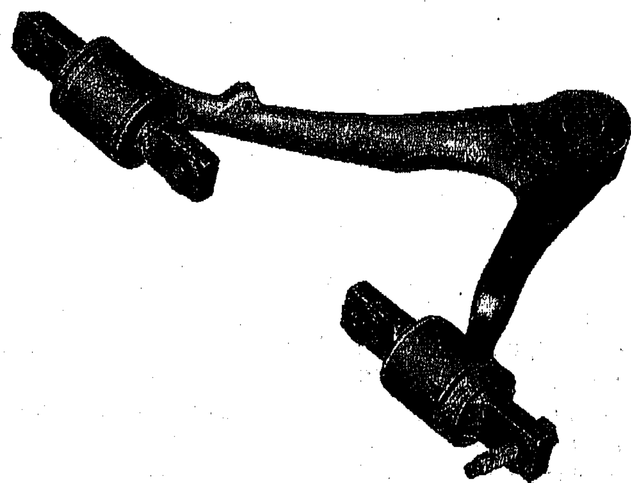


Figure 5. Front Suspension Upper Control Arm.

REAR SUSPENSION

OVERVIEW - The C5 rear suspension is shown in Figure 6. There is strong similarity to the front, with the common knuckle and semi-L shaped lower control arm. The lower control arm and toe link mount to a cast aluminum crossmember [5]. The upper control arm attaches to vertical brackets on the hydroformed frame rail. The knuckle is oriented with zero caster. The shock absorber has a clevis lower mount that mates to the lower control arm. The composite transverse leaf spring is clamped to the crossmember and attached to the lower control arm with rubber isolated tension links. The stabilizer bar is aft of wheel center, clamped to the crossmember (sharing one fastener with the lower control arm rear bushing) and is attached to the lower control arm with dual ball joint links. Steel plunging halfshafts with constant velocity joints deliver torque to the wheels.

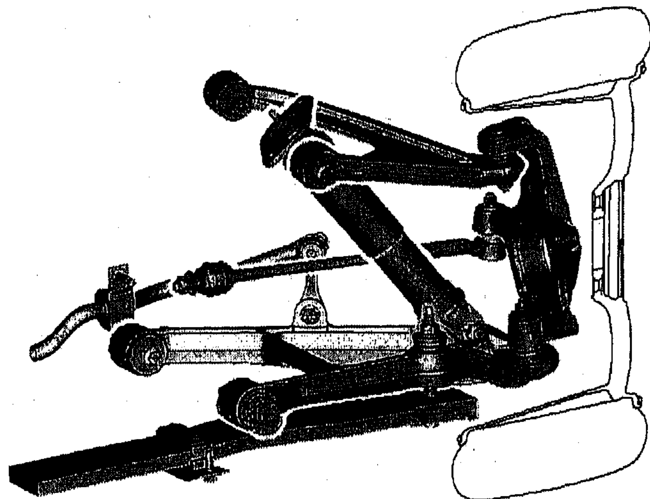


Figure 6. 1997 Chevrolet Corvette Rear Suspension.

C4 PERSPECTIVE - As the C4 matured over the years, it became evident that the rear suspension had reached it's full potential for ride and handling optimization. The fundamental layout dates to 1963, and exhibits very good packaging and unique dual-function use of the half-shaft as a suspension link. With the lateral suspension links attached to the differential, tuning the differential mounts for driveline isolation typically degraded handling, and vice-versa. Dual trailing links controlled the knuckle in the fore-aft and pitch directions. It was difficult to reduce suspension fore-aft stiffness without producing excessive knuckle caster change and ride steer variation. The spring and shock absorber were attached directly to the knuckle a significant distance away from wheel center in side view. The ride spring load added static pre-load to the trailing link bushings. It was difficult to tune the shock absorber because of series compliance and phase lag caused by caster deflection in the knuckle. The half-shaft as the upper control arm of the suspension resulted in a relatively short kingpin axis and sub-optimized roll center geometry. Ride camber geometry and robustness to camber alignment contributed to higher sensitivity to uneven and cambered roads.

ALTERNATIVE SELECTION

Criteria - Numerous packaging and SDF analyses were conducted for each proposal to optimize factors such as packaging, cost, ride and handling performance, interface loads, overall vehicle mass and shock absorber efficiency.

Suspension Type - A survey of all rear suspensions used in high end sports cars today reveals there are a multitude of architecture solutions in use. Some give priority to packaging, others to cost by sharing components with high volume vehicles, and a few to handling performance but typically at the expense of ride.

The top priority of the C5 rear suspension was ride and handling performance. With careful attention to detail, and recognizing that many design features that enable excellent handling also benefit ride, there was full expectation that outstanding ride could be achieved without degrading handling.

The design team concluded that a SLA suspension provided the only acceptable solution for the handling and ride requirements of the C5. A high degree of decoupling between front view and side view geometry, fore-aft and lateral stiffness, camber, caster and toe adjustment, ideal roll center control and robustness to all of the above may be realized in a carefully designed SLA. The C5 rear suspension does not share components with other vehicles, to eliminate any possible performance compromise.

The Corvette driver ranges from the daily commuter to the weekend racer. Toe and camber may be easily adjusted to maximize handling performance without significant impact on other suspension characteristics. Note that virtually every modern, purpose built independent suspension for racing is a SLA because of kinematic decoupling and adjustability.

Control Arms - The L shaped upper control arm design stabilized early. The stiff (handling) pivot bushing is on wheel center, and the soft (ride) pivot bushing is as far aft as could be packaged.

Lower control arm alternatives ranged from a conventional A-arm to a 3-link configuration (trailing link and two lateral links). Correspondingly, knuckle alternatives included sharing with the front or a new design for the 3-link lower arm. New knuckle designs were considered with and without direct attachment of the spring or shock absorber.

Direct attachment to the knuckle in a low profile package requires components to be ahead or behind the drive axle. The resulting fore-aft moment arm for spring or damping transient forces generates caster deflection in the knuckle and unwanted toe change, especially on rough roads. The pitch stiffness of the knuckle becomes a series compliance for the ride spring and shock absorber, degrading tunability.

The concept selected was a semi-L shaped lower control arm that carries the spring, shock absorber and stabilizer bar, mated to a common front/rear knuckle. Commonizing the knuckle eliminated the investment and engineering overhead of another major component. An adjustable, dual ball joint toe link located aft of wheel center simplifies alignment, and provides deflection understeer [2].

Shock Absorber - Following selection of the lower control arm, extensive packaging, load path and SDF analyses were used to locate the shock absorber. Travel ratio, inclination angle, shock tower mass and stiffness were considered. The low profile package was not compatible with a clevis around the drive axle. The shock absorber was attached to the lower control arm as close as possible to the lower ball joint to minimize the series compliance effect and induced loads in the soft ride bushing. Simultaneously, various spring alternatives were evaluated and their influence on shock absorber location.

Spring - Three basic spring configurations were considered: transverse leaf, coil over shock, and separate coil. Attachment was considered either to the knuckle or to a lower arm/link. Following the decision to attach the spring and shock absorber to the lower control arm, the composite transverse leaf emerged as the best alternative, permitting the best shock travel ratio, lowest mass and the smallest shock tower which contributed to improved trunk volume.

The transverse leaf spring is suspended below the lower control arm with rubber isolated tension links (Figure 3b). The links decouple the fore-aft stiffness of the spring from the lower control arm, for improved impact isolation and are threaded to allow custom trim set.

CASE STUDY: REAR LOWER CONTROL ARM

The lower control arm underwent extensive shape evolution to accommodate packaging, loads and manufacturability. Finite element models and sand cast prototypes were used to validate the final design.

The rear suspension lower control arm (RLCA) evolved significantly from Alpha to production design. During design development, the only imperative was to maintain the overall topology (pivot point and attachment locations). If the suspension performance parameters specified in the Subsystem Technical Specification [2] were preserved, the macro performance of the RLCA in the suspension would not change. The design team had complete freedom to revise the structural shape and attachments to meet mass, durability and packaging requirements with no risk to suspension performance.

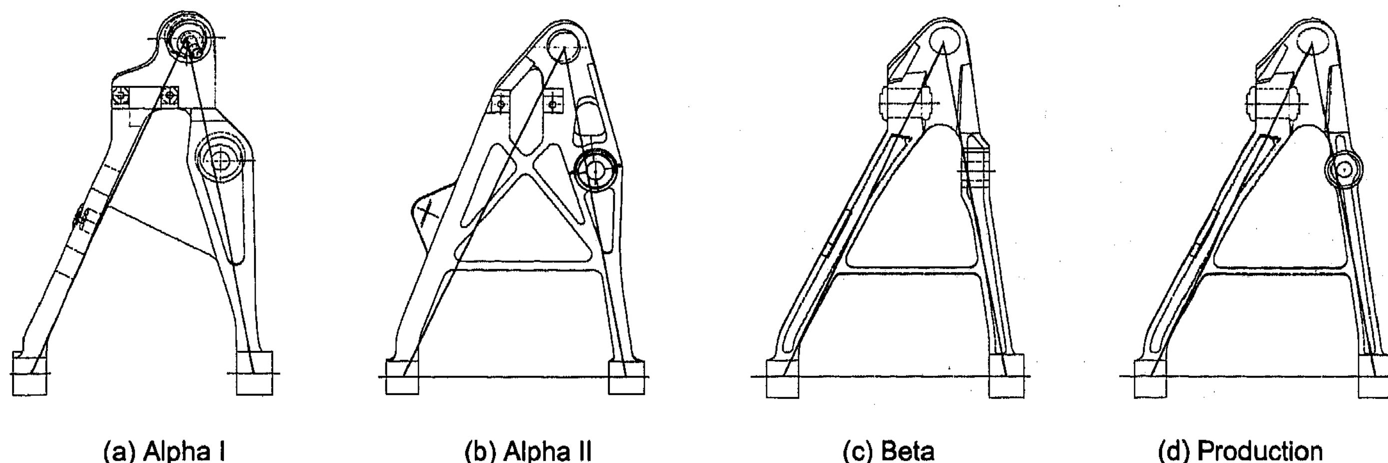
Several load cases were analyzed with FEA, including severe pothole, maximum braking and high cycle fatigue. The Corvette is required to pass the same Proving Ground durability tests as all other passenger cars and light trucks.

ALPHA DESIGN

Structural Shape - Figure 7a shows the Alpha RLCA. Alpha components are intended to prove system functionality and do not reflect any attempt at detailed optimization. Overall package optimization located the pivot points attachments for the shock absorber and spring, which did not change through to the production design.

Initial shape optimization was based on the Alpha design. Figure 7b shows the evolutionary Alpha II design. In plan view, lines of action are shown between the load reaction points (pivot bushings and ball joint). The web between the front and rear legs and the area around the shock absorber attachment was replaced with a truss type structure. The rear leg was revised to distribute material more symmetrically about the ideal load line in plan view. Clearance required for the shock absorber lower mount prevented using a significant amount of material directly on the ideal load line where the highest bending moments occur. Extensive FEA work on the Alpha II design concluded that a reasonable mass reduction could not be achieved.

Shock Absorber Attachment - With the integral jounce bumper, the highest loads applied to the RLCA are at the shock absorber attachment in a severe pothole event. The Alpha lower mount was a commonly used "bowtie" type. To accommodate this fastener configuration, material must migrate away from the ideal line of action. In rear view, the shock absorber is inclined significantly, and jounce loads into the RLCA have a large outboard component. In the Alpha design, two vertical 8 mm fasteners clamped the lower mount to the RLCA. In a severe pothole, the shear load in the interface exceeded the clamping capacity of the 8 mm fasteners. (Observation of an Alpha durability vehicle confirmed this.) To handle the shear load, 14 mm fasteners would have been required. To accommodate 14 mm fasteners, the lower mount would have to increase in size and more material be added to the RLCA away from the ideal line of action.



(a) Alpha I (b) Alpha II (c) Beta (d) Production

Figure 7. Rear Lower Control Arm Evolution.

BETA DESIGN

Criteria - At this time, three factors indicated a major revision to the shape of the RLCA was necessary. First, the desired mass reduction was not being realized in the basic Alpha shape. Second, selection of the cast preform-forging process was not well suited for the Alpha II design. Third, the shock absorber attachment was becoming large and heavy. To address these issues, a very different approach was taken for the Beta RLCA, as shown in Figure 7c.

Shock Absorber Attachment - The lower mount bushing was moved to the RLCA and a clevis added to the bottom of the shock absorber. Placing the bushing in the RLCA was well suited to the optimized section depth at this location. A single 16 mm fastener was used, which provides sufficient clamping force to withstand the severe pothole load. The number of fasteners per side was reduced from four to two, with the potential of reducing to one by using a weld nut on the clevis. In the first Beta design, a clevis attachment for the ride spring was also investigated.

PRODUCTION DESIGN - In the production RLCA shown in Figure 7d, the spring attachment was changed from the clevis concept to a grommet and link bolt. This method of attaching the spring has better articulation in all directions to decouple the fore-aft rate of the spring from the RLCA, and accommodate the lower control arm plan view displacement that occurs with camber adjustment. The stabilizer bar attachment was refined to allow use of the same link as the front suspension. The net mass savings to the vehicle for the production RLCA system was approximately 2 kg, compared to all other alternatives considered.

CONCLUSION

The suspension design of the 1997 Chevrolet Corvette was the product of a dedicated design team that used state-of-the-art computer techniques, prototyping, close cooperation with suppliers, laboratory and mule vehicles while following a logical and comprehensive methodology. The C5 systems engineering process began with the voiced desires of the customer and guided the design and development through to the production designs. The final suspension designs provide ride and handling characteristics that meet or exceed targeted performance. Improvements were made for durability, manufacturability, enhanced alignment and trim height control and low mass. Compared to C4, tolerances were reduced for trim height by 36%, bumper height by 60% and cross-camber and cross-caster by 50%.

The net result of the design effort for the 1997 Chevrolet Corvette suspension systems is an uncompromised balance of ride and handling that re-defines Corvette performance and continues it's unique heritage.

ACKNOWLEDGMENTS

Evolution of the C5 suspension designs were the result of an exceptional team effort by a group of individuals too numerous to recount. The authors would be remiss, however, if they failed to acknowledge the contributions of the following General Motors individuals: Al Bodner, Jerry Ferranti, Kasi Goundan, Cliff Jao, Rich Kowalczyk, Scott Larson, Jim Lloyd, Dave McClain, Mike Neal, Norm Sawaya, Rex Stump, and Dave Wickman.

REFERENCES

[1] "Objective Ride and Handling Goals for the 1997 Chevrolet Corvette," Ryan, J. P., Fuja, S. P., Schmid, H. A., SAE 970091.

[2] "Synthesis of Chassis Parameters for Ride and Handling for the 1997 Chevrolet Corvette," Fuja, S. P., Ryan, J. P., Schmid, H. A., SAE 970097.

[3] "Ride and Handling Development of the 1997 Chevrolet Corvette," Neal, M. W., Dona, M. A., SAE 970098.

[4] "The 1997 Chevrolet Corvette Structure Architecture Synthesis," Longo, S., Moss, E., Deutschel, B., SAE 970089.

[5] "The 1997 Chevrolet Corvette Suspension Crossmember," Brown, K., Juras, P., SAE 970370.

970094

Methodology for Effective Design Evaluation and Durability Approval of Car Suspension Components

Vatroslav V. Grubisic and Gerhard Fischer
Fraunhofer-Institut für Betriebsfestigkeit LBF

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

In this paper the criteria and the requirements for the approval of the service strength and the durability of vital car suspension components are treated. The overview of the main influences on the fatigue life is given and the methodology for the design evaluation and a reliable time and cost saving development, including durability approval, are described. These include:

- Description of representative, customer related loading
- Procedure for design evaluation based on customer usage
- Corresponding test programs for durability approval including the requirements for statistical validation and quality assurance of manufacturing.

On two examples - a wheel and a suspension arm - the procedure including individual steps is explained.

INTRODUCTION

Criteria and Requirements for Design and Durability Approval

Trends in vehicle development in recent time are influenced by opposing requirements - on one side, the vehicle should have high efficiency and performance with low impact on environment (low emission, noise, waist at the life end), and on the other side an improved reliability and safety should be realized. Finally, these achievements should be accompanied by low production and possibly also low maintenance (service) costs.

To satisfy all of these requirements a light-weight design is of paramount importance. But for the light-weight design approach it is essential that the required service life is assured under all possible usage conditions, especially for the components for which failure may cause accidents. These are most of the vehicle suspension components like wheels, hubs, spindles, suspension arms etc.

The most important criteria for the design of such vital components are the attainment of the expected **service life** with the required **reliability** and **safety,** whereby applicable standards and regulations must also be observed.

The **service life** depends primarily on the loading conditions in service [1, 2]. Figure 1 shows schematically how the character of service loads will influence the service life. In this figure the fatigue life *L* curves are presented in relation to the stress amplitude S_a for different stress (loading) spectra, as a typical S-N-diagram, whereby the stress amplitude $S_{a,l}$ is based on the components' endurance limit for which the value is $S_a = 1$. It can be seen that for a maximum stress amplitude $\bar{S}_{a,o}$ of the load-time history resulting in a spectrum, called also cumulative frequency distribution [3, 4, 5] of type *A* (constant amplitude loading) the fatigue life will be, for example, 200 times lower compared to the varying amplitudes with the spectrum *B* or 2000 times lower compared to the spectrum of the type *C*. For the design spectrum of the type *D*, consisting of two individual spectra that are valid for specific loading conditions, like cornering and straight driving for wheels [6] the fatigue life will be about 10^4 times higher than in the case of spectrum *A*.

If the comparison is made for an allowable stress in the case of a design life L_o (corresponding to e.g. 300 000 km for passenger cars which lead to about $2 \cdot 10^8$ cycles as shown in Fig. 1), which is related to the endurance limit ($S_{a,A} = 1$ being allowable stress for spectrum *A*) the values will be

$\bar{S}_{a,B} = 1.3$ means the allowable stress for spectrum *B* will be 30 per cent higher than the endurance limit,

$\bar{S}_{a,C} = 1{,}7$ the allowable stress for spectrum *C* is 70 per cent higher than the endurance limit, or even

$\bar{S}_{a,D} = 2.1$ the allowable stress for spectrum *D* is 110 per cent higher than the endurance limit.

These relations depend besides on operational loading also on material properties [2, 7] and design data (stress distribution, stress state). But for an optimum light-weight design the loading spectrum related to usage conditions is of paramount importance.

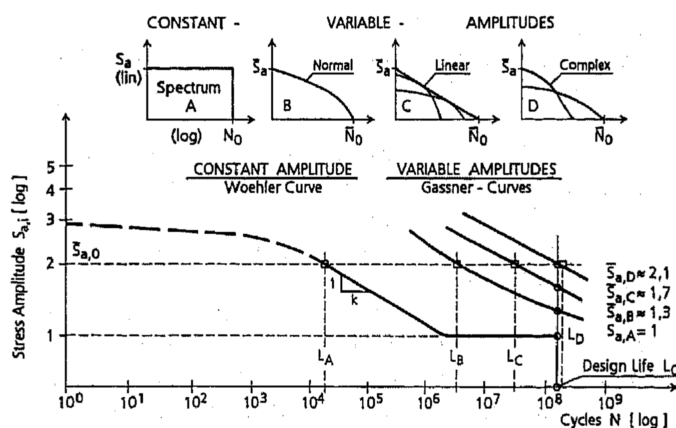


Fig. 1: *Influence of the Load (Stress) Spectra on the Fatigue Life and on the AllowableStress*

It must be regarded that for a reliable evaluation of the service fatigue life the probability of occurrence of the design spectrum must be defined [2]. The scatter in service loading is relatively high. This is influenced by the vehicle usage, driver habits and capability and operational conditions. Some vehicles are used predominantly for long distance driving on highways and good roads achieving several hundred thousand of kilometers (spectrum *a* in Fig. 2); others are used mainly on country roads with a high amount of bad road sections and bends, and finally some only in cities with a high number of braking and accelerating operations (spectrum *b* and *c* in Fig. 2).

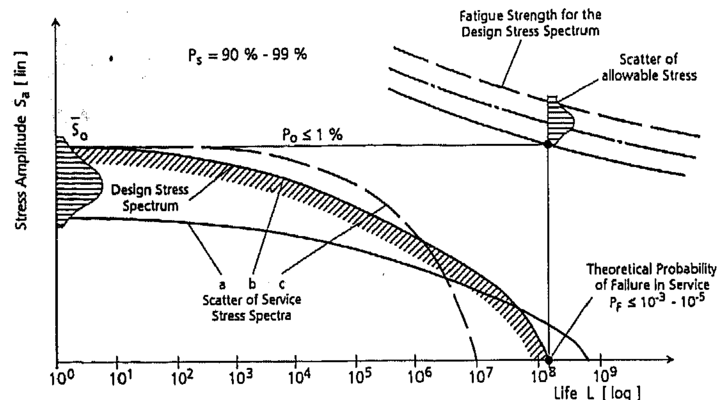


Fig. 2: *Scatter of Service Spectra and the Definition of the Design Spectra*

For reliability analysis a design spectrum must take into account both scatters. This should be done assuming a **design life L_o** and a **standardized spectrum** (e.g. spectrum *b*), whereby the scatter is determined by recalculation of the loading intensity (damage) of individual service usage spectra using the damage accumulation procedure (Fig. 4). For vital suspension components it is absolutely necessary to use a "hard" design spectrum, which has a very low probability of occurrence (*P_o < 1 %*) and which includes maximum values of operational loads $\hat{S}_a$ for each individual loading case; this is shown schematically on Fig. 2.

Based on the probability of occurrence P_o and the scatter of components fatigue strength, the **probability of failure P_F** in service could be determined by an adequate statistical evaluation:

If, based on the scatter of the fatigue strength, a fatigue life curve with a probability of survival *P_s = 90 per cent* (failure criterion: an initial crack of the component with additional fatigue life in crack propagation before total fracture) or *P_s = 99 per cent* (failure criterion: fracture of the component) is used, the resulting **theoretical** probability of failure for the component will be *$P_F < 10^{-3}$ to 10^{-5}*.

This approach assures that the component under consideration will not fail during the expected service life in a way that may cause an accident [2]. By this procedure also the maximum operational loads of normal usage are included in the design evaluation and test programs, being of importance with respect to the structural yield point and possible global plastic deformations of investigated components.

Additional to the design spectrum it is necessary for some vehicle component to take into account some specific overloads generated by special events or even accident-like loading. These loads are predominantly impact loading and do not influence directly the fatigue behavior; their influence must be separately controlled, as it is the case in regulations for cast aluminum wheels [6, 8].

To fulfill the requirements concerning the required durability life it is important that the failure criteria is adequately determined. A definition that "no fatigue cracks" should occur under the usage is not possible for an up to date design, due to technical and economical reasons. As shown on Fig. 3, for a representative load spectrum the durability life is determined based on a certain failure criterion, called technical crack, detectable with simple means like dye check or even by the naked eye.

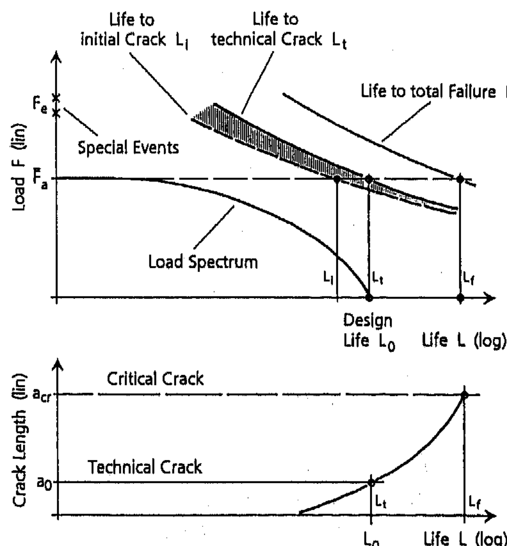


Fig. 3: Failure Criterion and Durability Life

The **safety**, an important criterion, is largely determined by the components ability to cope with unexpected over loads F_e without catastrophic failures. It is of decisive importance for the safety components that under overloading total fracture doesn't occur. This overloading, occurring outside the normal usage mission, is often defined as a **misuse**, and therefore assumed that it could be neglected by the design. This could be a great risk from the legal point of view if this concerns a vital, safety component. Generally, these high impact loadings, which occur e.g. by driving with higher than allowable speed over speed bumps or if a vehicle hits a curb in a bend, is defined as "a special event loading" and must be covered by the design and the material selection, nevertheless they do not influence the fatigue damage and are therefore not included in the design spectrum and durability approval.

For the criteria concerning the allowable technical crack in durability tests it should be kept in mind that, beside the requirement that such a crack should have a low crack propagation rate, it should also not lead to a component's total fracture if loaded by a **special event loading.** The requirements concerning the **special event loading** differ at individual vehicle companies and it will be beyond the scope of this paper to treat them.

Because the existing procedures for fatigue life estimation are of limited accuracy, the final durability approval must be carried out experimentally. This is done as well by driving with vehicles over proving grounds as by tests in laboratory. To decrease the development time and to achieve a light-weight design already in a very early stage, tests under simulated operational loading are necessary.

To carry out such tests an appropriate test facility, in which the **generated component's deformations correspond to that under operational loading** and a **reliable test program,** must be available.

A vehicle prototype for service measurements to derive a drive signal for testing is usually not available at this stage of development. Using standardized test programs [2, 6, 8 to 13] the preliminary durability approval can be carried out. Such programs are adapted to individual vehicles adjusting "basic program" by factors which depend on vehicle data and its representative usage.

When a vehicle prototype for road measurements is available, the assumed loads for durability tests can be controlled by simple short-time measurements with wheel load sensors [14] and the test data verified. If the used and measured data differ, it is possible in most cases by recalculation of achieved results, to estimate the expected fatigue life for the measured loads.

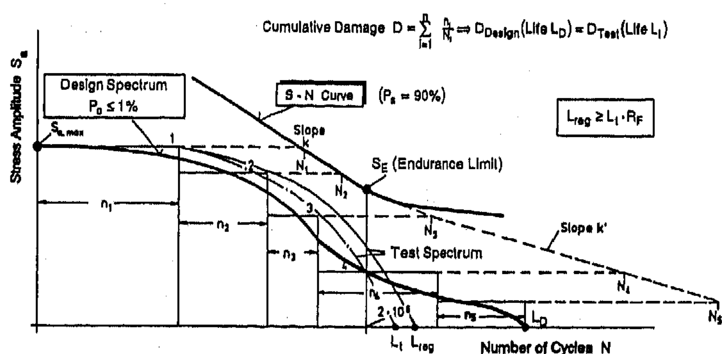


Fig. 4: Damage Fatigue Life Validation for Design and Test Spectrum by Modified Palmgren-Miner Damage Accumulation Calculation

To carry out the durability approval the damage accumulated in the test must be accelerated in relation to service [13]. The fatigue damage under accelerated test must correspond to the damage under customer usage. This means that the fatigue damage of the design and test spectrum must be equal, what could be verified by a relative fatigue damage calculation, as shown in Fig. 4. To take into account all possible damage influences, e.g. fretting corrosion, the test acceleration must be limited to a certain number of cycles, which should be achieved in tests to prove the durability life. For suspension components the required number of cycles should be not lower than $2 \cdot 10^6$ *to* $7 \cdot 10^6$ depending on the individual case [11 to 13].

Generally, the test acceleration should be based on the adjustment of the load spectrum in medium and high load levels and on omission of non-damaging low intensity loads, having high frequency of occurrence and predominantly originating from operational conditions during driving over smooth roads. The test acceleration by increasing maximum operational loads should be avoided, because it could change the fatigue damage development due to possible plastic deformation not occurring in service.
To meet the reliability requirements several durability tests should be carried out. Using a risk factor, based on the statistical approach [13] the required test life should be determined:

$$L_{req} \geq L_t \cdot R_F$$

L_t = Fatigue life under the test spectrum generating the damage equal to that under design spectrum

R_F = Risk factor, which takes into account possible misjudgment due to low number of tests and scatter of the components fatigue strength

Usually for durability life approval of vital suspension components at least three tests should be carried out, whereby a risk factor of $R_F = 1.4$ *to* 1.5 should be used [13, 16].

METHODOLOGY

The design evaluation and durability approval must include all parameters decisive for the service strength. An effective development must be a part of simultaneous engineering in the development, which includes a **design** and a **testing phase**. In the design phase, besides the stress analysis, the material selection including corresponding manufacturing procedure are also included. In the testing phase the approval of the design including the influence of the manufacturing on the components fatigue strength is carried out (Fig. 5).

This procedure guarantees a time saving, reliable design evaluation and durability approval including manufacturing influences.

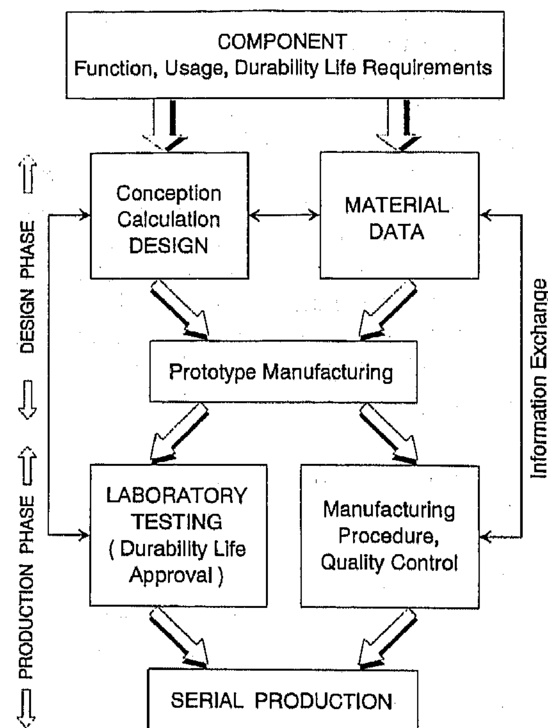


Fig. 5: *Procedure for an Effective Components Development*

To determine the durability life and to achieve an optimum design it must be known:

- the **operational loading** of the suspension under individual driving conditions
- the **allowable stresses**, which depend on material, manufacturing, design, and stress spectra.

OPERATIONAL LOADING -The main influences on the load spectra are: the **usage,** the **operational conditions** and the **dynamic behavior** of the vehicle including suspension. Decisive for the **usage** are vehicle utilization and driver, for the **dynamic behavior** it is vehicle design and tire properties, and for the **operational conditions** it is quality and type of roads.

If the data about usage and operational conditions are known, then, to determine the load spectra of a new vehicle, it is mainly necessary only to adapt the loads related to the vehicle data; the most important vehicle data are: dimensions, rated wheel loads, driving characteristics and vibratory behavior.

In service the suspension is subjected to variable loads, which originate from the road roughness and maneuvers like braking, steering or accelerating a vehicle (Fig. 6). The distinction between these sources is important for the derivation of the design or test spectra and for the design evaluation. The operational loads could be determined either by calculation based on the vehicle data or experimentally using a sensorized suspension or a measuring wheel (Fig. 6), whereas for a reliable load determination the experimental approach is often necessary.

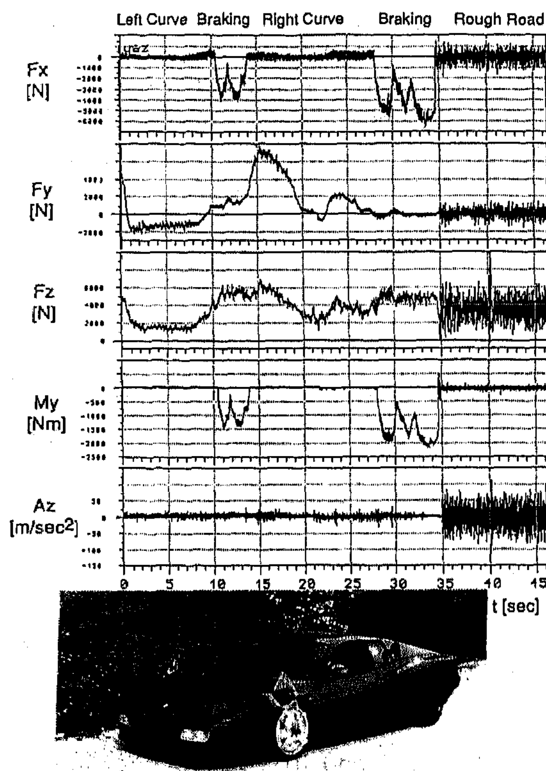


Fig. 6: *Operational Wheel Loads Sensed by Measuring Wheel VELOS 8 J x 18*

For the design evaluation the load-time histories, as shown on Fig. 6, must be transferred to the stresses generated on individual points of components. The stress-time history on a component depends both on the external loads and the responses of individual components to these loads. Basically, we have different response of the suspension rotating and non-rotating components as it can be seen on Fig. 7. For the rotating components the stress-time history is influenced by the cycles generated through rotation (these stress cycles are generated even at driving on ideal flat road surface at which condition non-rotating suspension components are stressed only by a possible prestress due to the payload) and superimposed stresses due to the variable operational loads. Also the mean stress of the rotating components doesn't change in maneuvers like cornering while this could be significantly changed on the non-rotating components, as it can be seen on Fig. 7.

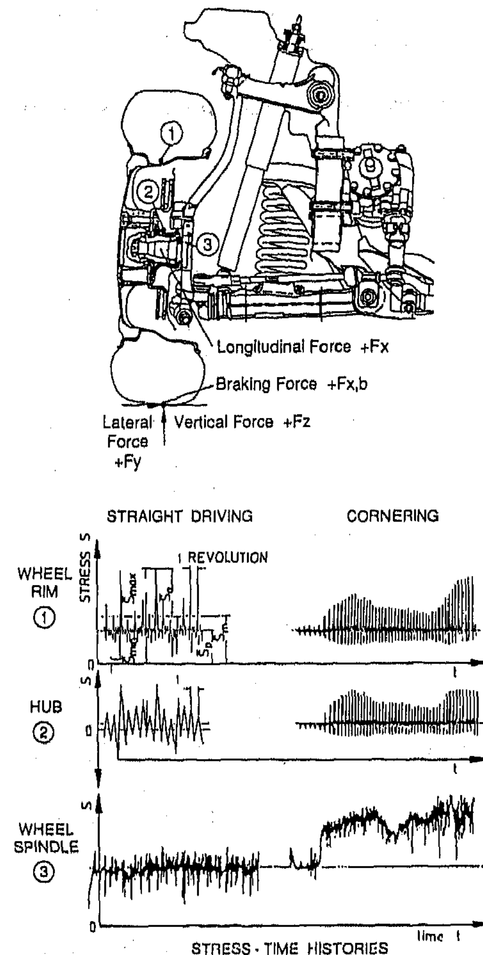


Fig. 7: *Stress-Time Histories on Individual Suspension Components*

Because the operational stresses vary at random, their evaluation is based on statistical counting methods [3 to 5]. The counting methods being of importance for the fatigue life evaluation and presentation of operational spectra are range pair ("rain flow counting") and level crossing counting, by which the ranges can be derived. It would be beyond this paper to treat the counting methods, used for fatigue life evaluation, in detail. It could be summarized that for the fatigue evaluation of suspension components, achieving more than 10^8 cycles during required operational life, a spectrum based on amplitude distribution (Fig. 8) fulfills the conditions for a reliable fatigue evaluation. This spectrum shows how often (value N) the stresses (stress amplitude $S_{a,i}$) have been reached or exceeded so that the analyzed stress-time history are classified only according to magnitude and number of occurrences. For the stress-

time histories with the mean value shift, like that shown for wheel spindle in the cornering (Fig. 7), the amplitude spectra of the range pair and the level crossing counting will differ, so that for the fatigue life evaluation (damage calculation) the mean value of both spectra should be used.

In Fig. 8 the main parameters of such a spectrum are presented.

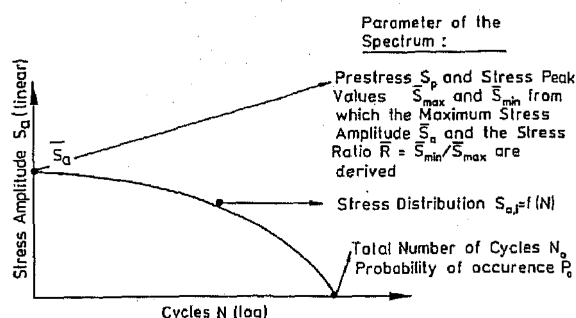


Fig. 8: Design Spectrum

The extensive research including the measurements and the evaluation of failures (damage, initial cracks, fractures) under proving ground tests, public road approval and customer usage was carried out to derive a representative design spectrum for wheels and hubs [2, 6, 8, 10].
On Fig. 9 the basic data necessary to derive the design spectra for **wheels and hubs** under the mission profile for West-European roads are presented. The operational stresses on wheels and hubs are mainly the result of the vertical wheel forces F_z and the lateral wheel forces F_y [6, 10]; longitudinal wheel forces F_x doesn't generate fatigue relevant stresses and should be considered only for the wheels with spokes.

	Loads		Spectrum Parameters			
Condition	F_z	F_y	N	N_β	N_e	Distribution
Straight Driving :	$F_{z,s}$	$\pm F_{y,s}$	$0{,}96 N_t$	$0{,}48 N_t$	$05 \cdot 10^{-6} N$	Linear
Cornering :	$F_{z,c,a}$	$F_{y,c,a}$	$0{,}04 N_t$	$0{,}02 N_t$	$12{,}5 \cdot 10^{-6} N$	Normal

Total Frequency $N_t = \frac{10^3}{2 \cdot r_{dyn} \cdot \pi} \cdot z \cdot L$

L = Design Life in km
r_{dyn} = dynamic tire radius in m
z = Number of cycles for a wheel revolution
N_β = Frequency of the basic stresses $S_{a,stat}$ under rated wheel load $F_{z,stat}$
N_e = Frequency of Maximum Values

Fig. 9: Basic Data for the Derivation of Stress Spectra on Wheels and Hubs

The decisive load cases for the fatigue life of wheels and hubs are **straight driving** and **cornering** [6]. The loads which determine the stress at **straight driving** $S_{a,s}$ are maximum vertical load $F_{z,s}$ and in-phase lateral loads, directed inside ($+F_{y,s}$) and outside ($-F_{y,s}$) the vehicle; in **cornering** of importance are only the maximum vertical ($F_{z,c,a}$) and lateral ($+F_{y,c,a}$) load acting on the wheel positioned on the bend outer side by which the maximum cornering stress $S_{a,c}$ is generated.
The operational stresses of wheels and hubs correspond to the oscillation of the basic stress amplitude $S_{a,stat}$ due to the wheel rolling under its static vertical load $F_{z,stat}$ (the rated wheel load) and by superimposed stresses generated through variable service loads acting in straight driving and cornering.
Figure 10 shows in generalized form the spectra derived by the data given in Fig. 9.

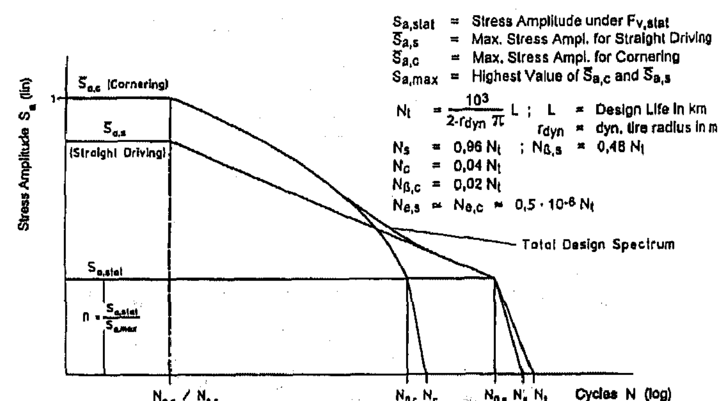


Fig. 10: Generalized Description of Wheel and Hub Design Spectra

The data on Figs. 9 and 10 cover the existing recommendations for the durability approval of wheels and hubs from SAE, JASU, TÜV [6, 8, 10], so far in using test facilities a service-like deformation in critical areas is achieved, which for rotating bending test facilities is seldom the case. Also the known requirements for durability approval on proving grounds or public roads by individual vehicle companies in most cases are covered too.

As the state-of-the-art for **non-rotating suspension components**, for the preliminary design evaluation and the durability approval by tests, the data in Fig. 11, based on existing data and experience [11, 12, 15 to 17], can be used. The decisive load cases for the fatigue life of these components are: **straight driving, cornering** and **braking / accelerating** is by experience of importance only in the case of driven axles. The decisive **straight driving** loading conditions to be taken into account are: simultaneously acting maximum vertical force ($F_{z,s}$) and lateral load

directed inside ($+F_{y,s}$) or outside ($-F_{y,s}$) of vehicle or longitudinal load acting toward vehicle ($F_{x,s}$). The maximum stress amplitude for straight driving $S_{a,s}$ is derived from the maximum and minimum stress, whereby $S_{min,s} \leq 0$.
For **cornering** the loading conditions on the wheel at the outer side of a bend (loads $F_{z,c,a}$ and $+F_{y,c,a}$) and on the inner side (loads $F_{z,c,i}$ and $-F_{y,c,i}$) must be taken into account.
For **braking and accelerating** only the loads acting at maneuvers in straight driving are decisive, so that only a simultaneous combination of vertical loads ($F_{z,b}$) and the longitudinal loads ($F_{x,b}$) are of importance.

The number of cycles for individual spectra N and for maximum values N_e are given as values related to the total number of cycles N_t. The total number of cycles vary depending on dynamic behavior, but for fatigue evaluation, as a mean value for cars 500 cycles per km can be used.

	Loads				Spectrum Parameters		
Condition	F_z	F_y	F_x	$F_{x,b}$	N	N_e	Distribution
Straight Driving :	$F_{z,s}$ $F_{z,s}$	$\pm F_{y,s}$ –	– $F_{x,s}$	–	$0,96 \cdot N_t$	$N \cdot 10^{-6}$	Linear
Cornering : outer side inner side	$F_{z,c,a}$ $F_{z,c,i}$	$F_{y,c,a}$ $F_{y,c,i}$	– –	– –	$\frac{0,04 \cdot N_t}{20}$	50	Normal
Braking :	$F_{z,b}$	–	–	$F_{x,b}$	$3 \cdot 10^5$	$3 \cdot 10^2$	Normal

with Total Frequency $N_t = L \cdot l$
N = Frequency of individual spectrum, N_e = Frequency of extreme values
L = Design Life
l = Number of cycles per km

Fig. 11: *Basic Data for the Derivation of Stress Spectra for Suspension Components of Cars*

It is still in validation how well this data covers the durability test requirements on proving grounds and the public roads used for durability approval by individual companies, but it covers the known laboratory test requirements of certain individual companies as well for the case of the multiaxial loads as for uni-axial load testing.

The graphical presentation corresponds to the form given in Fig. 10 for wheels and hubs, where in the periodic stress amplitude $S_{a,stat}$, generated by the rolling wheel with the frequency of occurrence N_β, is to be dismissed. The spectra in a log (life, cycles) - linear (stress or load) diagram between the cycles N_e for maximum values and the total number of cycles N are described by a linear distribution for straight driving and a normal distribution for cornering and braking/ accelerating maneuvers.

ALLOWABLE STRESSES - The allowable stresses depend on **operational loading** and on **material properties, manufacturing procedure** and **design details**. Therefore, for a reliable design evaluation before any durability approval is made, a good knowledge about individual influences is necessary. Due to the manufacturing procedure the hardness distribution, the grain size, the surface roughness, and the residual stresses can be decisively changed and thus the fatigue strength. It is known that through surface treatment an increase of the fatigue strength up to 300 per cent could be achieved, depending on material properties, stress distribution, type and level of stress-time history.
The design details can influence the stress concentration (notch factor K_t), the stress distribution in the components crack propagation section and the stress state (bending, tension, multiaxial stress state), which all have effect on the fatigue strength too. Generally, it should be regarded that the increase of the material ultimate (static) strength doesn't improve the fatigue strength proportionally and depends decisively on the design details as well as on the manufacturing process.

As an example the fatigue strength data for the cast aluminum wheels (Al Si7 Mg, T6) are summarized by Fig. 12. The fatigue strength for different wheel areas differ due to influence of prestress ($R = 0.2$ for rim and $R = -1$ for disc areas), stress distribution and the damage mechanism.

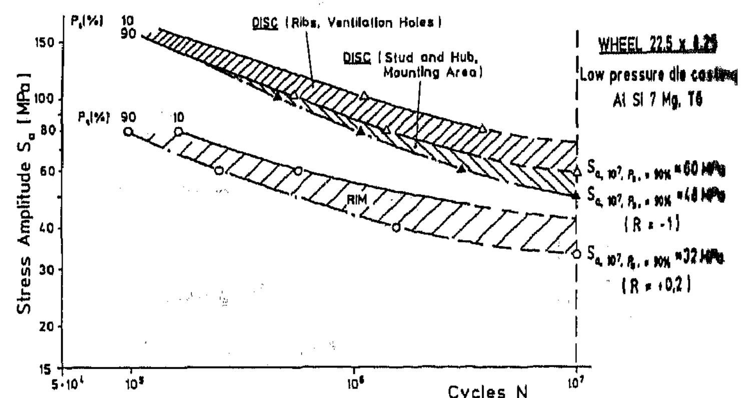


Fig. 12: *Fatigue Strength Data for Different Wheel Areas of Low-Pressure Die Casting Al Si7 Mg, T6*

More data and details concerning the fatigue strength of materials and suspension components including influences and procedures to improve the fatigue strength can be found in corresponding literature and data banks.

As a general comment it must be understood that a fatigue evaluation of a design, based merely on material fatigue properties determined under strain controlled cyclic loading, can't deliver reliable results, if additional influences on components fatigue strength are not taken into account.
It must be emphasized that a good design is of paramount importance. We have to keep in mind that due to the exponential relation between fatigue life and the stress

$$N_x = N_o \cdot \overline{S}_{a,x}^{-k} \text{ (see Fig. 13)}$$

a decrease of the stress by 10 per cent can increases the fatigue life roughly by 100 per cent. It means that a small change in design shape or dimensions can satisfy the requirements for an increase of the fatigue life. The relation between the fatigue life and the dimensions depend, on the stress distribution in the critical components section and is described by the equation:

$$t_o = t_x \, (N_o \, / \, N_x)^{\frac{1}{k \cdot n}},$$

where:
- N_o = required fatigue life
- N_x = achieved fatigue life
- t_o = required thickness to reach N_o
- t_x = existing thickness
- k = slope of the fatigue life curve (usually between $k = 4$ to $k = 7$)
- n = ratio of the stress change through thickness ($n = 2$ for pure bending and $n = 1$ for pure tension/compression stress).

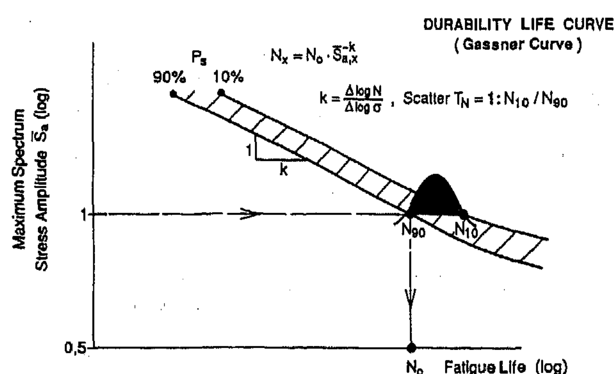


Fig. 13: Relation between Operational Stress and Fatigue Life

Finally it should be mentioned that for a state-of-the-art fatigue life scatter in the production of suspension components [2], depending on the production technology, a value $T_N = 1 : 3 \text{ to } 1 : 4$ (see Fig. 13) should be achieved. This data should be used as a basis to determine the requirements (risk factor R_F) for the durability approval of the suspension components.

DURABILITY APPROVAL

Because the existing procedures don't satisfy the requirements for a reliable fatigue life prediction the durability approval of vital suspension components generally is made experimentally by tests in laboratory as well as on roads (proving grounds). To guarantee a time and cost effective development the experimental tests on components should be made in an early development stage. To carry out such tests appropriate test facilities including load programs [8 to 12, 15 to 18] must be available. A general approach to the durability approval of rotating components (wheels, hubs) and non-rotating suspension components will be presented.

PROCEDURE FOR DURABILITY APPROVAL OF THE ROTATING SUSPENSION COMPONENTS (WHEELS, HUBS) - To assure a service-like durability approval of rotating suspension components the tests are carried out in a wheel/hub multiaxial load test facility. This test facility, which is increasingly used at vehicle and wheel producer companies, allows an effective and reliable durability approval, whereby a wheel with mounted tire runs inside a drum and simulated loads are generated by servohydraulic cylinders.

There exist basically three different variations of this facility, which all have in common the same basic loading principle, shown on Fig. 14a; more details can be found in the literature [16].
The time history of the loading program used in such a test facility, which takes into account the mission profile for the West-European loading and usage conditions and the generated strain in a point of the wheel disc are presented in Fig. 14b.

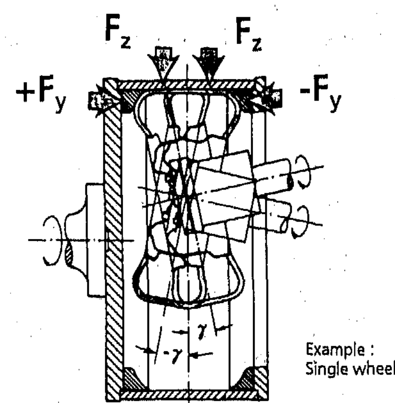
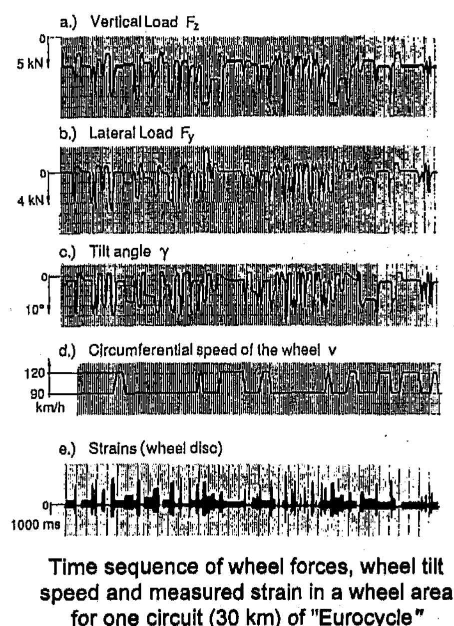


Fig. 14a: Loading Principle for the Testing of Rotating Suspension Components



Time sequence of wheel forces, wheel tilt speed and measured strain in a wheel area for one circuit (30 km) of "Eurocycle"

***Fig. 14b:** Standardized Load Program for Rotating Suspension Components*

In Fig. 15 the design spectrum derived by the data on Fig. 9 and the test spectrum for the generated stress in the wheel disc are presented. The requirement is that the fatigue damage induced by test program D_T has the same value as the damage of the design spectrum D_D ($D_D = D_S + D_C$, whereby D_S represents the damage at straight driving and D_C at cornering):

$$D_T = D_D.$$

This is realized by the standardized test spectrum, called "Eurocycle", whereby a test distance of 10,000 km is needed for aluminum wheels, taking into account that three tests are carried out for each wheel type and a corresponding safety factor R_F, as discussed in Chapter 1.

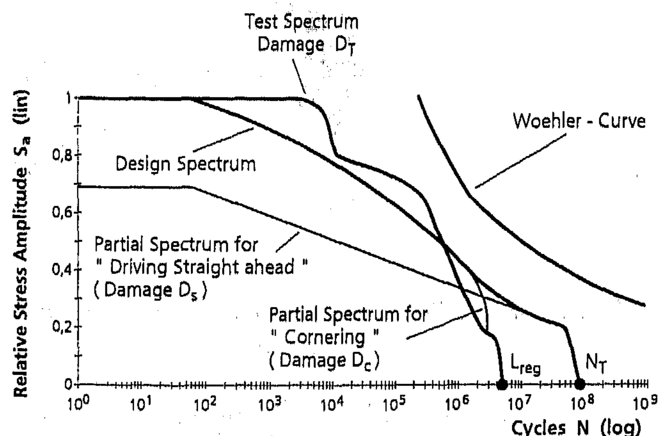


***Fig. 15:** Design (300 000 km) and Test (10 000 km) Spectrum for Light Metal Car Wheels*

As an example the results of a durability approval on a cast aluminium wheel 6 J x 14 are presented. The wheel has as a result of its design relatively high local stresses in the connection area between disc and rim on the cooling hole (spot A1, Fig. 16), whereby the nominal stress in that section is relatively low being primarily bending. At Area *B* the local surface stress on the inner side is about 30 % lower compared to spot A1 (Fig. 17), but the nominal stress level is higher because in this section primarily tensile or compressive stresses occur during wheel rolling. These stress conditions result in the following failure mode development (Fig. 17):

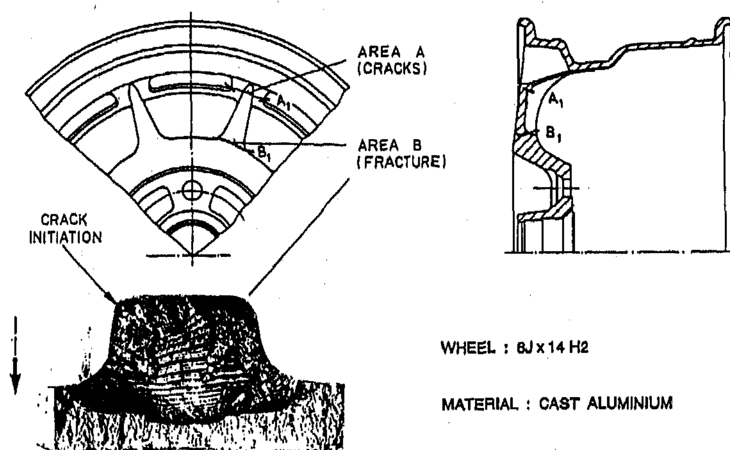


***Fig. 16:** Fracture under Durability Tests in the Biaxial Wheel Test Facility*

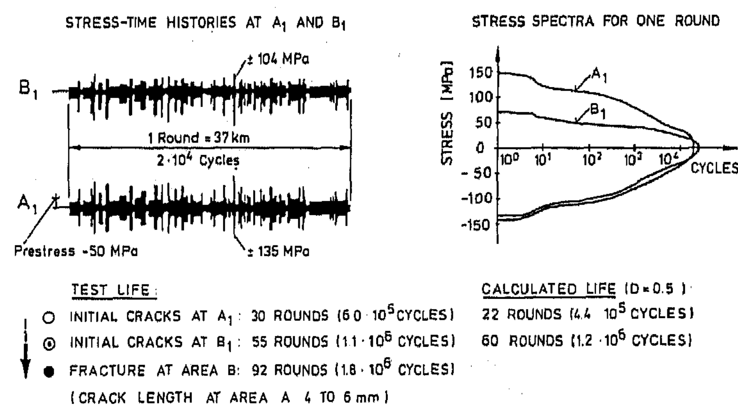
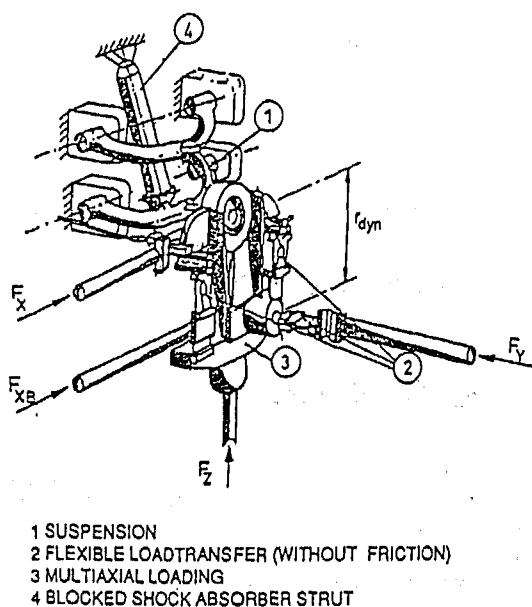


***Fig. 17:** Stress-Time Histories, their Spectra and Fatigue Life*

Initial cracks of approximately 1 mm are registered after about 1/3 for the total test life in the Area *A*. After 2/3 of the total test life the cracks occur also in Area *B*. The cracks propagate in this area quickly while in Area *A* very slowly.
Finally, the tests are finished because of the global deformation of the wheel due to very deep cracks in Area *B* (see Fig. 16), while the cracks in Area *A* are almost stopped after achieving a length of 4 to maximum 6 mm. So the improvement and redesign of the wheel is more important in Area *B* than in Area *A*, nevertheless, the local stress in this area is about 30 % higher.

PROCEDURE FOR DURABILITY APPROVAL OF THE NON-ROTATING SUSPENSION COMPONENTS - A schematic layout of the test facility for the durability approval of the non-rotating suspension components, from the spindle up to the shock absorber and a standardized load program are presented on Fig. 18a and Fig. 18b. The suspension components at this test are assembled to a "half axle". The four relevant loads, acting at the wheel, are generated by servohydraulic actuators. These loads are transferred through rods, joints and a wheel adapter to the wheel base (vertical, lateral and braking loads) or to the wheel centre (longitudinal loads generated by rough road and acceleration) and further transferred to the suspension components which have to be tested. The excitation of the servohydraulic actuators are made to be load controlled.



1 SUSPENSION
2 FLEXIBLE LOADTRANSFER (WITHOUT FRICTION)
3 MULTIAXIAL LOADING
4 BLOCKED SHOCK ABSORBER STRUT

Fig. 18a: *Layout of the Test Facility for Durability Approval of Suspesnion Components*

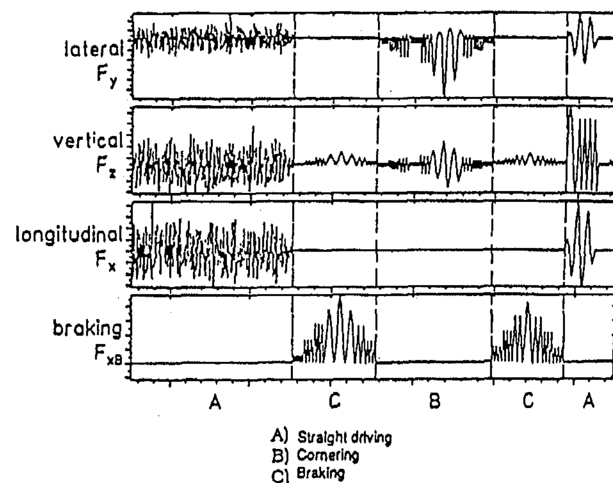


Fig. 18b: *Standardized Load Program for Durability Approval of Suspension Components*

The procedure for the design evaluation and durability approval will be presented on an example. On Fig. 19 a front suspension lower arm, manufactured by welding of steel sheet parts is shown. Based on vehicle data the loads for individual conditions - straight driving, cornering, braking - were determined and used as well for calculation of stresses in individual areas as for the adaptation of the drive signals (test program).

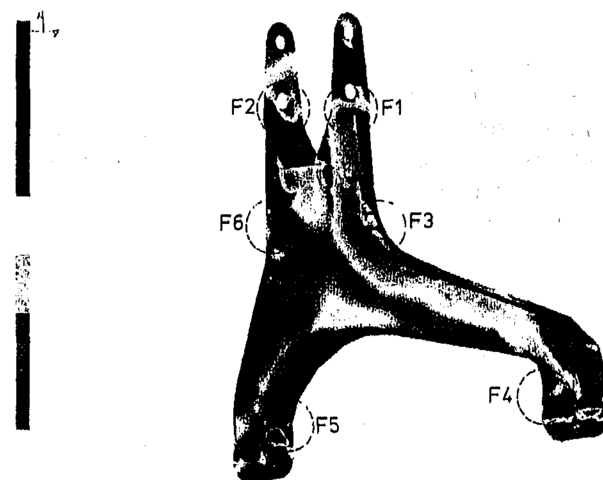


Fig. 19: *Front Suspension Lower Arm and Possibly Critical Area F1 to F6*

Based on preliminary theoretical design evaluation the prototypes were manufactured and experimentally investigated. In individual areas *F1 to F6* strain gages

were positioned and stresses analyzed under individual load case and compared with the stresses during durability tests.
An excerpt of the measured stresses in areas *F3*, *F5* and *F6*, superimposed to the pre-stresses, induced by static suspension load, are reproduced in Fig. 20. It can be seen from this recordings that the highest stresses in the areas *F3* and *F6* are generated by braking, being mainly bending stress at the sections *F3* to *F6*, while in the area *F5* the higher stress, being generally much lower than in areas *F3/F6*, are induced through straight driving.

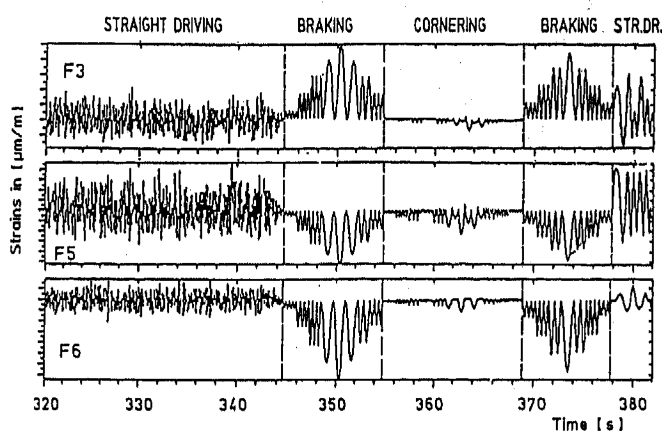


Fig. 20: *Measured Strains in Areas F3, F5 and F6 at Durability Test*

Based on the data in Fig. 11 the stress spectra for individual load cases are derived and compared with the test stress spectra; the results obtained for area *F3* and *F5* are shown on Fig. 21 and Fig. 22.

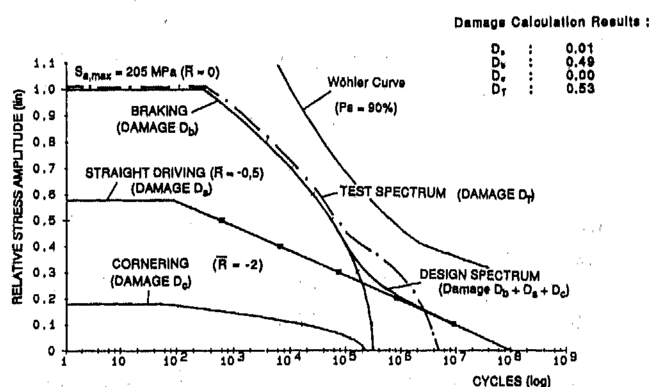


Fig. 21: *Comparison of the Design (300 000 km) and the Test Spectrum - Area F3*

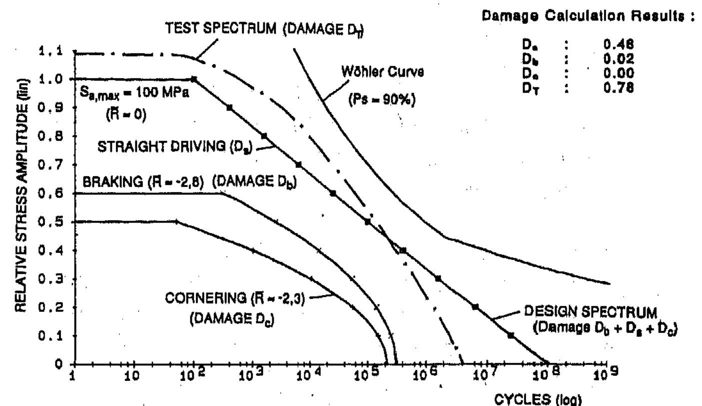


Fig. 22: *Comparison of the Design (300 000 km) and the Test Spectrum - Area F5*

For the design evaluation also the analysis of the fatigue damage induced by individual loading were carried out simultaneously. These analysis delivered the following results:

Area F 3:

The fatigue damage in this area is influenced almost only by braking (partial damage of braking spectrum $D_b \approx 0.49$, damage at straight driving $D_s = 0.01$ and cornering $D_c \approx 0.00$). This means that if a redesign should be necessary, it should be concentrated mainly to decrease the stresses due to the longitudinal forces.
To achieve the required fatigue life based on an allowable total damage $D \leq 0.5$ an endurance limit of $S_e \geq 0.42 \cdot S_{a,max}$ taking into account a slope $k = 4$ *to* 6 valid for welded steel sheet components. For the maximum stress amplitude $\overline{S}_{a,max} = S_{a,b} = 205$ *MPa* an endurance limit $S_e \geq 86$ *MPa* is needed in this area, what could be realized by good manufacturing for welded steel sheet components.
Based on the damage calculation the damage in the durability test corresponds quite well the damage for the design spectrum ($D_D \approx 0.5$ and $D_T = 0.53$). Similar results are obtained for areas *F2* and *F6.*

Area F 5:

The damage in this area is mainly influenced by straight driving $D_s = 0.48$; the damage due to the braking is about $D_b \approx 0.02$, while in cornering the value is very low $D_c \approx 0.00$. Based on an allowable total damage $D = D_s + D_b + D_c \leq 0.5$ an endurance limit (under same conditions as for area *F3*) of $S_e \geq 0.43$ $S_{a,max}$ should be realized;

for the value of $S_{a,max} = S_{a,s} = 100\ MPa$ an endurance limit of $S_e \geq 43\ MPa$ is necessary, which could be easily achieved.
The damage of the test spectrum in this area is $D_T \approx 0.78$ being higher than the damage of the design spectrum ($D_D \approx 0.5$), but due to low stress level without influence on the validation of durability life.

The result of durability tests show that the required fatigue strength in individual areas and corresponding durability satisfy the demand, whereby initial fatigue cracks appeared after achieving the required test life in areas F1, F3 and F6. On Fig. 23 an example of the typical failures during durability tests are shown. The cracks occurred in the areas F3 and F6, near the stabilizer fixture, and F1 at the weld toe. The achieved results verify that the test requirements based on load assumptions are satisfied.

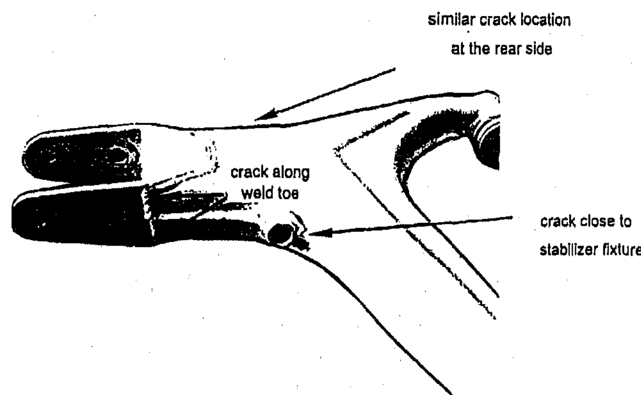


Fig. 23: Damage Locations at Durability Approval

To verify these results the measurements concerning expected operational loads (as shown on Fig. 6) and the stresses were carried out with the prototype vehicle; the stress measurements were carried out with the suspension arm used for stress analysis in the test facility.
The results of these measurements have shown that the assumed loads for individual conditions generally correspond to the measured loads with cornering and braking slightly lower than calculated. The achieved stresses in individual areas correspond also generally to the simulated stresses, whereby related to fatigue damage non-significant differences between measured and simulated stresses in individual areas were found.

CONCLUSION

For a reliable approval of vital suspension components the following conditions must be fulfilled:

1. **The representative loading, which include the maximum values, frequency of occurrence distribution and the total number (called load spectrum) for the specific mission profile (customer usage) must be defined.**
The load data should be a part of the specification to be used for the design evaluation and the durability approval. The load data must take into account all possible loading conditions, which maximum values during a customer usage would be seldom achieved; this is of decisive importance for the suspension safety components.

2. **The test programs for the durability approval should assure that the fatigue damage under operational loading is adequately verified.** To fulfill this requirement it is not recommended [13] to accelerate the durability approval by increasing the maximum test loads in relation to operational loads, but the acceleration is achieved by the omission of low intensity loading occurring predominantly at straight driving and by increasing the frequency of occurrence of the medium loads. Nevertheless, the total frequency of the test cycles should be not lower than $3 \cdot 10^6$ to $7 \cdot 10^6$ to simulate all fatigue damage phenomena, being of importance for suspension components.

3. **The design and durability approval should be a part of the simultaneous engineering process, in which the evaluation of material behavior and manufacturing procedures as well as design selection for prototypes to be proved, is carried out.**
For suspension safety components all influences on the strength and the fatigue life, especially for the light weight design, like material behavior under operational conditions, influence of the manufacturing process and components behavior under special event loading must be taken into account. It must be known that the most expensive failures, if a component doesn't meet the requirements, are made during design and approval stage

4. **The requirements for the quality assurance and control should be fixed based on results of design evaluation and durability tests.**
These requirements should guarantee a reliable service strength and durability.

Finally it must be emphasized that the future requirements for optimal light-weight design and improved reliability and safety will increase the demands on procedures being used to determine the service strength of suspension components. Therefore, the

further development is expected regarding standardized test procedures at which computer assisted design evaluation will be simultaneously carried out. This will allow an effective, time and cost saving control of the new designs and developments and yield also the data which may be important, should there be a question of product liability.

ACKNOWLEDGMENTS

This paper includes individual results of the project »Design and Dimensioning of Highly Stressed Welded Light Alloy Suspension Components«, supported by the European Community Project No. BE5249 (Contract No. BRE2-CT92-0181) and co-operating industrial partners AUDI-Ingolstadt, FERRARI-Maranello, and SPEEDLINE-Tabina di Caltana. We express our thanks to all of them.

REFERENCES

[1] Gassner, E.; Schütz, W.:
Evaluating Vital Vehicle Components by Program Fatigue Tests
9th FISITA Congress, London (1962), Proceedings, pp. 196-205

[2] Grubisic, V.:
Criteria and Methodology for Light-Weight Design of Vehicle Components Subjected to Random Loads
SAE Technical Paper 850367

[3] Buxbaum, O.:
Random Load Analysis as a Link between Operational Stress Measurement and Fatigue Life Assessment
ASTM STP 671 P.K.

[4] Donaldson, K.H.:
Field Data Classification and Analysis Techniques
SAE-IP-82/109 - No. 820 685

[5] Tucker, L.E.:
Cumulative Damage Analysis
SAE-IP-82/109 - No. 820 686

[6] Grubisic, V.; Fischer, G.:
Automotive Wheels, Method and Procedure for Optimal Design and Testing
SAE Technical Paper 830135,

[7] Gassner, E.:
Fatigue Resistance of Various Materials under Random Loading
In: Fatigue Life of Structures under Operational Loads; Proceedings of the 9th ICAF-Symposium, Ed.: O. Buxbaum and D. Schütz, Fraunhofer-Institut für Betriebsfestigkeit LBF, Darmstadt, Report No. TB-138 (1977), pp. 3.5/1-34

[8] Fischer, G.; Grubisic, V.:
Cast Aluminium Wheels for Trucks and Buses
- Testing and Evaluation -
SAE Technical Paper 841705

[9] Jaeckel, H.R.:
Design Validation Testing
SAE-IP-82/109, No. 820 690

[10] Grubisic, V.; Fischer, G.; Heinritz, M.:
Design Optimization of Forged Wheel Hubs for Commercial Vehicles
SAE Technical Paper 841706

[11] Neugebauer, J.R.; Grubisic, V.; Fischer, G.:
Procedure for Design Optimization and Durability Life Approval for Truck Axles and Axle Assemblies
SAE Technical Paper 892 535

[12] Müller, A.; Grubisic, V.; Fischer, G.:
Fatigue Evaluation of Passenger Vehicle Suspension Components using a Generalized Multiaxial Load Programme
In: Proceedings of the 4th Int. Conf., Firenze, March 16-18, 1994
Ed. by ATA, Orbassano and DMTI, Firenze; Vol 2 (1994), pp. 879-898

[13] Grubisic, V.:
Determination of the Load Spectra for Design and Testing
Journal of Vehicle Design 15 (1994), No. 1/2, pp.8-26

[14] Rupp, A.; Grubisic, V.; Neugebauer, J.:
Development of a Multi-Component Wheel Force Transducer - A Tool to Support Vehicle Design and Validation
SAE Technical Paper 930258

[15] Feitzelmayer, K.; Breitling, U.:
Development of Commercial Vehicle Driven Axles on Multi-Axis Test Set-Ups
Paper of the Conference "Prediction and Simulation of In-Service Conditions", London, 22/23 May 1995, I. Mech. E. (1989), C 113/85, pp. 131-138

[16] Biaxial Wheel/Hub Test Facility
Proceedings of the 1st Int. User Meeting, Sept. 30, 1993, Darmstadt
Ed. by V. Grubisic and G. Fischer, Fraunhofer-Institut für Betriebsfestigkeit (LBF), Darmstadt, TB-204 (1994)

[17] Bigliani, V.; Demichelis, G.; Jacoby, G.:
A Rig for Vehicle Suspension Durability Test
In: Proceedings of the 4th Int. Conf., Firenze, March 16-18, 1994
Ed. by ATA, Orbassano and DMTI, Firenze; Vol 2 (1994), pp. 857-877

[18] Fischer, G.; Grubisic, V.; Gu B.L.; Lu S.C.; Li M.:
Determination of Design and Test Spectra for Vehicle Components under Operational Conditions in China
Proceedings of XXV FISITA Congress, 1994, Beijing
Volume 2, pp. 57- 67

970096

Design and Operation of a New Vehicle Suspension Kinematics and Compliance Facility

Tony Best and Steve J. Neads
Anthony Best Dynamics

John P. Whitehead and Ian R. Willows
MIRA

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A Kinematics and Compliance (K&C) facility is a test rig which undertakes quasi-static testing of vehicles to measure the K&C characteristics of the suspension and steering systems.

This paper describes a new K&C test facility which is a four wheel station machine designed to accommodate automobiles and light trucks. It is electro-mechanically operated, fully automated and computer controlled, and will complete a full suite of measurements in hours rather than days.

The first installation, at MIRA England, was commissioned in January 1996. Since then it has been in almost continuous use and a variety of vehicles have been tested; typical results are presented.

A second machine is now installed at the facilities of Goodyear Tire & Rubber Company in Akron, Ohio, USA.

INTRODUCTION

Kinematics is defined as the study of motion without reference to mass or force. Compliance is the deflection resulting from the application of force (ie the inverse of stiffness). These two words are particularly apt in describing a machine which measures kinematic characteristics due to suspension and steering system geometries and compliances due to springs, anti-roll bars, elastometric bushes and component deformations.

The study of suspension kinematics and compliances is not new. However, measurement rigs have improved over the years, not only because of improved technology, but also because of the increasing importance placed by the vehicle manufacturers on the need for complete and high-quality measurements.

With the increased thrust to reduce time from concept to production of a new vehicle, more and more use is being made of computer analysis and prediction techniques. Initial design parameters may be based on experience but, increasingly, use is being made of measurements on previous models and on competitor "target" vehicles. Suspensions are refined with the aid of computer ride and handling models for which measured data is essential; also managers need to ensure that development work is carried out only on vehicles that are known to be built to specification. All this can be achieved only by making the relevant measurements.

The test facility described here is a machine that measures the wheel displacements, and the forces and moments occurring at the tyre to road interface, caused by bounce, roll and pitch of the vehicle body and also by the application of simulated cornering, braking and traction forces.

The essence of the machine is that it subjects the vehicle suspension and steering system to a variety of forces and displacements applied slowly, so as not to excite any inertia forces from accelerating masses or any damping from shock absorbers or elastomers, whilst recording all the relevant deflections and loads.

BACKGROUND

Various systems for measuring suspension kinematic and compliance characteristics have been designed and installed around the world, some in vehicle manufacturing companies, some in component manufacturers and universities. Many are in-house designed and built and therefore each is often unique to its own institution.

However, to the authors' knowledge, there are three main threads to the development of a more widely adopted solution.

The advent of the radial ply tyre in France in the late 1940's created a requirement for more precise control of suspension kinematics, and early work in this area was done by Citroen in close collaboration with Michelin. The need to introduce longitudinal compliance whilst retaining good wheel control saw Peugeot collaborate with Michelin during the next decade.

As the popularity of the radial tyre grew worldwide, Michelin extended its "Suspension Static Test Facilities" first into the UK and later into the USA. Traditionally these rigs have moved the body of the car against a fixed ground plane, with the wheels supported on air bearings to allow free changes of track, wheelbase and steer.

Probably the most well known machine is the Chevrolet VHF (Vehicle Handling Facility) designed and built by General Motors in the USA [1]. This became operational in the early 1970's. Subsequently a number of machines have been based on the design of this facility. In this design the body is held stationary and the wheel platforms are moved in the vertical direction. In the plane of the ground the wheel platforms are servo controlled to provide forces or moments if desired, or to provide zero forces and moments to allow unrestrained track, wheelbase and steer changes.

The third line of development with the title SPMD (Suspension Parameter Measuring Device) started in the 1960's at Cranfield Institute of Technology in England under the direction of Professor J R Ellis [2]. In 1983 the National Highway and Traffic Safety Administration (NHTSA) of the US Federal Department of Transport sponsored an SPMD which was installed at the NHTSA operation in East Liberty, Ohio [3]. The design was further developed for an SPMD which was purchased by Goodyear Tire and Rubber Company and installed at its site in Akron, Ohio [4]. In this version, the SPMD moves the vehicle body against a fixed ground plane and uses servo controlled wheel pads in the ground plane to provide forces and moments.

CHOOSING A FACILITY FOR MIRA

Initially MIRA considered the feasibility of producing a facility from an in-house design. This would have drawn inspiration from previous experience of MIRA staff and the then-current methods of K&C testing. However, it was realised that this would have involved a degree of "re-inventing the wheel", and consequently a lengthy and expensive development process.

A commercially available solution was sought, but none (with service support in Europe) was found in a fully developed and production ready state. The survey included visits to the SPMD facility of the Goodyear Tire & Rubber Company in Akron, Ohio. This facility impressed the MIRA engineers because:

- The SPMD design had been under development for 30 years.
- At the time of the decision, the SPMD facility had been in successful operation for the previous four years.
- The SPMD is highly automated and can produce a complete map of suspension characteristics within a day or two (vehicle attachment or non-routine testing may increase this time).
- The SPMD produces data at the end of each individual test sequence so that results may be viewed as the test programme progresses.
- A number of US vehicle manufacturers have found the Goodyear facility useful and competent.
- Although retaining high accuracy and ease of use, the measurement system is simple in concept.
- The SPMD concept is based on the use of commercially available actuators and transducers, and the software control and analysis routines are constructed around commercially available packages.

SUPPLIER OF THE FACILITY - The then-current SPMD was a two wheel station device that required the test vehicle to be measured one axle at a time. From the outset it was envisaged that the MIRA K&C rig would be a four wheel station machine enabling two-axle operation and, hence, a further reduction in test time from that available on the SPMD facility. It was also envisaged that the rig would be capable of accommodating all vehicles from the smallest passenger cars to the largest light trucks.

These changes, and the consequent changes in test control requirements, demanded a radical re-think of the design. Goodyear had also been thinking of upgrading to a new generation machine along similar lines. With this in mind, it was decided to seek a designer and builder for the machine. The essential criteria for selection were that the supplier should have experience in designing, developing and marketing test equipment, and experience in design and development of vehicle suspension systems. The company chosen was Anthony Best Dynamics Limited (ABD).

THE DESIGN PHILOSOPHY BEHIND THE NEW SUSPENSION PARAMETER MEASURING MACHINE (SPMM)

In the concept stage there were three guiding requirements:

1) The design should be cost effective, simple and compact.
2) In use, the machine should be efficient, requiring the minimum number of operators and the minimum of maintenance.
3) MIRA was insistent that the machine should test both axles simultaneously.

There was a fundamental question to be answered: should the machine move the vehicle body or move the road? There are two arguments which favour moving the vehicle body:

- To apply the correct tyre loads, with the centre of pressure in the correct position for a vehicle running on a level road, the relationship between the tyre and ground should be correctly simulated. The simulated road surface beneath all wheels should lie in one plane whatever the position of the body in roll or pitch. If the body is held stationary and the suspension is exercised by vertical actuators effectively moving the road, then each of the wheel pads has to tilt both about a roll axis and a pitch axis and be controlled to lie in a plane. Conversely, if the body is moved, then the simulated road surface beneath each wheel does not have to be moved in the vertical direction, and all the surfaces will remain in one plane.

- Tyre characteristics such as camber are quoted relative to the ground plane. Therefore camber and steer angles should be quoted relative to the ground plane. This is more easily achieved by keeping the ground plane and the measuring transducers stationary.

Having decided to move the body against a stationary ground plane the question remained: how to do it?

This test philosophy implies that the machine must be able to move the vehicle body in pitch. The SPMD, which only moves the body in roll and bounce, uses a fixed pivot about which the roll and bounce frame is rotated. Extending this concept to add a pitch pivot proved very unattractive. To design any gimballed arrangement that would have sufficient stiffness to restrain the body satisfactorily, was difficult, and mechanically the concept seemed complex.

A number of alternative solutions were considered. The most mechanically attractive was to eschew the idea of fixed pivots and allow the table the full six degrees of freedom whilst ensuring that the table was properly constrained by six separate actuators. This meant that even for a pure vertical bounce movement all six actuators would have to be moved and the precise deflections for each actuator would have to be computed. However, today, appropriate computing power is available.

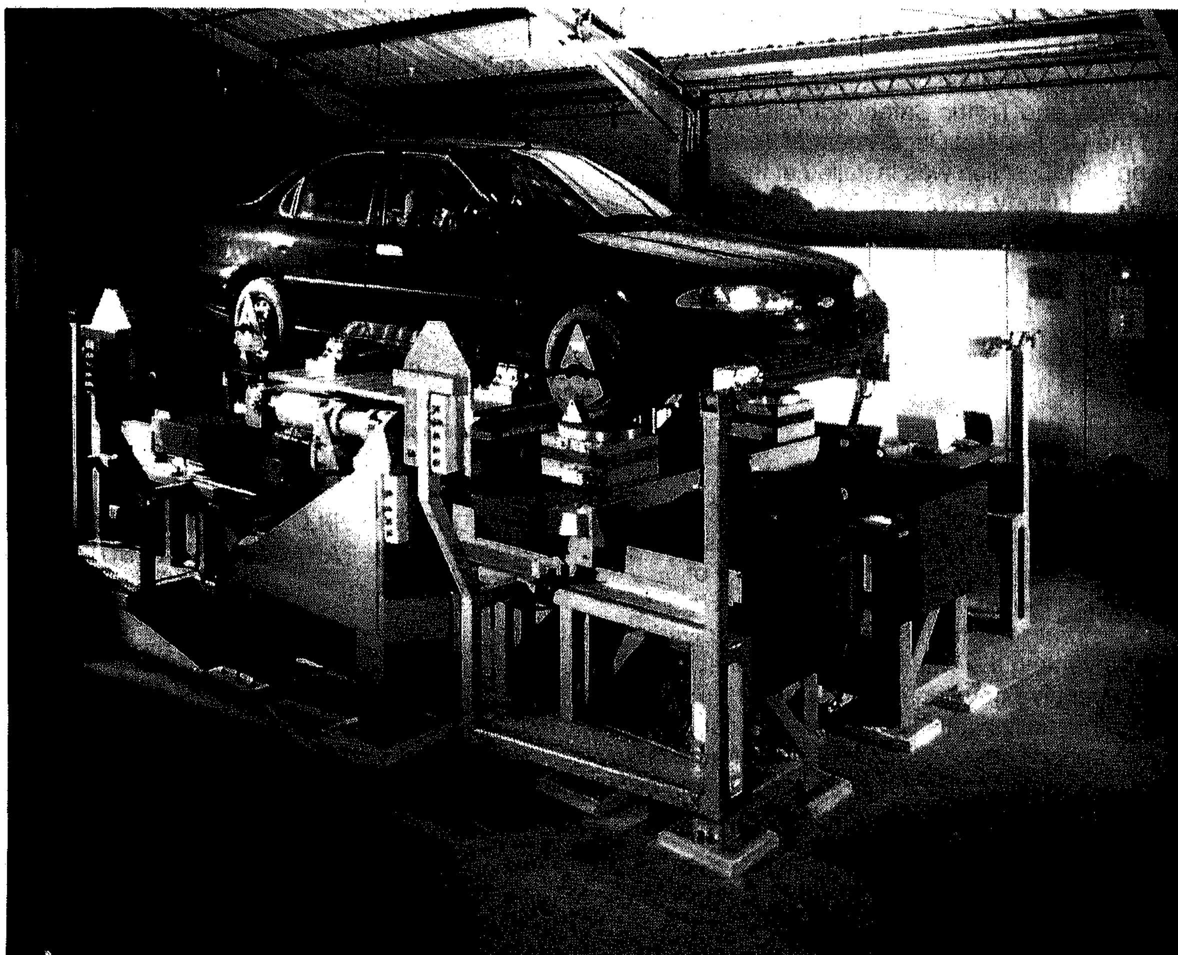


Figure 1 - SUSPENSION PARAMETER MEASURING MACHINE (SPMM)

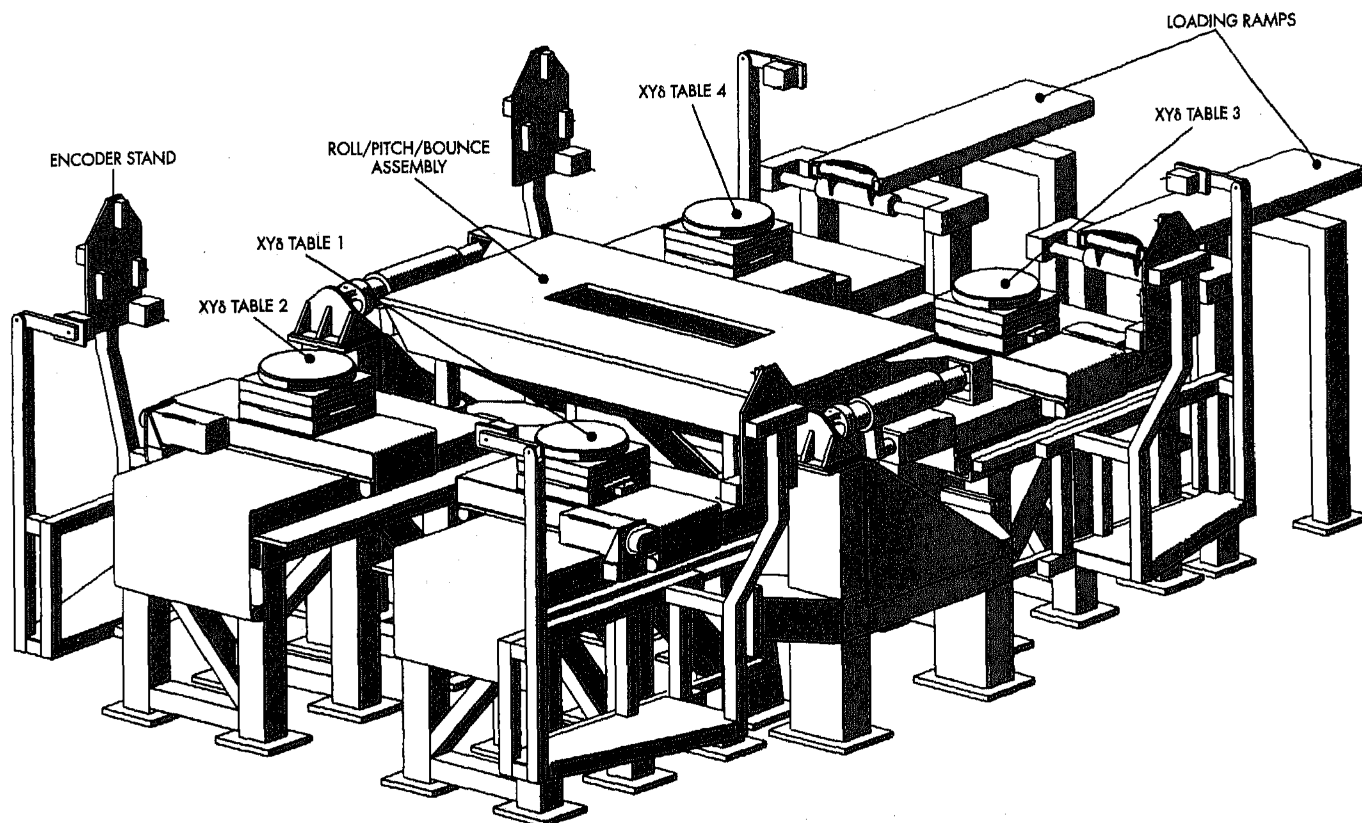


Figure 2 - SPMM GENERAL ARRANGEMENT

DESCRIPTION OF THE SPMM

OVERVIEW - The main structure of the machine consists of three different sub-assemblies: the Roll/Pitch/Bounce assembly, the XYδ tables and the loading ramps. Each XYδ table is bolted to the roll/pitch/bounce table, two in front and two behind. The loading ramps are attached behind the two rear XYδ tables. The arrangement is shown in Figures 1 and 2.

ROLL/PITCH/BOUNCE ASSEMBLY - The heart of the machine is the roll/pitch/bounce assembly (Figure 3) which controls the motion of the body of the vehicle under test. It comprises two rigid steel frames, the base frame (which is bolted to the floor and to the XYδ tables), and the moving frame, which is connected to the base frame by six "Rolaram®" linear actuators. The Rolarams® are connected to the base frame by special high rigidity universal joints, which allow them to pivot in any direction, but prevent rotation about their axes, and are connected to the moving frame via spherical joints. Three of the Rolarams® are oriented so that they act principally in the vertical direction, two so they act principally in the fore and aft direction, and one laterally. Accurate motion of the assembly is achieved by synchronised movement of all six actuators.

The Rolaram® actuators are commercially available units, driven by brushless DC brake-motors, which convert rotary motion to linear motion by means of a roller-screw assembly. This is intrinsically more accurate and rigid than re-circulating ball screws.

XYδ TABLES - Each of the four XYδ tables has a turntable upon which the test vehicle's wheel rests (see Figure 4). The turntable assemblies can be driven in the fore and aft (X) and lateral (Y) directions, and can be driven about their central vertical axis (δ). These motions enable horizontal forces or aligning torques (steer torques) to be applied to the wheel. In addition, the X and Y motions provide rapid computer-controlled adjustment to accommodate different vehicle wheelbases and tracks.

The X and Y axes are driven by brushless - DC motors via recirculating ball screws, and are mounted on pre-loaded rolling linear bearings, for maximum rigidity and minimum friction. The δ axis is driven, via a clutch, by a "pancake" DC motor and a Harmonic Drive gearbox, which provides a high reduction ratio with high stiffness and zero backlash.

Force measurement - Integral to each XYδ table is the transducer system for measuring the forces acting on the wheel. This eight channel piezo-electric system, manufactured by Kistler, enables the three orthogonal forces and moments about the pad centre to be determined. From these outputs, the location of the line of action of the resultant force, and the pure steer torque on the wheel are also determined.

The piezo-electric transducer system was chosen in favour of a strain gauge system for its high accuracy and low inherent crosstalk characteristics, its stability and proven reliability, its ability to provide both high loads and high resolution, and for its compact

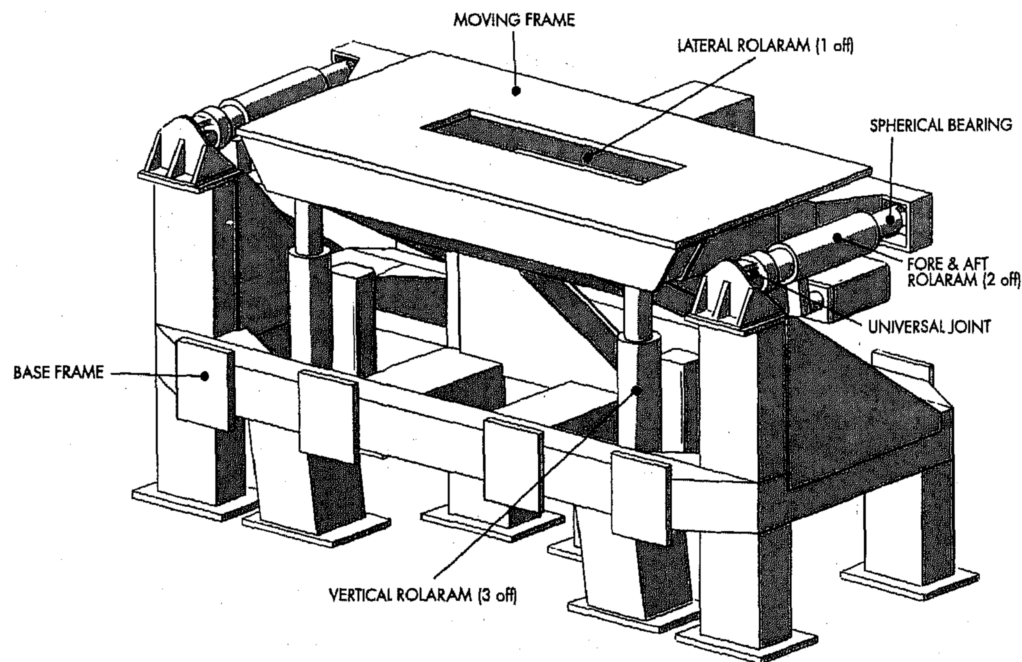


Figure 3 - THE ROLL/PITCH/BOUNCE ASSEMBLY

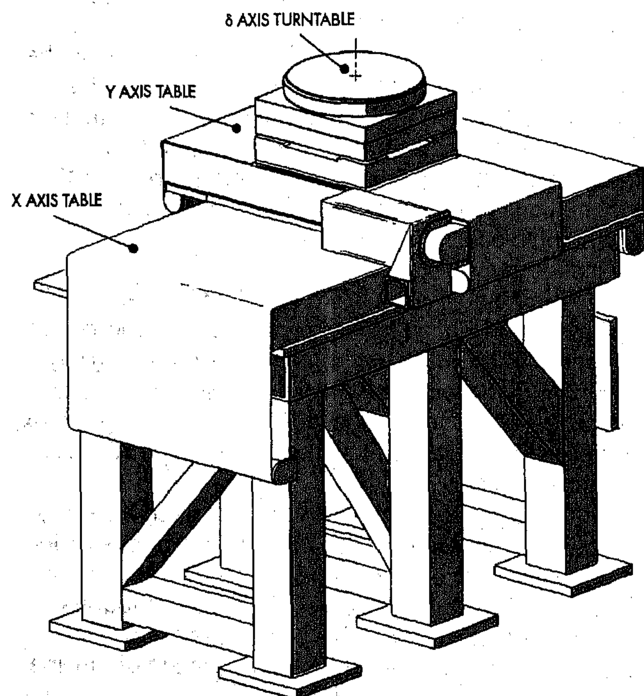


Figure 4 - THE XYδ TABLE ASSEMBLY

design. These advantages far outweigh the well known DC drift difficulties. The drift in a typical test period was predicted to be insignificant and this has now been proved in use

Wheel Position Measurement - The deflections of the wheels are measured by an array of optical "string" encoders which are mounted on a stand beside each XYδ table (see Figure 2). The encoder stands can be moved in the fore and aft direction on linear bearings to allow adjustment for different wheelbases.

Each encoder consists of a fine stainless steel cable, which is wound around a spring-loaded precision spool. As the cable is extended, the spool rotates an optical encoder which provides a digital position signal. The cables are hooked-up to a plate which is attached to the vehicle wheel in a known position. The plate is free to rotate about the axle line, and is weighted to hang vertically like a pendulum. As the wheel moves, each encoder of the array also moves. The encoder outputs are processed by the computer to determine the X, Y, Z, steer and camber deflections of the wheel. The castor change of the suspension is measured by a separate inclinometer which must be attached to the wheel carrier.

The wheel deflection measurement system has proved to be accurate and reliable. The system is easy to fit to the vehicle and is easy to understand and troubleshoot.

LOADING RAMPS - The ground plane of the rig stands approximately 1.5 metres above the floor, allowing easy inspection of the vehicle whilst under test. A standard garage lift is used to raise the vehicle to enable it to be loaded onto the rig. The two ramp

assemblies, attached to the rear XYδ tables, bridge the gap between the machine and the lift. The ramps are telescopic, and are pulled out and attached to the rear turntable assembly for loading and unloading the vehicle. This process is a computer co-ordinated automatic sequence, which enables vehicles to be loaded and ready for testing in less than one hour, and unloaded in about half an hour. During testing, the ramps are disconnected from the turntable assemblies and retracted (the positions shown in Figure 2)

STEERING WHEEL MOTOR - The characteristics of a vehicle's steering system can be investigated by testing with a drive-motor assembly attached to the steering wheel. This assembly uses a DC pancake motor and harmonic drive gearbox to provide motion, and has an integral transducer to measure the steering torques. The unit is designed to allow for eccentric steering column motion without distortion of the torque output, and reacts the torque onto the doors of the test vehicle.

CONTROL SYSTEM - A sophisticated machine-tool / robotic controller system (PMAC®) is used to synchronise the motions of the rig. The PMAC® system is designed to be interfaced with PC's, and consists of a number of plug-in boards which are installed within the PC. These boards control the feedback system for the SPMM's motors and capture all data from the measurement transducers.

The SPMM has 19 servo-controlled axes of motion, enabling it to subject a vehicle to a wide range of motions and forces. Up to 128 channels of data-capture measure simultaneously the forces on, and deflections of, the wheels, the body and the suspension.

SOFTWARE SYSTEM - The kernel of the SPMM software is its database, written in MAGIC®, which provides the principal user interface. The database stores all the set-up information for each vehicle tested, as well as all instrumentation data. It manages the processes of loading and unloading vehicles, performing tests and carrying out calibration routines, and monitors the status of the rig at all times. The database also catalogues the tests which have been performed for each vehicle, and interfaces with the post-processing software.

An individual test is defined by a single screen of numeric data, which specifies the motions and force limits for each axis of the machine. Tests are built up into sequences of one or more tests, which are run consecutively. A sequence of tests can be run fully automatically, or set to pause between tests to enable the results to be examined before continuing with the next test. Utilities provided in the system allow new sequences to be built up by copying from standard "templates", which define a particular type of test, or by copying sequences of tests from previous vehicles. The database will allow up to 26 sequences of tests, each containing up to 99 individual tests, to be defined for each vehicle, and will allow each of these to be run up to 99 times!

The standard post-processing supplied with the machine uses the commercially available plotting and mathematics package DIA®. This is used in two modes, interactive and batch. The interactive mode is used for examining and plotting test results on the computer screen, and can be used immediately a test is complete. It allows any measured channel, as well as a large number of derived data channels, to be plotted against any other channel, by means of an easy-to-use menu system. Multiple plots can also be overlaid. The batch mode can be used when a sequence of tests has been completed, to automatically perform standard post-processing and plotting for each test in the sequence.

Test results are stored in a standard binary format and can be transferred to other post-processing packages if required.

TESTING STRATEGY - In the current testing strategy, all tests are cyclic, with the cycle time being specified by the user (typically between 45 and 120 seconds per cycle).

Each axis of the rig can move independently from each other axis, and each axis has both displacement and force limits defined for it in the test definition. All axes can be set to "displacement" control, and all axes, except body roll, can also be set to "force" control. Axes under displacement control are cycled at a constant velocity between the displacement limits specified, with their motions synchronised. Motions can be either in-phase or in anti-phase. Axes under force control are programmed to maintain a constant force (or torque) level throughout the test cycle.

The system allows up to five consecutive cycles to be performed during a single test, and captures 256 points of data for each channel, each cycle. This produces a results file of approximately 1 Mb. The results for each cycle can be plotted independently, or all the cycles averaged. However, the normal testing method is to perform two cycles for each test and to plot the results of the second cycle (the first cycle being a "break - in" cycle). This gives data files of about 400 kb in size.

The SPMM has been designed to allow testing to be carried out with steel wheel substitutes fitted, although, for most tests, it is considered preferable to test with the normal road wheels and tyres fitted. The turntables of the wheelstations have high-friction surfaces which enable nearly 1 'g' to be generated in horizontal compliance tests. However, in these circumstances, there is often some tyre slip which results in there being residual horizontal forces at the end of the test cycle. The SPMM overcomes this by automatically performing a "lift cycle" to reset the vehicle between each test. In this cycle, the vehicle is lifted clear of the XY turnables, which are then moved to the correct positions to allow the vehicle to be lowered back to its datum ride height. As the vehicle is lowered, the XY turnables are servo-controlled to maintain zero horizontal force and to ensure that the tyres are correctly centred on the tables.

During steady-state cornering, an unrestrained vehicle rolls about its roll axis, maintaining the total vertical load at each "axle" constant. To replicate this behaviour during a roll test, and to obtain correct roll stiffness and lateral weight transfer information, the SPMM moves the vehicle simultaneously in roll, pitch and heave.

RANGE OF MEASUREMENTS

The SPMM offers a wide range of standardised and customised measurement possibilities.

The standard range of measurements includes the following:

- Roadwheel angular and rectilinear movements due to suspension system kinematics and compliances.
- Wheel vertical rates, tyre radial rates, suspension roll stiffness distribution and suspension horizontal compliances.
- Roll centre heights and anti-dive and anti-lift tangents.
- Steering ratio, steering system stiffness, Ackermann curves, kingpin offset, scrub radius and mechanical trail.

More detailed information is given in Appendix A.

SOME EXPERIENCES OF USING THE SPMM

At the time of writing, MIRA had been operating the SPMM for seven months, during which nearly fifty vehicles had their kinematic and compliance characteristics measured. Having gained a reasonable amount of experience in using the SPMM the following notes have been formulated.

PURPOSE OF MEASUREMENT - The SPMM provides the engineer with an important data set relating to vehicle ride and handling performance. The purposes of obtaining this data set can be broadly classified as follows:

Benchmarking - quantification of a competitor/class leading vehicle's kinematic and compliance characteristics in order to obtain a better understanding of its performance, or measurement of a prototype/production vehicle to verify its characteristics against the design intent.

Modelling support - verification of suspension system models against vehicle characteristics measured on the SPMM and modelling of whole vehicle dynamic behaviour using suspension derivatives from actual vehicle measurements.

Vehicle development - correlation of track based objective and subjective data with vehicle kinematic and compliance characteristics gives the development engineer a greater understanding of the influence of suspension system behaviour on the whole vehicle ride and handling performance. In particular the sensitivity of whole vehicle characteristics to variations in different suspension parameters can be explored, thus defining both the "ideal" characteristics and a tolerance bandwidth.

TEST CONFIGURATIONS - An initial examination of the test configurations listed in Appendix A may give the impression that a typical programme of kinematic and compliance measurements is reasonably short and concise. However, the SPMM is extremely flexible in allowing the user to drive any of its axes either singly or in parallel with each other whilst imposing displacement or force offsets on one or more axes. Hence, for example, a series of lateral force tests could be performed with the vehicle at different fixed roll angles and with the force input at a series of different trail values. Add to this the possibility of varying the force inputs in terms of amplitude and phasing in an attempt to simulate a "real" driving condition, and the test programme can quickly grow to a size where the volumes of data produced are large, unwieldy and of limited utility. The real value of the SPMM is that it allows the engineer to separate and identify the different suspension system kinematic and compliance parameters, and the test configurations should be structured to reflect this.

TURNAROUND TIMES - An original requirement specified by MIRA was a total turnaround time from the commencement of loading, through testing, to completion of unloading of not more than two days. This allowed for a test programme consisting of eight to ten tests assembled from those listed in Appendix A. In practice the SPMM has exceeded this requirement by a large margin with test programmes of around twenty five tests routinely completed within six to eight hours. One engineer with occasional technician help during the loading and unloading cycles can achieve this throughput.

CORRELATION - During and since commissioning of the K&C rig, correlation exercises have been successfully completed against the facilities of a number of vehicle mnanufacturers.

TEST RESULTS

As Appendix A indicates, a suite of tests on one car yields a large quantity of results, and it is beyond the scope of this paper to present more than a sample of the results available.

The authors have concentrated on presenting a small number of results which demonstrate interesting phenomena or trends.

Table 1 lists a selection of current European market sedans, giving vehicle type, drive configuration and principal dimensions.

Table 2 provides a comparison of some suspension stiffness and hysteresis characteristics associated with these vehicles, of all classes, and highlights some interesting phenomena.

Given that all the vehicles possess relatively "normal" weight distributions, there are fairly significant

differences in the ratios of front to rear vertical stiffness and, hence, in suspension frequencies.

Excessive wheelrate hysteresis is often detrimental to smooth road ride performance. In MIRA's experience, a target of 200 N or less is beneficial in this respect. It was surprising therefore to discover that many of the vehicles exhibited hysteresis values significantly larger than this. In general, those vehicles with "simple" suspension systems possessed the lowest values. One of the MacPherson strut derived systems displayed the lowest value for the front suspensions, although others of this type were considerably inferior. This indicates, perhaps, that spring offsets are incorrect or compromised by working space or roadwheel interference considerations. At the rear a twistbeam-derived suspension has the least hysteresis of the vehicles shown. The performance of the more "sophisticated" multilink types vary between 130N and 290N.

There is a large variation in rear suspension toe stiffness with respect to the lateral force, probably reflecting the decision of the designer either to hold the wheel rigidly or to employ some form of passive steer control.

Roadwheel longitudinal compliance is a significant contributor to impact isolation. Some of the vehicles in the "luxury" class possess an indifferent performance in comparison to the "cheaper" cars in this respect. Indeed the lowest stiffness/highest compliance is exhibited by a MacPherson strut type suspension system fitted to Car 3, a medium saloon.

From the roll test, one of the results produced is of the "Static Roll Weight Transfer Coefficient" or SRWTC. The SRWTC is calculated as:

$$SRWTC = \frac{2.dZf}{Zf} - \frac{2.dZr}{Zr}$$

Where Zf = front "axle" weight

Zr = rear "axle" weight

dZr = weight transfer across front "axle"

dZr = weight transfer across rear "axle"

The SRWTC is plotted against roll angle, and the graph produced can be interesting from two points of view. Firstly the gradient of the characteristic gives an immediate visual indication of which way the suspension roll stiffness is biased compared to the static weight distribution of the vehicle.

For example, a positive gradient indicates that the front "axle" reacts proportionately more lateral weight transfer during roll than it bears in static load; this is normally regarded as an understeering trend. A negative gradient would normally indicate an oversteering trend. Rear wheel drive cars tend to produce SRWTC graphs with positive gradients, whereas front wheel drive cars tend to produce SRWTC graphs with small negative gradients. In practice, the roll stiffness distribution may vary with roll angle and, again, the SRWTC graph gives a visual indication of this by showing a gradient which changes with roll angle.

The roll stiffness distribution can be calculated from the SRWTC gradient and from knowledge of the total roll stiffness of the vehicle (also measured during the roll test). This may be done to obtain an overall figure using average gradient and stiffness values, or to obtain a point figure using gradient and stiffness values at a particular roll angle.

The roll stiffness distribution may be calculated as follows:

$$\frac{Sf}{S} = \frac{t.Zf.Zr}{2Z} \times \frac{dSRWTC/d\theta}{dM/d\theta} + \frac{Zf}{Z}$$

Where Sf = front suspension roll stiffness

S = vehicle total roll stiffness

t = wheel track

Z = vehicle total weight

θ = roll angle

dSRWTC/dθ = gradient of SRWTC versus roll angle graph

dM/dθ (=S) = gradient of rolling moment versus roll angle graph

The second aspect for which SRWTC plotted against roll angle provides a very sensitive indication is an effect of friction in the suspension system. The introduction of friction turns a line graph into a hysteresis loop. Large disparity between the amounts of friction in the front and rear suspensions, combined with a low SRWTC gradient, tends to produce large hysteresis. The effect of this is to make the roll stiffness distribution indeterminate during transient handling manoeuvres involving roll angles within the hysteresis loop. For such vehicles in such manoeuvres, the understeer/oversteer effects of roll stiffness distribution may be unpredictable.

Table 1. Vehicles Subjected to Kinematic Compliance Test

	Vehicle Type	Drive Configuration	Wheelbase, m	Track, m Front	Track, m Rear	Test Weight, kg Front	Test Weight, kg Rear
Car 1	*Small*	*Front*	2.45	1.43	1.37	675	419
Car 2	*Small*	*Front*	2.58	1.46	1.42	720	434
Car 3	*Medium*	*Front*	2.61	1.48	1.48	783	593
Car 4	*Medium*	*Front*	2.57	1.46	1.44	674	493
Car 5	*Medium*	*Front*	2.58	1.45	1.42	761	483
Car 6	*Medium*	*Front*	2.54	1.45	1.45	738	481
Car 7	*Medium*	*Front*	2.7	1.5	1.47	943	561
Car 8	*Medium*	*Front*	2.64	1.45	1.45	827	570
Car 9	*Sport Saloon*	*Front*	2.55	1.47	1.46	835	608
Car 10	*Sport Saloon*	*Rear*	2.82	1.51	1.5	872	831
Car 11	*Sports*	*Rear*	2.58	1.5	1.49	885	791
Car 12	*Large*	*Rear*	2.99	1.5	1.49	926	832
Car 13	*Large*	*Rear*	2.93	1.54	1.54	995	932
Car 14	*Large*	*Rear*	2.83	1.54	1.52	825	740

Table 2. Kinematic and Compliance Characterists Summary

		W/C Vertical Stiffness N/mm	Hysteresis N	Roll Stiffness kNm/deg	SRWTC Gradient	SRWTC Hysteresis	W/C Lateral Stiffness kN/mm	Camber Stiffness kN/deg	Toe Stiffness kN/deg	W/C Longitudinal Stiffness kN/mm
Car 1	*Front*	18	190	0.56	-0.039	0.014	4	4	55	0.15
	Rear	16	90	0.43			5	2.5	33	0.77
Car 2	*Front*	21	155	0.83	-0.14	0.008	40	4	110	0.25
	Rear	21	80	0.95			2.5	2.5	9	0.35
Car 3	*Front*	21	126	0.63	0.03	0.003	20	5	60	0.07
	Rear	20	130	0.37			14	1.6	30	0.25
Car 4	*Front*	25	200	0.56	-0.03	0.006	5	3.3	33	0.3
	Rear	20	190	0.52			6.3	3.3	50	0.5
Car 5	*Front*	25	260	0.56	-0.04	0.019	10	4	142	0.38
	Rear	20	200	0.52			2	2.8	7.7	12
Car 6	*Front*	18	250	0.63	-0.03	0.007	25	3.8	15	0.47
	Rear	20	170	0.5			2.2	4.2	10	0.9
Car 7	*Front*	21	180	0.94	-0.08	0.025	50	4	45	0.2
	Rear	20	250	0.88			1.2	2.5	71	0.4
Car 8	*Front*	24	275	0.81	-0.012	0.018	440	4	20	0.52
	Rear	19	260	0.6			1.5	3	227	0.23
Car 9	*Front*	19	367	0.71	-0.077	0.004	17	4	45	0.27
	Rear	23	290	0.83			2	3.5	12.5	0.38
Car 10	*Front*	26	180	0.82	0.042	0.002	3.4	5	N/A	0.17
	Rear	27	280	0.53			4	4.5	15	0.35
Car 11	*Front*	16	300	0.78	0.005	0.027	83	4.2	14	0.31
	Rear	32	200	0.67			55	5.3	50	0.4
Car 12	*Front*	15	240	0.65	0.013	0.038	12.5	N/A	8	0.2
	Rear	27	120	0.5			33	N/A	50	0.3
Car 13	*Front*	24	360	0.5	0.03	0.013	8	5	33	0.16
	Rear	29	290	0.54			5	5	20	0.33
Car 14	*Front*	21	420	0.96	0.045	0.026	53	4	14	0.18
	Rear	25	290	0.56			25	3	11	0.58

Notes: 1. W/C = Wheel Centre
2. SRWTC = Static Roll Weight Transfer Coefficient (see text)

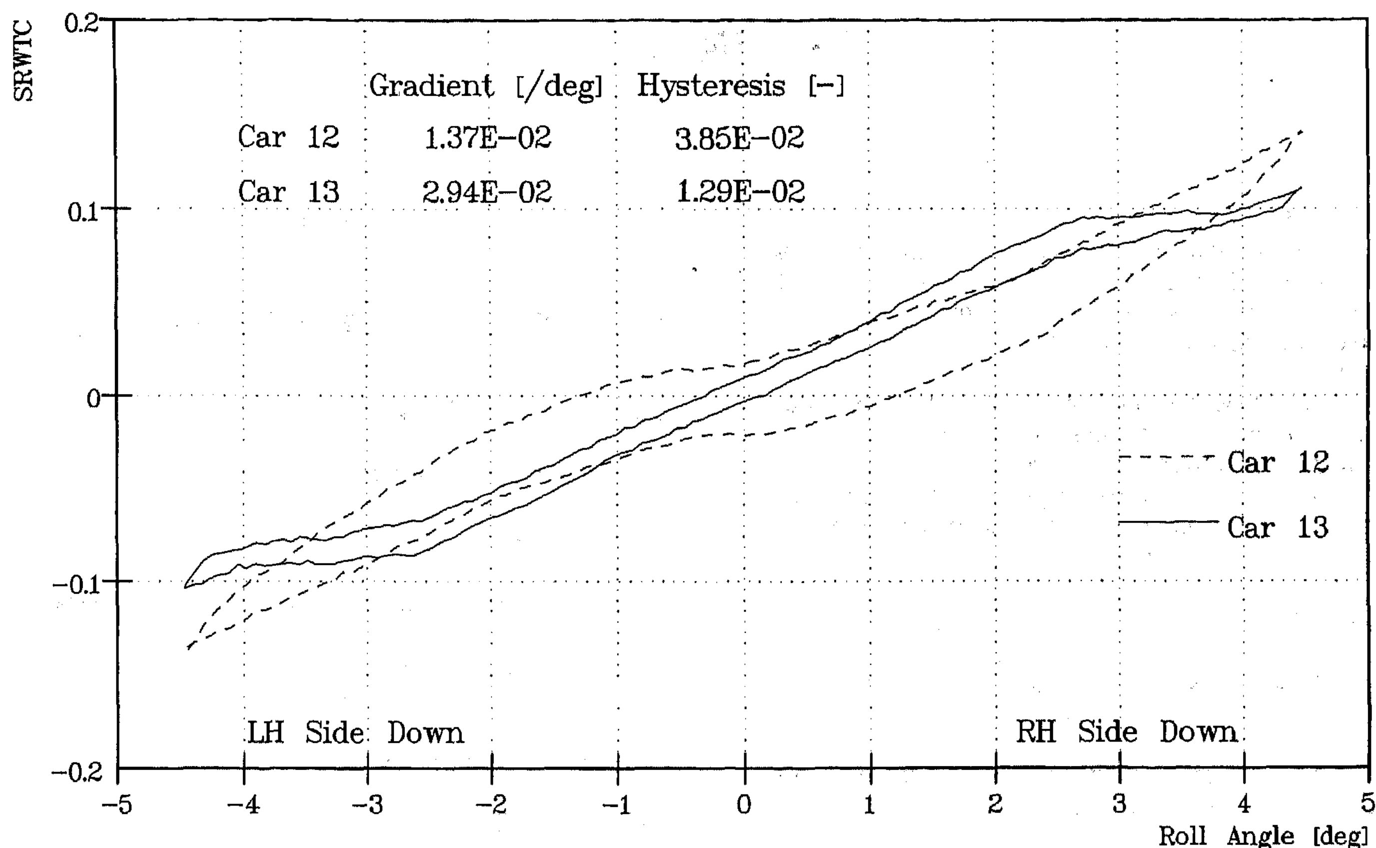


Figure 5 - STATIC ROLL WEIGHT TRANSFER CO-EFFICIENT v ROLL ANGLE

Figure 5 presents SRWTC plotted against roll angle for Cars 12 and 13 (reference Tables 1 and 2). For Car 12, the SRWTC hysteresis is high at 0.038. The ratio, front to rear, of suspension hysteresis is 2:1, and the SRWTC gradient over the central, low roll angle, region is low at 0.013. By contrast, the SRWTC hysteresis for Car 13 is only one third that of Car 12. Although Car 13 has greater suspension hysteresis, the ratio front to rear is 1.2:1 and the SRWTC gradient over the central region is much higher, at 0.03. Figure 5 also shows characteristics of distinctly different shapes for these two cars. For Car 12 the gradient increases with increasing roll angle, indicating an increasing bias of the roll stiffness distribution towards the front with increasing roll angle. For Car 13, the SRWTC curve shows the opposite trend.

CONCLUSIONS

In seeking to enhance its suspension design and development services, MIRA saw the need for an efficient, state-of-the-art kinematics and compliance measurement facility. After surveying the market, the decision was taken to order a new machine. Although it would have some similarities with earlier machines it should represent a significant advance.

MIRA collaborated with Anthony Best Dynamics (ABD), on the specification of the new machine. The Suspension Parameter Measurement Machine (SPMM) maximises the use of commercially available hardware and software. The use of commercially available components minimises both the initial and the maintenance costs.

Since entering service, the SPMM has more tha fulfilled expectations, particularly in ease and speed of use and in quality of output. K&C tests typically have taken two weeks to complete. The new machine enables most cars to be tested in one or two days, depending upon difficulty of vehicle attachment.

REFERENCES

[1] Nedley, A.L., Wilson, W.J., "A New Laboratory Facility for Measuring Vehicle Parameters Affecting Understeer and Brake Steer", SAE Pape 720473.

[2] Ellis, J.R., " A Study of Suspension Mechanisms", Cranfield Institute of Technology, ASAE Report N 5, 1969.

[3] Ellis, J.R., et al, "The Design of a Suspension Parameter Measurement Device", SAE Paper 870576.

[4] Coovert, D.A., et al, "Design and Operation of a New Type Suspension Parameter Measurement Device", SAE Paper 920048.

APPENDIX A

LIST OF RESULTS

The format for presentation of results is specified by the operator. Results are available at the end of each test sequence on PC screen display. Printed results are available immediately upon completion of a test programme.

The following is a list of the "standard" package of tests and results parameters, although it is acknowledged that many clients will have custom requirements, particularly with regard to presentation of results. The facility and service operation is flexible to cater for these requirements.

Vertical Test

- Wheel Vertical Rate (contact patch vertical force versus wheel centre vertical displacement).
- Tyre Radial Rate (contact patch vertical force versus vertical compression displacement).
- Bump Steer (roadwheel steer change versus wheel centre vertical displacement).
- Bump Camber (roadwheel camber change versus wheel centre vertical displacement).
- Bump Castor (roadwheel castor rotation versus wheel centre vertical displacement).
- Track Change (contact patch lateral displacement versus wheel centre vertical displacement).
- Wheelbase Change (contact patch longitudinal displacement versus wheel centre vertical displacement).
- Kinematic Roll centre height for each "axle"
- Kinematic Anti-lift or Anti-drive tangent for each "axle"

Roll Test

- Roll Stiffness (roll moment versus roll angle).
- Wheel loads (roadwheel vertical loads versus roll angle).
- Roll Stiffness Distribution (ratio of individual axle roll stiffness to total vehicle roll stiffness).
- Roll Steer (roadwheel steer change versus roll angle).
- Axle Steer (steer change of beam axle versus roll angle).

Steering System Tests

- Roadwheel Steer versus Hardwheel Steer.
- Ackermann curves.
- King Pin offset, scrub radius and mechanical trail.
- Handwheel Steer versus Handwheel Torque for various control strategies of Roadwheel Steer.
- Steering System stiffness.

Lateral Compliance Tests

- Lateral Compliance (roadwheel centre lateral displacement versus contact patch lateral force).
- Lateral Compliance Steer (roadwheel steer change versus contact patch lateral force); this may be a family of curves with different effective longitudinal moment arms acting in the contact patch.
- Camber Lateral Compliance (roadwheel camber angle change versus contact patch lateral force).
- Roll force centre height for each "axle".

Longitudinal Compliance Tests

- Longitudinal Compliance (roadwheel centre longitudinal displacement versus contact patch longitudinal force).
- Longitudinal Compliance Steer (roadwheel steer change versus contact patch longitudinal force).
- Castor Longitudinal Compliance (roadwheel castor angle change versus contact patch longitudinal force).
- Force derived Anti-lift or Anti-dive tangent for each "axle".

970097

Synthesis of Chassis Parameters for Ride and Handling on the 1997 Chevrolet Corvette

Steven P. Fuja, Henry A. Schmid, and Joseph P. Ryan
General Motors Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper describes the performance attributes of the all-new front and rear SLA (short-long arm) suspensions, steering system, and tires of the 1997 Corvette. The process by which these subsystem attributes flowed down from vehicle-level requirements for ride and handling performance is briefly described. Additionally, where applicable, specific subsystem attributes are rationalized back to a corresponding vehicle-level performance requirement. Suspension kinematic and compliance characteristics are described and contrasted to those of the previous generation (1984 to 1996 Model Year) Corvette. Both synthesis/analysis activities as well as mule-level vehicle development work are cited for their roles in mapping out specific subsystem attributes and related vehicle performance.

INTRODUCTION

The all-new 1997 Chevrolet Corvette has been internally dubbed "C5", shorthand for 5th Generation Corvette. Likewise, the previous generation Corvette (1984 to 1996 Model Year) is referred to as C4 (4th Generation Corvette). These terms will be used throughout.

This paper is one of a set of four SAE papers which discuss the evolution of the C5 Corvette's ride and handling characteristics. Reference [1] is sequentially the first of the four, and describes how vehicle level targets for vehicle dynamics performance were defined. The paper you are now reading is second in the series, and describes how these vehicle targets resulted in specific subsystem attributes for suspensions, tires, and steering system. The third paper in the series (Ref. [2]) describes the execution of the front and rear suspensions in production hardware. The last paper (Ref. [3]) deals with the final development tuning of those suspensions, as well as that of the tires and steering system. Together, the four papers describe the derivation of the C5's vehicle dynamics attributes, from the broadest expression of customer desires and preferences to the specific design details of the final product to be delivered to those customers.

SUBSYSTEM ATTRIBUTES

ROLLDOWN PROCESS - The C5's suspension/steering system design and functionality are the result of a structured flowdown process dictated by General Motors' 4-Phase Vehicle Development Process. This process may be summarized as one in which Voice of the Customer (VOC) inputs flow down to Vehicle Technical Specification (VTS) requirements, which in turn flow down to Subsystem Technical Specification (STS) requirements. See Figure 1. This process is a systems engineering-based method which captures customer inputs and then translates these desires into measurable engineering metrics. Competitive vehicles are benchmarked for these engineering metrics, and the measured results form one set of inputs to the negotiated vehicle-level targets which ultimately become the VTS. (Other inputs to the VTS may include such items as relevant C4 historical data, technology rollout considerations, etc.) These vehicle-level specifications are then rolled down to subsystem and component-level specifications which must be satisfied to enable the vehicle level goals to be met.

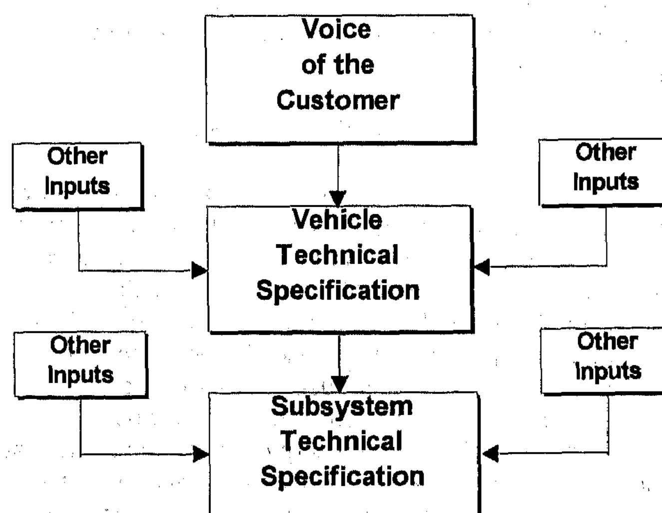


FIG. 1: Requirements Flowdown

In the case of the C5 suspensions and steering system, this rolldown of vehicle-level requirements to specific suspension/steering system attributes was accomplished by a variety of methods, including extensive computer simulation, actual hardware builds (mules, alpha, and beta vehicles), and knowledge by recognized experts in the field (both inside and outside of General Motors).

In systems engineering, equally important to the rolldown process is the process of confirming that subsystem and component goals are met, then re-assembling those validated subsystems back into an overall vehicle, and finally checking the performance of the vehicle against the original goals. Again, the C5 team used both analytical models of subsystems as well as subsystem hardware tests to demonstrate capability, then assembled these subsystems into more complex vehicle models or actual vehicles. Objective vehicle testing against VTS metrics completed the systems engineering process. The result is a requirements-driven vehicle system design enabling world-class ride and handling capability.

SUSPENSION STATIC DESIGN FACTORS - Necessary to understanding a suspension's influence on overall vehicle dynamics is a system of describing that suspension's kinematic and compliance characteristics. At GM, these suspension attributes are known as Static Design Factors (SDFs). Over 60 SDFs are defined and are monitored analytically as well as (in some cases) experimentally as a suspension design evolves. As is the case with any complex system, modifications meant to improve one particular suspension attribute may cause degradation in another area. Packaging changes and other considerations will also drive changes which must be quantified, evaluated, and balanced in terms of impact to overall suspension and vehicle performance. This continuous, proactive modeling, measuring, and balancing of suspension SDFs is key to arriving at a vehicle which delivers capable, pleasible dynamic performance.

FRONT SUSPENSION AND STEERING

TOPOLOGY - Figure 2 shows an isometric view of the C5 front suspension. In the tradition of the earlier generation Corvette, the upper and lower control arms and knuckle are all cast or forged from lightweight aluminum to enable low unsprung mass for good wheel control and impact isolation. Also similar in concept to the previous Corvette is the use of a composite transverse leaf spring and monotube shock absorber. The lower control arms and steering rack mount to a cast aluminum crossmember, which in turn is hard mounted to the frame rails. The crossmember, with its machined interfaces, serves as a precision locating surface for the suspension/steering system components. The upper control arm attaches directly to the hydroformed frame rail. The steering rack, which features Magnasteer™ II (a variable effort steering system developed by GM's Delphi Saginaw Division), is located forward of wheel center. It should be noted that Magnasteer™ was considered part of vehicle program content from the very beginning, and that the front suspension was designed accordingly to complement the Magnasteer™ II system.

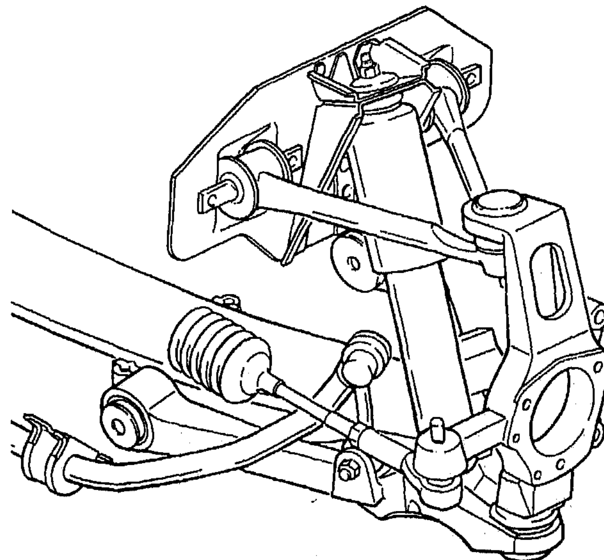


FIG. 2: C5 Front Suspension (Alpha Build)

KINGPIN GEOMETRY - C5's front-view kingpin geometry is set up to enhance steering system performance, with the aim of satisfying vehicle-level requirements for improved steering feel and better mid-range handling. C5 features a relatively short spindle length of 63mm (vs. 93mm in C4), and a relatively upright kingpin (8.8 degrees for C5 vs. 16.0 degrees for C4). Since the spindle length acts as a moment arm for both Fore/Aft and Vertical road input, reduction of this length results in lower kingpin moments due to rough road events, thereby reducing the resulting steering system transient forces and enabling better rough road handling. Reduced spindle length also results in a lower camber moment to be reacted by the upper and lower ball joints; these resulting reduced ball joint lateral forces act to reduce kingpin friction and hysteresis, which may further improve steering feel.

In the side view, C5 features 6.5 degrees of caster angle and 36mm of caster trail. By contrast, the C4 design features 5.9 degrees of caster angle and 45mm of trail, with a 12mm side view spindle offset. The reduction in caster trail was aimed at reduced front cornering compliance to enable a vehicle-level goal of improved yaw damping and better midrange handling. An additional input parameter to the determination of caster geometry, however, was the vehicle-level goal of good steering wheel road feel, with a desire for relatively high handwheel torque gradients (steering wheel effort as a function of lateral acceleration). Since the caster trail is the moment arm for steering system loads resulting from lateral acceleration, this desire for high torque gradients is in conflict with reduction in caster trail desired for improved yaw damping. The need for

balance between these two conflicting goals was recognized early on, and drove extensive computer modeling to explore the relationship among front cornering compliance needs, torque gradient needs, and the interaction of the suspension with steering system tuning.

Even the earliest C5 computer simulations and front suspension mules focused on upright kingpin geometry and carefully balanced caster geometry as an enabler of good steering feel. Competitive vehicles were measured for their kingpin geometric features and were also quantified at a vehicle level for their steering feel performance attributes. Specific steering attributes of some of these competitive cars were eventually factored into C5's VTS.

CAMBER GEOMETRY - C5's camber geometry is shown relative to that of C4 in Figure 3. C5 has a static negative camber setting to allow enhanced cornering power under maximum lateral acceleration. The slope of the camber curve around the layout design position is less aggressive than that of C4, providing less camber change and tread change for low lateral acceleration road disturbances, improving rough road handling and tracking. Roll center height and roll center height migration as a function of ride travel are similar to the C4 design. See Figure 4. Roll stiffness of the new car is somewhat lower than that of C4; accordingly, roll steer is likewise reduced.

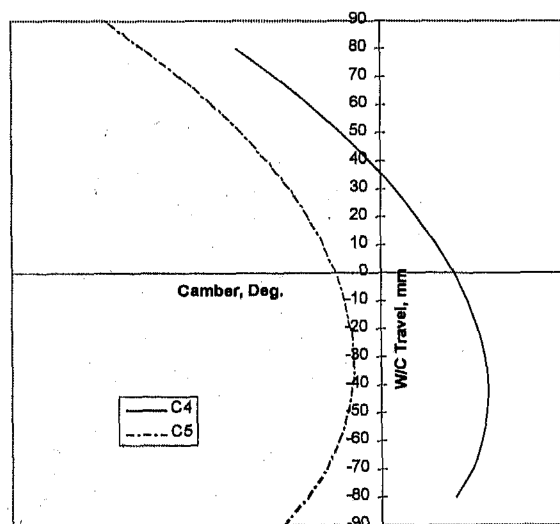


FIG. 3: Front Suspension Camber

Reduction in roll stiffness is part of an overall vehicle level strategy of allowing somewhat more compliance and higher gains, but with greater levels of damping. This approach works to the benefit of both ride and handling by better managing tire normal load variations due to road inputs. The occupants' perception of higher gain may be minimal because they are less sensitive to absolute gain (controlled by stiffness) than they are to rate of change (controlled by damping).

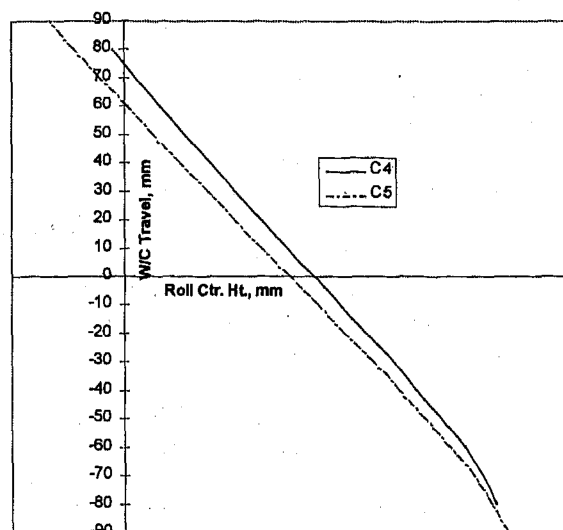


FIG. 4: Front Suspension Roll Center Height

Early suspension mules experimented with different camber curve slopes around zero wheel center travel. These cars, with their adjustable levels of roll camber, were subjectively evaluated for such difficult to quantify attributes as rough road wheel kick. Computer simulations both preceded the hardware development as well as tracked it through development-led changes. The resulting camber geometry and roll gain offers good down the road midrange handling without intrusive wheel kick, yet enables the high maximum lateral acceleration capability necessary in a Corvette.

RIDE QUALITY FEATURES - One message which came through very strongly from the Voice of the Customer was the desire for improved ride quality, particularly in the area of impact isolation. Several suspension features were aimed at satisfying this requirement without sacrificing the handling capability necessary in a high-end sports car.

The lower control arm is set up as a modified L-shaped arm, rather than an A-arm, featuring a stiff bushing close to wheel center for good lateral stiffness, and a soft "ride" bushing aft of wheel center for reduced fore/aft stiffness. These bushing locations and compliances are carefully matched to enable good isolation, yet result in minimal toe change under transient road inputs, to the benefit of rough road handling.

Another ride "enabler" is the transverse fiberglass leaf spring. The packaging efficiency of this springing medium allows the shock absorber to be placed further outboard laterally than could be achieved with a conventional coil spring over shock absorber arrangement; the result is a better shock lever arm ratio (1.34:1), which enables better wheel damping authority for both isolation and roadholding.

A key vehicle-level feature which enables good ride performance out of the suspension is the exceptional rigidity of the C5's unique body structure.

Body stiffness unparalleled in an open roof production car provides a firm reaction surface so that springs, dampers, and bushings in the suspension can function properly without excessive series compliance.

Early mule work also focused on suspension enablers for improved ride quality. Different variants of the leaf spring were tried, as were other springing systems (including a bell crank system with horizontal coil springs over shocks!). Impact harshness considerations were developed in hardware by use of an early mule which had adjustable fore/aft stiffness down to very low levels, as well as adjustable side view geometry; this vehicle was the hardware realization of extensive lumped parameter simulation work which focused on the same attributes.

BRAKE STEER CHARACTERISTICS - Front suspension braking performance characteristics were dictated by a philosophy of making the vehicle insensitive or neutral to acceleration/deceleration events. As such, brake steer characteristics for both symmetric and asymmetric events were kept to a minimum by matching carefully balanced control arm bushing compliances with plan view tie rod angle. These neutral steer characteristics, combined with a wheelbase increase of over 200mm and a slightly positive scrub radius (10mm), provide a significantly greater envelope of vehicle capability under braking.

Brake performance sensitivity to scrub radius was evaluated in hardware by the use of suspension mules using tire/wheel combinations with several different wheel offsets. The wheel offsets used allowed evaluation of changes to both direction and magnitude of scrub radius. This sensitivity work was augmented and enhanced by computer simulation provided by GM's NAO Brake and Bearing Systems group.

STEERING SYSTEM INTERACTIONS - A strong message from the Voice of the Customer was the desire for improved steering feel. As such, the C5 steering system and how it interacts with suspension parameters and tire characteristics was the subject of extensive computer simulation before hardware was ever built. The GM NAO Vehicle Handling Lab has developed sophisticated vehicle models capable of simulating the complex nonlinear functions of the steering system. Steering system characteristics such as hydraulic boost, friction, lash, etc. may be modeled based on physical tests at a subsystem and component level. These analytical tools may be used to simulate full vehicle tests which are statistically correlated to customer preferences. Thus, given good vehicle level targets of steering performance, the team could use these analytical tools to synthesize a design capable of meeting these goals before hardware is built. Subsequent hardware build approximates the desired capability, and is available for development engineers to tune to final configuration using both objective and subjective evaluation.

C5's Magnasteer™ II system is capable of varying handwheel efforts with both vehicle speed and lateral acceleration, enabling a wide tuning range of effort characteristics for varying vehicle regimes (parking lot vs. highway vs. aggressive high lateral acceleration events). This capability was first simulated in the aforementioned steering vehicle models to assess the bandwidth of their effect on customer-perceived attributes; subsequent vehicle development in hardware focused on mapping out performance requirements within this bandwidth to result in a subjectively pleasing package for the customer.

The earliest Magnasteer™ development work in hardware was actually performed in C4 vehicles. Because C5 has different kingpin and steering geometry as well as different tire requirements, the C4 Magnasteer™ experience was not directly applicable to C5. Accordingly, simulation work significantly preceded hardware development work on C5 steering performance. Simulation first focused on modeling and understanding the C4 system to serve as a hardware correlation point. A C5 model was then developed, and this model was used to set subsystem performance attribute targets which were needed to enable vehicle steering goals as defined in the VTS. When actual C5 hardware arrived later in the vehicle program, the C5 models were benchmarked against measured vehicle performance, with the result being exceptionally good correlation.

REAR SUSPENSION

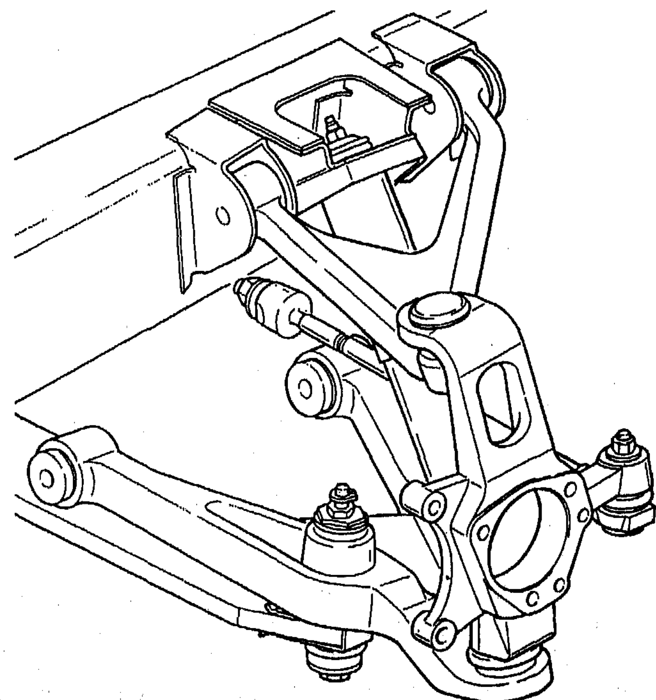


FIG. 5: C5 Rear Suspension (Alpha Build)

TOPOLOGY - Conceptually, the C5 rear suspension resembles the C5 front suspension and is a significant departure from the C4 rear suspension. Like

the C5 front, the C5 rear is a true SLA suspension with a modified L-shaped lower control arm for decoupled lateral and fore/aft compliance. It uses cast aluminum components and features a transverse fiberglass leaf spring and monotube shock absorbers. See Figure 5. Steel plunging halfshafts with constant velocity joints deliver torque to the 18" tire/wheel combination. As in the case of the C5 front suspension (and in contrast to the C4 rear suspension), the lower control arms and tie rods mount to a cast aluminum crossmember with machined interfaces. The upper control arms attach to the hydroformed frame rail. The C5 rear suspension's kingpin geometry is conceptually similar to that of the front due to the fact that knuckle castings are identical, front to rear; angular orientation differences enable front to rear caster and kingpin angle differences. By contrast, the C4 rear relies upon a non-plunging halfshaft to act as the upper control arm; as such, the outer Cardan joint of this halfshaft acts as an effective upper ball joint. The C5 geometry, with separate components (plunging halfshaft and true upper control arm) providing separate functions (torque delivery and suspension control, respectively) enables the reduced scrub radius and spindle length summarized in Table 1.

SDF	C4	C5
Caster Angle (degrees)	1.2	0.2
Kingpin Inclin. (degrees)	-7.1	9.2
Scrub Radius (mm)	162.3	1.0
Spindle Length (mm)	123.0	50.7

TABLE 1: Rear Suspension Kingpin Geometry

CAMBER GEOMETRY - C5's separate upper control arm also allows a more optimal tuning of its camber geometry. See Figure 6. Note also negative static camber setting for improved maximum lateral cornering power. C5's reduction in camber change as a function of ride travel around zero wheel center travel and reduction in tread deviation with ride travel is considered to be directionally correct for improving rough road handling and minimizing truck groove wander. An additional benefit of C5's improved camber geometry is a reduction in roll center height, and an improvement of roll center height change as a function of ride travel; the lower roll center height will result in less jacking during cornering, and the improved slope will result in the roll center height being fixed to the vehicle during ride/roll events, enabling more predictable roll performance. See Figure 7.

Even the earliest mules were SLA architecture, with a separate upper control arm to enable the desired camber geometry. Various levels of roll camber were subjectively evaluated for rough road and truck groove performance. Other factors such as tire size, alignment settings, etc. also influence wander behavior, so these considerations had to be kept in mind when making these evaluations. The resulting combination of camber geometry and roll gain represent the necessary balance between good down the road performance on less than optimal pavement and the necessary high-g cornering power required.

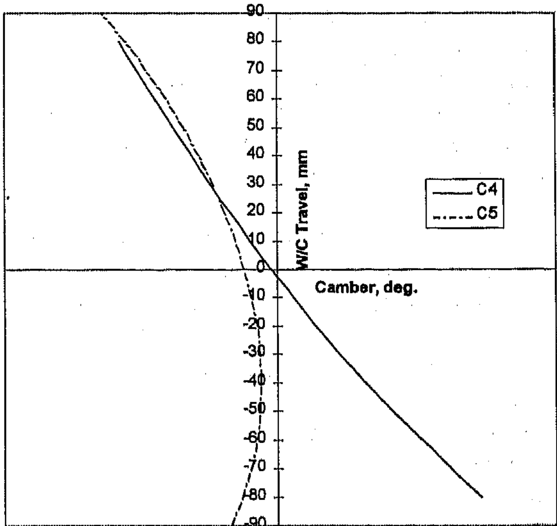


FIG. 6: Rear Suspension Camber

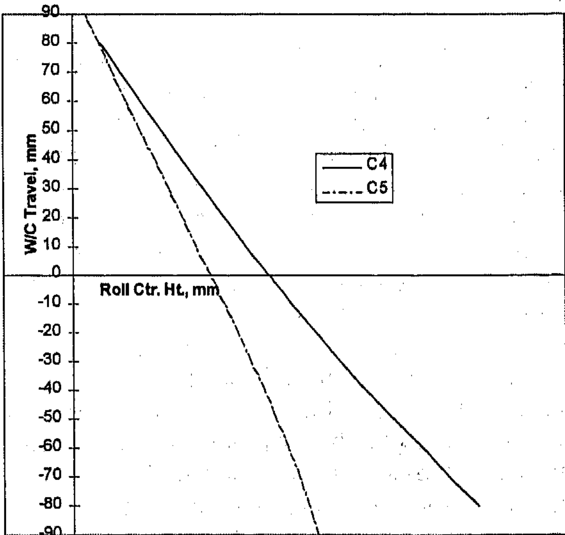


FIG. 7: Rear Suspension Roll Center Height

TOE CHANGE UNDER ACCEL./DECEL. EVENTS - A further benefit of the C5's upper control arm and the resulting reduced spindle length and scrub radius is their influence on the tractive and braking steer

performance of the car. A large reduction in spindle length and scrub radius results in corresponding reductions in kingpin moments due to acceleration/deceleration events. These reduced kingpin moments are then reacted through control arm bushings and a tie rod plan view angle carefully matched to minimize toe change during these events. This is in keeping with the previously stated vehicle-level philosophy of neutral steer characteristics under fore/aft inputs. These suspension features, combined with C5's wheelbase increase, result in clearly superior performance under drop-throttle events.

Toe change under accel./decel. conditions was the focus of considerable early mule work. Early analytical work had focused on minimizing these toe changes as a key enabler of achieving the neutral character desired in the car. Early hardware execution, however, pointed out the importance of such subtleties as body and component series compliance, as well as bushing non-linearities. Each of these additional compliances were tracked down in turn and were either minimized or managed through complementary hardware and simulation work. Resulting toe change under free-rolling conditions was then predicted analytically as well as measured experimentally on suspension compliance measurement systems such as GM's Vehicle Handling Facility and a similar facility at Goodyear's Akron Technical Center. The resulting toe change under fore/aft inputs is exceptionally low, resulting in the desired vehicle level performance under accel./decel. events.

This careful management of toe change under fore/aft road inputs is also a key enabler of ride isolation improvement. Because steer characteristics are so well controlled, greater compliance can be allowed without sacrificing handling precision, to the benefit of impact harshness.

LATERAL FORCE STEER CHARACTERISTICS - C5's use of a rear-mounted, stiff tie rod (toe link) allows it to toe in slightly under the application of a cornering load. This lateral force steer is in the understeer direction, and thereby reduces rear cornering compliance. This in turn enhances lateral acceleration response time and thereby addresses Voice of the Customer inputs which demand responsiveness in midrange handling events.

C5's lateral force understeer is enabled by virtue of the body-side toe link attachment to a rigid crossmember, rather than to a rubber isolated differential (as is the case with the C4 rear suspension). The direction of toe change under the application of lateral load was the focus of considerable debate, test, and simulation work by experts both within and outside of GM during the early phases of the C5 vehicle program. All sides agreed that very small, controlled levels of lateral force steer were appropriate for the C5, with the difference of opinion arising in regard to the sign of that toe change (toe in or toe out). Tuning philosophy, subjective preferences, and development heritage played a strong role in the determination of opinions on this topic. In the end, the very small amounts of toe-in exhibited by C5 were considered appropriate as an enabler of mid-range responsiveness, as well as being subjectively pleasing to the driver.

SIDE-VIEW GEOMETRY - C5 features a relatively long side-view swing arm (2.1 meters) and a somewhat reduced side-view swing arm angle (5.2 degrees for C5 vs. 7.8 degrees for C4). This results in the reduced levels of Percent Anti-Squat (PAS) and Percent Anti-Lift (PAL) shown in Figures 8 and 9. Long side-view swing arms with relatively low angle changes as a function of ride travel result from C5's SLA architecture; trailing link suspensions such as C4 usually result in shorter side-view swing arms with more angle change with ride travel.

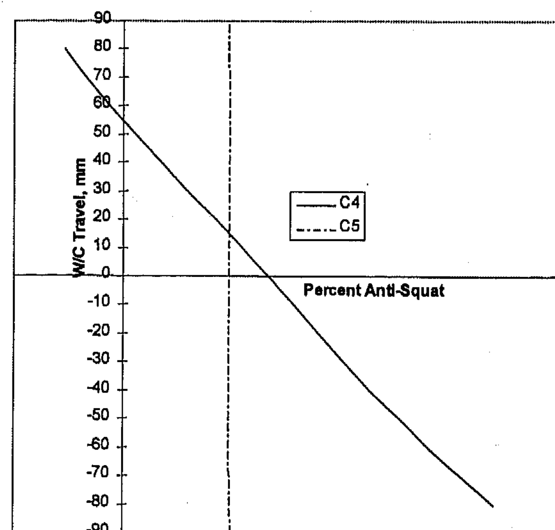


FIG. 8: Rear Suspension Percent Anti-Squat

C5 is able to reduce PAS and PAL and still retain good pitch control because of its increased wheelbase. Equally important to absolute pitch gain is the driver's perception of pitch under accel./decel. events. In a two-seat car such as the Corvette, the occupants sit closer to the rear axle than would be the case on a car which has a back seat; rear suspension anti features may therefore be more perceptible to occupants than would be the case on other types of vehicles.

Earliest mule work focused on very low levels of anti features, with the mule also incorporating the C5's extended wheelbase. Since this earliest mule was deemed too "pitchy" under acceleration, subsequent development work focused on a mule with adjustable anti features. This vehicle was used to more precisely map out driver preferences with respect to pitch control. Obviously, spring rate and shock authority also contribute to pitch perception, so tuning considerations had to be kept in mind when making these evaluations.

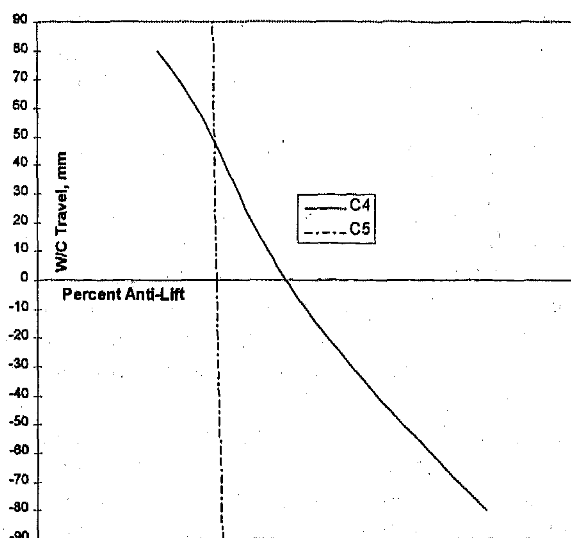


FIG. 9: Rear Suspension Percent Anti-Lift

SPRING AND SHOCK CONSIDERATIONS - The rear leaf spring is packaged beneath the lower control arm, suspended by rubber isolated tension links. These links serve to decouple the leaf from the lower control arm in the fore/aft direction, preventing the leaf spring from contributing to fore/aft suspension stiffness, and thereby reducing impact harshness.

The choice of a leaf spring for the C5 rear suspension was influenced by needs of packaging efficiency. Benefits offered include increased trunk volume as well as enhanced shock lever arm ratio. The decision to mount the spring and shock to the lower control arm (as opposed to mounting directly to the knuckle) was a conscious tradeoff of what would have offered an even greater potential improvement in shock lever arm ratio. Mounting these elements to the lower control arm results in a knuckle which is decoupled from the shock and spring transient forces which might result in undesirable toe changes, particularly under rough road events.

Additionally, mounting the shock to the lower control arm provides benefit by decoupling fore/aft stiffness from the shock absorber's load path. This reduces series compliance to the damper and thereby enhances shock authority. In suspensions which mount the shock directly to the knuckle, a "bell crank" mechanism can result, whereby reducing the fore/aft stiffness to improve impact harshness can result in more effective compliance acting in series with the shock absorber. This in turn results in diminished capability for wheel control. An early C5 candidate rear suspension which featured a multi-link lower control arm exhibited this undesirable coupling in analytical modeling.

TIRE CONSIDERATIONS

EXTENDED MOBILITY TIRES - Standard on the C5 is the use of Goodyear Eagle F1 GS Extended Mobility Tires (EMT), otherwise known as "run-flat" tires. These tires offer the capability of driving with zero air pressure; this capability, combined with C5's Tire Pressure Monitor system, obviates the need for a spare tire, benefiting the customer both in terms of increased trunk space and reduced overall vehicle mass.

While EMTs offer obvious benefits to the customer, they also present unique challenges to the suspension design. EMT construction which allows run-flat capability also generally results in higher vertical tire spring rates and increased unsprung mass. Both of these characteristics tend to work against good impact isolation and good wheel control. Fortunately, C5's suspensions were designed with EMTs considered standard equipment from the very beginning, so these tire attributes were recognized and comprehended early in the suspension design. Additionally, C5 benefited from lessons learned on the C4 Corvette, which offered an earlier version of EMT tires with one suspension option package. Accordingly, the C5 suspensions were designed to provide relatively low fore/aft stiffness (with corresponding neutral steer characteristics) to minimize impact harshness. Additionally, C5's exceptional body stiffness at local shock/spring/bushing attachment points allow optimal isolator and damper performance to the benefit of both isolation and wheel control.

Early suspension mules relied upon C4 EMT tires. Since these earlier variants of EMT technology are somewhat stiffer and more massive than those which ultimately were developed for C5, they served to provide a worst case scenario for considerations such as impact harshness, wheel control, and peak jounce bumper loads. Continuously during the course of the vehicle program, development engineers from both GM and Goodyear worked suspension compliances, shock and spring combinations, and tire characteristics in a complementary fashion to arrive at the resulting total C5 package.

LATERAL REQUIREMENTS - Tire characteristics form the basis of front and rear cornering compliances and thereby vehicle yaw plane behavior; suspension attributes serve to modulate those tire characteristics. Accordingly, proper specification of tire handling characteristics such as cornering coefficient, aligning torque coefficient, camber stiffness, weight transfer sensitivity, etc. is critical. Targets for these parameters were rolled down very early in the program via the use of coarse, lumped parameter vehicle models. These targets were then monitored and modified as appropriate as the design continued to evolve and mature. The C5's differential tire sizes front-to-rear are critical to enable proper front and rear cornering coefficients for desired yaw plane behavior; the C5 front

tires are P245/45ZR17, and the rear tires are P275/40ZR18. Goodyear was identified early on as a potential source for C5 tires and their representatives were ongoing participants in the C5 Ride and Handling Product Development Integration Team.

In addition to their role at the highest level of vehicle yaw plane dynamics, tire lateral characteristics also play a key role in steering system performance and feel. Likewise, tire asymmetries and suspension alignment also influence customer perception of steering feel. As the C5 design matured, the influences of such parameters on steering performance were comprehended in detailed vehicle models.

SUSPENSION INTERFACE STIFFNESS REQUIREMENTS

GLOBAL STIFFNESS REQUIREMENTS - The C5 is unprecedented in terms of overall global stiffness in an open roof production car. First body bending and torsional natural frequencies are in the range more commonly associated with those of good fixed roof coupes and sedans. The development of C5's unique body structure was the result of extensive up-front specification, benchmarking, and simulation work and is documented in Ref. [4].

This up-front focus on exceptional global body stiffness works to the benefit of both ride and handling performance. It also serves to illustrate one of the many interfaces between the chassis/suspension activity and simultaneous work in other areas of the car. Each of these interfaces had to be carefully mapped out and managed, with each of the interfacing disciplines imposing requirements upon its interface counterpart. In the case of the body/chassis interface with respect to vehicle modal performance, it meant that each of the suspension modes (hop, tramp, and fore/aft) was assigned a frequency "slot", into which it had to fall. This allowed suspension modes to be decoupled from body and powertrain modes to enable good noise and vibration performance. As an example, consider suspension wheel hop frequency. In a high end sports car, this typically falls in the range of 12-15 Hz. Unfortunately, typical open roof vehicles usually have body torsional modes in that same frequency band. This can result in coupling between suspension and body modes and poor vibration performance. In the case of C5, however, this assigned slot of 12-15 Hz for suspension hop is far below the C5's body torsional mode, allowing decoupled modal performance with good isolation. This pro-active modeling and managing of modal requirements is another key contributor to C5's significant improvement in ride quality and isolation performance.

LOCAL STIFFNESS REQUIREMENTS - Equally important to global stiffness, however, is local stiffness requirements at suspension interfaces. In the case of C5, some of these interfaces are upper control arm to frame rail, shock absorber to shock tower, and lower control arm to the cast aluminum crossmember. Additional interfaces are steering rack to (front) crossmember, and rear suspension tie rod to (rear) crossmember. The suspension/tire local stiffness interface is through the bearing and wheel stiffness. Targets for each of these local stiffnesses were assigned, tracked, and managed through the course of C5 development.

Defining interface stiffness results in several obvious benefits as well as some more subtle. One obvious benefit is the suspension tuning flexibility afforded by stiff suspension interfaces. Isolation elements such as bushings are able to function to their maximum capability if unwanted series compliance effects are minimized. Likewise, shock authority is not limited by series compliance effects in the shock tower.

Another obvious advantage of stiff suspension interfaces is the beneficial effect on road noise. Early analytical work identified important noise paths, then local stiffness targets were assigned to allow the isolator to provide an appropriate input mobility mismatch across the interface.

A more subtle benefit, not always recognized, is the fact that good stiffness at suspension interfaces allows up-front synthesis/analysis to be more accurate and effective. Early in a vehicle program, analytical models are of necessity more coarse in order to quickly evaluate numerous design alternatives without the benefit of large amounts of design data. Some types of suspension models used in this phase are rigid body models, without structural compliance effects explicitly modeled. If suspension interfaces are eventually managed to be appropriately stiff, these early approximations of rigid behavior are much closer to the mark than would be the case if significant series compliance effects resulted. Obviously, as a vehicle program progresses, series compliances can be tracked, modeled, and managed, but if those contributions are kept to a minimum, the resulting vehicle design will perform much closer to those early predictions.

CONCLUSION

An exhaustive team effort over the course of several years led to C5's ride and handling chassis systems as they are now defined. Early vehicle level specification work set the stage with extensive customer input as well as that of recognized experts. A critical program enabler was clear management direction at an early juncture that the C5 would be allowed all-new suspension designs, to be shared with no other GM vehicle. Computer simulations of vehicle performance as well as those of the subsystems which enable that performance led to design solution requirements; that effort was augmented by mule vehicle level development work. The use of premium and innovative materials and components allowed design solutions to be executed in hardware. Development engineers then tuned the whole package into a pleasible, capable vehicle for the customer. The result is a no-compromises balance of

ride and handling which extends and expands upon the Corvette's unique heritage.

ACKNOWLEDGMENTS

The evolution of the C5 ride and handling chassis systems was the result of an exceptional team effort by a group of individuals too numerous to recount. The authors would be remiss, however, if they failed to acknowledge the contributions of the following individuals: Al Bodnar, Dave Wickman, Jim Lloyd, Dick Topping, Rich Kowalczyk, Mary Dona, Scott Allman, Rex Stump and Mike Neal.

REFERENCES

[1] "Objective Ride and Handling Goals for the 1997 Chevrolet Corvette", Joseph P. Ryan, Steven P. Fuja, Henry A. Schmid, General Motors Corporation, SAE Technical Paper 970091, 1997

[2] "Design Synthesis of Suspension Architecture for the 1997 Chevrolet Corvette", Henry A. Schmid, Steven P. Fuja, Joseph P. Ryan, General Motors Corporation, SAE Technical Paper 970092, 1997

[3] "Ride and Handling Development of 1997 Chevrolet Corvette", Michael W. Neal, Mary A. Dona, General Motors Corporation, SAE Technical Paper 970098, 1997

[4] "The 1997 Chevrolet Corvette Structure Architecture Synthesis", Stephen D. Longo. Edward D. Moss, Brian W. Deutschel, Midsize/Luxury Car Group, General Motors, SAE Technical Paper 970089, 1997

DEFINITIONS, ACRONYMS, ABBREVIATIONS

SLA: Short-Long Arm

C4: Fourth Generation Corvette (1984 to 1996 Model)

C5: Fifth Generation Corvette (1997 Model)

VOC: Voice of the Customer

VTS: Vehicle Technical Specification

STS: Subsystem Technical Specification

Mule: Earliest vehicle build, consisting of components or subsystems built into a modified current production vehicle.

Alpha: An early vehicle build in which major subsystems are first integrated into a vehicle.

Beta: A later, more refined vehicle build featuring updates from previous build cycles.

SDF: Static Design Factor

PAS: Percent Anti-Squat

PAL: Percent Anti-Lift

EMT: Extended Mobility Tire

Magnasteer™ II: GM Delphi Saginaw Division's Variable Effort Steering System. It allows steering wheel effort as a function of vehicle speed and lateral acceleration.

970098

Ride and Handling Development of the 1997 Chevrolet Corvette

Michael W. Neal and Mary A. Dona
General Motors Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper describes the ride and handling development process used for the 1997 Corvette. Three levels of suspension are available for the 1997 Corvette: base (FE1), sport (FE3) and RTD or Real Time Damping (F45) suspensions. All suspensions will be discussed in this paper. A review of the development and vehicle integration tradeoffs for each of the specific chassis components is included. Control arm bushings, springs, jounce bumpers, anti-roll bars and insulators, tires, shock mounts, shock absorber valving, real-time damping, steering development, alignment and measurements are discussed.

INTRODUCTION

OVERVIEW - The fifth generation Chevrolet Corvette introduced in 1997 (C5) is a completely new vehicle, featuring significant design changes in areas compared to the fourth generation Chevrolet Corvette of 1984-96 (C4).

Four papers have been written that summarize the Systems Engineering approach to the suspension design of the 1997 Chevrolet Corvette. The first paper [Ref. 1] reviews the customer research, Engineering Direction, human factors and analytical evaluation that established the objective Ride and Handling targets for the vehicle. The second paper [Ref. 2] describes the "roll down" of the vehicle level performance targets to specific subsystem attributes and parameters for the suspension, steering, and tires. The third paper [Ref. 3] reviews the hardware that complies with the specifications established in the previous two steps. This paper summarizes the final development tuning of the suspension, steering and tires. Careful execution of the procedures and techniques discussed in this series of papers was critical in producing the highly optimized suspension of the 1997 Chevrolet Corvette and for assurance of high customer satisfaction for ride and handling performance.

DEVELOPMENT PROCESS - C5 ride and handling goals specified significantly improved ride, but not at the expense of handling. In fact, handling should be improved as well. To achieve both meant that suspension compliances had to be present for enhanced ride quality, but needed to be managed for nimble and predictable handling. The chassis development process optimized the suspension components to obtain the specified linear and non-linear range handling characteristics while achieving exceptional ride quality. Enablers for this process were an exceptionally stiff structure (450% stiffer in torsion than C4) to mount the springs, shocks and bushings, and a new optimized suspension geometry specifically designed to manage kinematics and compliances and allow each component to function properly.

CHASSIS INTERFACE

STRUCTURAL FOUNDATION - The C5 structure was a clean sheet of paper design [Ref. 4]. The new structure enabled the suspension to be tuned for improved ride and handling by decoupling the structure and suspension dynamically. Control arm bushings, shocks, springs and anti-roll bars were tuned independent of structural interaction, unlike C4, which had to account for structural interaction. C5 structure had four primary design specifications to decouple structure and suspension, as shown in Table 1.

Table 1. Body Structure Specifications.

Mode	Freq Range (Hz)	Supports	Specification
Static Stiffness	0-5	Handling	720 N-mm/deg (>10X roll couple distribution)
1st Torsional	5-50	Structural	23 Hz (separation from suspension tramp mode)
		Vibration	Nodal Lines @ front edge of seat, vehicle centerline
1st Bending	5-50	Structural	21 Hz (separation from suspension hop mode)
		Vibration	Nodal lines @ front and rear shock towers
Input Mobility	50-800	Noise	20 dB separation from mount

STATIC STIFFNESS - The static stiffness specification insured that the torsional stiffness of the structure did not interact with the roll stiffness distribution of the front and rear suspensions. Thus, anti-roll bars and ride springs were very efficient, and did not require oversizing to compensate for torsional deflection of the structure.

STRUCTURAL 1st TORSIONAL MODE - The 1st torsional mode of the structure was decoupled from the suspension by frequency and mode shape. The 23 Hz frequency provided adequate separation from suspension tramp (16 Hz), so tramp mode vibrations were not amplified by the structure. By placing the 1st torsional mode nodal line at the front edge of the seat and center line of the vehicle, perception of shake was minimized at the seat track and steering wheel. Shock tuning could focus primarily on wheel control without compromise to control structural shake.

STRUCTURAL 1st BENDING MODE - Similar to the 1st torsional mode of the structure, the 1st bending mode was decoupled from the suspension by frequency and mode shape. The 21 Hz frequency provided adequate separation from suspension hop (15 Hz), so hop mode vibrations were not amplified by the structure. By placing the 1st bending mode nodal lines at the front and rear shock towers, the structural bending mode was further isolated from road induced excitation. The primary excitation by the suspension for stuctural bending vibrations occurs at the least sensitive area of the structure. As with structural 1st torsion, shock tuning could focus on wheel control without compromises to control structural shake.

INPUT MOBILITY - Specifying 20 dB separation between the suspension isolation components and the attaching structure enabled interior road noise specifications. The mobility specifications insured the effectiveness of each isolator by forcing each to be the weak link in the system.

Due largely in part to the comprehensive specifications created for each major subsystem at the beginning of the program, chassis development was able to achieve all the targets for ride, handling, vibration, and noise. Only predetermined tuning elements were adjusted, no suspension or structure revisions or patches were required (for all three levels of suspension).

CHASSIS COMPONENT DEVELOPMENT

This section describes the development process for optimizing the individual chassis components of the FE1, FE3, and F45 suspension packages.

A minimum amount of effort was required to tune springs and anti-roll bars. Their development was straightforward, and once established, did not have to be revisited at a later date to "fix" subsequent issues.

Emphasis was placed on control arm bushing development because of their key role in achieving the C5 ride and handling goals. Jounce bumper development was a major subset of chassis development, because of it's role in managing the highest suspension and structure loads.

DEVELOPMENT TIME LINE - The ride and handling process was very iterative and was active over the entire C5 program. Different types of vehicles were used to prove different objectives during that time.

Mule vehicles built from C4s were used initially to prove suspension concepts and help set goals and specifications for the C5.

Pre-Alpha vehicles were the first chassis built from the ground up and closely represented design intent at that time, except styling. These vehicles were used to prove out structural design and suspension isolation capabilities and helped establish the Alpha and Beta tuning libraries for chassis components.

Alpha vehicles were the first completely new vehicles built from the ground up that represented 100% mainstream design. They included functionally correct components in all areas. These vehicles and the previously established tuning libraries were used to establish the preliminary chassis tuning of the complete, design intent vehicle.

TIME (arrow pointing down)

VEHICLE LEVEL	CHASSIS DEVELOPMENT ACTIVITY
MULES	SUSPENSION CONCEPTS
PRE-ALPHA	STEERING
	RTD
ALPHA	STEERING
	TIRE MOLD SHAPE
	SHOCK VALVING
	RTD
	SHOCK MOUNTS
	CONTROL ARM BUSHINGS
BETA	STEERING
	TIRE TREAD PATTERN
	SHOCK VALVING
	RTD
	SPRINGS
	JOUNCE BUMPERS
	STAB BARS
	SHOCK VALVING
PROTO	TIRE CONSTRUCTION
	SHOCK VALVING
	RTD
	STEERING
	SHOCK MOUNTS
	ALIGNMENT

Figure 1. Chassis Development Timeline.

Beta vehicles were also built from the ground up, with the first production intent components, and were used for the majority of chassis tuning.

Prototype vehicles were built with production parts and were used to refine and finalize the tuning process. The timeline in Figure 1 shows the development sequence and the iterative nature of the development of the C5 suspension.

CONTROL ARM BUSHINGS - Due to the robustness of the suspension design, only one control arm bushing package was necessary for all suspension levels. The assembly plant was willing to proliferate, but the suspension design did not require it.

The C5 uses Short-Long Arm (SLA) suspensions front and rear, as shown in Figures 2 and 3. The control arms were designed as close to an "L" shape as possible, with a stiff "handling" bushing near wheel center and soft "ride" bushing aft of wheel center. The front upper control arm is a symmetric A-arm, due to packaging constraints of the front suspension during full lock steering. The "separation of function" concept for the control arm bushings worked very well in hardware execution.

Except for the front upper control arm (which was subject to very tight package constraints), all other control arm bushings were sized sufficiently to accommodate the anticipated rate of each, using a conventional rubber/elastomer bushing. Bushing durability, linear and non-linear rate characteristics, contribution to ride rate, and other factors were considered. Higher cost bushing alternatives (i.e. lubricated journals, voided or non-symmetric, cross-axis ball joints) were considered, but determined to not be necessary with the C5 suspension designs.

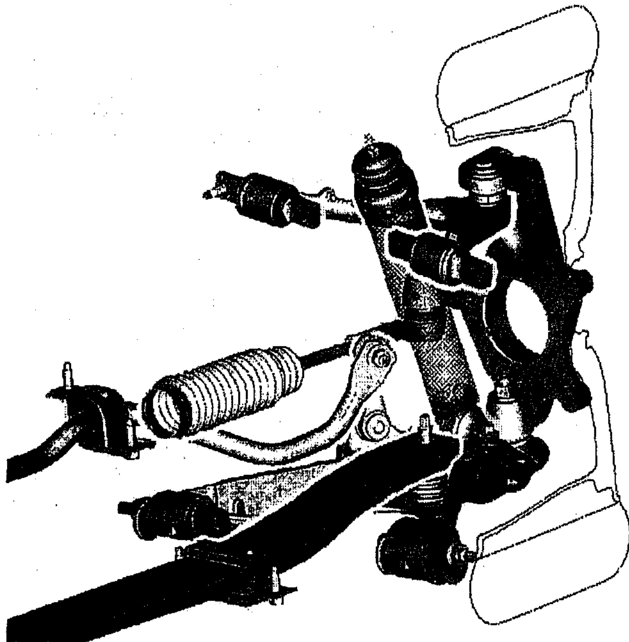


Figure 2. 1997 Chevrolet Corvette Front Suspension.

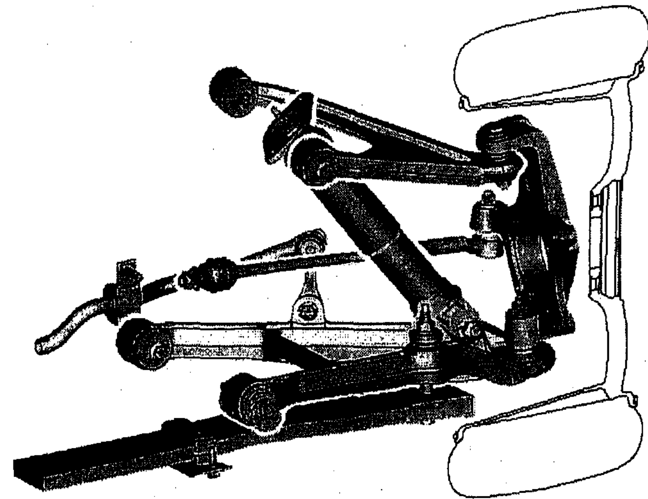


Figure 3. 1997 Chevrolet Corvette Rear Suspension.

Ride Bushings - Reducing the rate of any ride bushing improved impact hardness or reduced interior noise but not have a significant impact on handling. The suspension kinematics were robust to normal changes in bushing rates, compliance SDFs were somewhat decoupled, and primarily, unwanted toe change was non-existent. The suspension elasto-kinematics were designed for near-zero toe change with fore-aft displacement. Development experience with the ride bushings confirmed that large reductions in rate still did not introduce unwanted toe change. Handling modes such as rough roads, split-coefficient braking, braking-in-a-turn, accelerating in a turn, and dropped throttle were not significantly affected, a tremendous advantage for tuning suspension fore-aft compliance.

If the ride bushings were stiffened, there was a negligible improvement in handling, and always a significant degradation to ride quality and interior noise. There was little to be gained from specific bushing packages for each suspension option. Only one control arm bushing configuration was required to obtain the desired handling characteristics plus refined ride for all three suspension levels.

Handling Bushings - Increasing the rate of any handling bushing would improve multiple areas of handling but not significantly degrade impact hardness or increase interior noise. Handling bushings were extremely stiff, typically 3 to 6 times higher rate than ride bushings.

Analysis Support - A system to objectively quantify the kinematic and compliance characteristics of a suspension is necessary to understand, measure and simulate a suspension's influence on overall vehicle dynamics. At General Motors, these suspension attributes are known as Static Design Factors (SDFs). Over 60 SDFs were defined in analytical models, many are used in handling models and are monitored on the vehicles. As may be the case with any complex system,

modifications that improve one suspension attribute may degrade another [Ref 2].

Chassis Analysis synthesized a set of nominal bushing rates to meet the suspension SDF targets that in turn were synthesized to achieve the handling requirements. Using the nominal bushing rates as a basis, a tuning library of 20% softer and 15% stiffer bushing rates was obtained and subjectively evaluated by Chassis Development.

To assist the development process, a bushing rate sensitivity study was run which examined the effect on handling of varying each bushing rate individually or several simultaneously. Eighty-one combinations of nominal, soft, and stiff bushing rates were analyzed for the front and rear suspensions. Seventeen compliance related SDFs were tracked.

An example of the analysis is shown in Figure 4, where 1 of the 17 SDFs (Lateral Force Compliance Steer) is shown for 11 of the 81 bushing combinations.

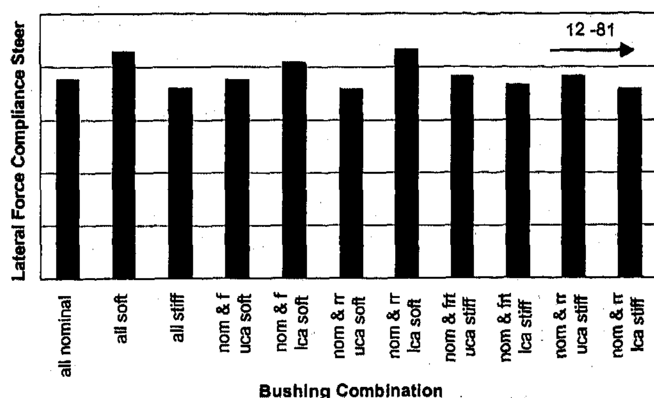


Figure 4. Sensitivity of Lateral Force Compliance Steer to Various Bushing Rate Combinations

The SDFs found to be most sensitive to bushing rate changes in the front suspension were: Lateral Force Compliance Steer, Stiffness at Wheel Center (Lateral, Fore-aft, and Pitch) and Brake Steer. The rear suspension showed the same sensitivities as the front, plus Tractive Force Steer.

The magnitudes of the steer SDFs were in all cases very small (again, a key feature designed into the C5 suspension). For example, in the final tuning of the rear suspension, Tractive Force Steer was reduced 65% over C4, while simultaneously reducing Wheel Center Fore-aft Stiffness by 10%.

Results of the bushing rate sensitivity study were used to reduce the number of bushing configurations to be subjectively evaluated. Combinations were evaluated for handling, impact hardness, structure borne road noise, on-center handling, lane change and recovery, braking in a turn, braking on split coefficient surfaces, any many more.

Using information from the analytical studies and subjective evaluation, Development rapidly arrived at a final control arm bushing tuning for the C5.

Figure 5 show the relative magnitude of each bushing rate for the initial synthesized values and the final developed values. The rate difference between the ride and handling bushings of the L shaped arms (all except front upper) may be seen. For the lower control arms, the handling bushings got stiffer and the ride bushings got softer. All the upper control arm bushings got softer.

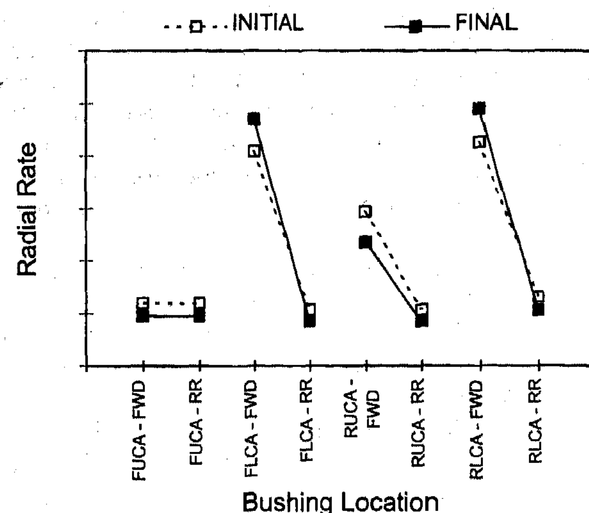


Figure 5. Control Arm Bushing Rates - Initial and Final.

Lower Control Arm Bushings - Bushing development of both the front suspension and rear suspension lower control arms followed nearly identical sequences.

Stiff (Handling) Bushings - The forward bushing of each arm was considered the "handling" bushing (closest to wheel center in plan view). For each lower control arm, the handling bushing was approximately 5 times stiffer than the ride bushing. From the library of bushings available, the rates selected by Chassis Development were each 15% stiffer than the original synthesized values. There were improvements in all areas of handling, particularly on-center. The control arms pivot about the handling bushing in plan view, and there was minimal degradation in impact hardness. There also was minimal increase in structure borne noise transmission, due primarily to the high impedance and damping of the crossmembers [Ref. 5].

Soft (Ride) Bushings - The aft bushing of each arm was the "ride" bushing, which absorbs the fore-aft input of a road disturbance. The rates selected by Chassis Developmentdeveloped rates were each 20% lower than the original synthesized values, which yielded a significant reduction in impact hardness. Very low fore-aft suspension rates were achieved, yet did not create additional toe change under fore-aft disturbances, braking, or acceleration, due to the carefully managed

elasto-kinematics of the suspension [Ref. 2]. C5 is significantly better than C4 for dropped throttle characteristics in a corner.

Upper Control Arm Bushings - The front upper control arm was a symmetrical design due to packaging constraints [Ref. 3] and was not able to separate bushing function. Both forward and aft bushing rates were the same. The rear upper control arm was a true L shape, with the forward bushing directly on wheel center. The rate of the handling (forward) bushing was approximately 3 times higher than the aft bushing.

For all upper control arm bushings, the developed rates were 20% lower than the original synthecrease in road noise. The upper control arms are attached to the side of the main frame rails (one piece hydroformed tube). The frame rails, being large cross-section and continuous tube, can potentially carry road noise to the occupants if the source is not properly isolated. Softer bushings for the upper control arms improved the impedance mis-match across each noise path.

There were no significant changes in handling. There was no impact to on-center performance, because of the low friction ball joints that were specified by early steering system synthesis and verified by development (see steering section). Control arm bushings can contribute to steering compliance if high steering forces are required to overcome high (rotating) ball joint friction. There was no effect on handling or impact hardness, fore-aft loads are relatively small in the upper arms. The control arm bushings were developed in Alpha vehicles. There was sufficient lead time to redesign the upper control arms to accommodate a larger outer diameter for the lower radial rates.

SPRINGS - With continuation of the composite transverse leaf spring for C5, there was a 15 year history to draw upon. FE and F45 share the same rates, and the FE (sport suspension) had higher rates, as shown in Table 2. Ride frequencies were developed to meet subjective requirements for controlling rigid body motions of the sprung mass, primarily flat ride. Required spring rates were then determined from the ride frequency targets.

Table 2. Spring Rates and Ride Frequencies.

	Front Suspension		Rear Suspension		
Suspension Option	Spring Rate (N/mm)	Ride Frequency (Hz)	Spring Rate (N/mm)	Ride Frequency (Hz)	Ride Frequency Ratio
FE1 & F45	77	1.15	103	1.36	1.18
FE3	93	1.20	113	1.45	1.21

Ride motion flatness was defined as a minimum pitch motion as the vehicle settled from cresting a swell. The front and rear spring rate balance was developed without shocks or jounce bumpers on the vehicle, since damping and jounce bumpers also affect ride motion flatness. This enabled the front and rear spring rate balance to be optimized for ride motion flatness ensuring that shocks and jounce bumpers weren't compromised by doing unnecessary work to control ride flatness motion.

The FE1 and F45 rear spring rate and ride frequency were increased beyond synthesis predicted values to control rear ride motions. The front ride frequency was then selected by evaluating ride motion flatness. The front spring rate increased 15% over the synthesis prediction, due to the lower torsional contribution of the developed control arm bushings.

The FE3 rear spring rate was selected from the synthesis prediction of rear ride frequency. The front rate was selected by evaluating ride motion flatness as in the FE1 and F45 suspensions. The developed front spring rate was 14% greater than the synthesis prediction.

JOUNCE BUMPERS - C5 jounce bumpers are integral with the shock absorbers, and must comprehend the load multiplication effect of the shock travel ratios. Jounce bumpers were developed for subjective ride entry feel, ride motion flatness and managing shock tower loads for a severe pothole event that is part of standard Proving Ground vehicle durability testing. The Corvette must pass the same Proving Ground durability tests as all other General Motors passenger cars and light trucks.

C5 had higher shock tower loads to manage than the C4 due in part to the stiffer body structure. The C5 rear suspension also had the added challenge of having a higher wheel/shock travel ratio (1.43), as compared to the C5 front suspension (1.34), which was a multiplication factor for the loads into the jounce bumper.

Keeping shock tower loads below the target levels led to an in depth study of jounce bumper energy absorption capabilities. Analytical modeling identified that shock tower loads management depended on a system of free travel, shock valving, and jounce bumper configuration. A jounce bumper configuration was characterized as a force-displacement curve. Assuming shock valving and free travel remained constant, a proposed jounce bumper configuration for loads management could be rapidly determined.

Coordinated Test, Analysis, and Development - A physical jounce bumper test was developed for evaluation of different configurations. In th bench test, a mass was dropped onto the jouce bumper configuration and drop height adjusted until the peak load correlated with the measured vehicle load. The same drop height and mass combination was then used to assess all subsequent jounce bumper configurations.

The bench test was a repeatable and fast method of quantifying each jounce bumper configuration's ability to absorb the amount of energy present in a severe pothole event. The jounce bumper configurations included variations in height, diameter, shapes, containment, and materials. An example of a

force-displacement curve for two typical jounce bumper configurations is shown in Figure 6. The measured force-displacement curve from the bench test as well as free travel and shock damping curves were then used in analytical models to predict and guide jounce bumper selection for loads management.

From initial analytical predictions, it was determined that the lowest peak loads could be obtained from a tall, uncontained jounce bumper. This jounce bumper configuration had good entry feel but ride motion flatness was unacceptable. The tall bumper decreased free travel and interfered too much in ride motion and wheel control development work. This required the shock damping to be revised for spring rate plus jounce bumper rate instead of spring rate only.

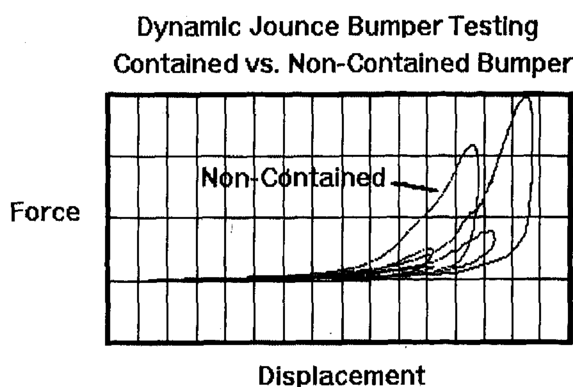


Figure 6. Jounce Bumper Force-Displacement Comparison.

The high shock travel ratio caused rear jounce bumper contact during roll at approximately 0.5 g lateral acceleration. This increased rear roll rate and upset the Tire Lateral Load Transfer Distribution (TLLTD). The nonlinear range understeer balance of the vehicle changed during limit handling maneuvers. To alleviate this phenomenon, a smaller rear anti-roll bar with 15% less vehicle roll stiffness was necessary to balance the vehicle.

To remedy the compromises in the shock damping and anti-roll bars caused by the long jounce bumper, a shorter jounce bumper coupled with a slight trim height change was developed which kept the suspension from contacting the jounce bumper during limit handling maneuvers. Jounce bumper configurations and revised shock damping recommended by Chassis Development were bench tested and checked in the analytical model to ensure shock tower loads were maintained below the target values. When a suitable jounce bumper configuration was found, the original rear anti-roll bar was installed and the handling characteristics were noticeably more linear (roll stiffness now relied on the linear rate elements versus the nonlinear rate of the jounce bumper).

Shock damping curves were then tuned to provide better wheel control over chatter bumps, improved impact isolation and flatter motion over swells The revised shock damping was included in the analytical model, and shock tower loads were further improved. Final changes to the jounce bumper configuration included redesigning the shock dust tubes to allow the largest jounce bumper diameter that could be packaged. To increase durability of the jounce bumper, some containment was found to be necessary. This increased impact hardness and bumper entry feel. A five percent less dense bumper material regained previous impact feel.

The final jounce bumper package was bench tested, and the rate curve combined with the final shock valving and free travel to predict shock tower loads, which were confirmed to be acceptable. Final validation was done by installing the new components on the loads measurement vehicle and checking shock tower load for the severe pothole, which was confirmed to be below the target value.

ANTI-ROLL BARS AND INSULATORS - Anti-roll bars were selected to achieve the target vehicle roll gain and TLLTD. TLLTD influences the nonlinear range understeer balance, which on C5 was more constant and predictable than most comparison vehicles because of the careful synthesis and design efforts to maintain SDF linearity [Ref. 2].

Rear Anti-Roll Bar - The original nominal tuning of the rear anti-roll bar was used during early tire construction development. During jounce bumper tuning the size was decreased to balance the vehicle for limit handling maneuvers. After the jounce bumper tuning was completed, the original size was re-instated. This size was used for the majority of shock valving and tire construction development. The production supplier was selected and available tube sizes identified.

Final nonlinear range understeer balance was developed by evaluating production available bar sizes with different durometer insulators.

Front Anti-Roll Bar - The front anti-roll bar did not change during the chassis development sequence, except to accommodate tube sizes available for production. This is just one of many examples that reflect the accuracy of the target setting and synthesis processes used on C5. On-center performance was enhanced with relatively stiff front bar insulators.

TIRES - Extended Mobility Tires (EMT) were adopted very early for the C5. A spare tire is neither needed nor available, and there are no optional tires. The tire mold shape, tread pattern, construction and compound were developed specifically for the C5. The result was the Goodyear Eagle F1 GS tire.

Tire specifications that were addressed in design and development included targets for cornering coefficient, aligning torque, wet and dry traction, mass,

rolling resistance, uniformity, durability, air borne noise, and many others.

Development of the tire's handling characteristics included subjective evaluations of transient and steady state handling characteristics such as: linearity, recovery, cornering power, on-center feel, steering sensitivity and many others. Evaluation rides were made in both dry and controlled wet environments at the Goodyear San Angelo Proving Ground. Subjective ride evaluations were made for ride characteristics such as impact isolation, impact noise, damping and tread noise at the GM Proving Ground.

The development process typically evaluated several tire submissions at one time on a C5 development vehicle. The evaluation information from the subjective rides was sent back to the Goodyear engineers to aid in designing the next round of tire submissions. In-between submissions, the best tire from each round was retained by the C5 development team for further suspension development work. This iterative development process for the tires was critical in the final optimization of the C5 suspension.

SHOCK ABSORBER MOUNTS - Upper and lower shock absorber mounts (bushings) were identical for the FE1 and FE3 suspensions, and unique for the F45 shocks (a different design and supplier). The series spring rate of both the upper and the lower shock mounts affect wheel control. A relatively high radial rate for the lower bushing was developed. The lower bushing conical rate was minimized to reduce bending loads in the shock during fore-aft suspension recession. This allowed the axial rate of the upper mount to be kept low for good coarse road isolation without sacrificing wheel control. As with the lower bushing, the conical rate had to be managed to minimize shock bending loads.

SHOCK ABSORBER VALVING - Each of the three suspension levels has unique shock absorber design and valving. This section describes the FE1 and FE3 suspensions.

The FE1 suspension uses a monotube design with a 36 mm piston. The FE3 suspension also uses a monotube design with a 45 mm piston. Both FE1 and FE3 shock absorber valvings were developed concurrently. Early valving iterations were unrefined but of adequate control to make the vehicle useful for other development work. Once the jounce bumper development was finalized, shock valving began in earnest.

Force-velocity curves were measured from the 1996 C4 FE1 and FE3 shock absorbers. The curves were adjusted for differences in wheel/shock travel ratio, to yield similar body and wheel control as C4. Different valvings were measured on a shock dynamometer until the target damping curves were matched. Shock valving development was based primarily on subjective rides that totaled over 175 for both FE1 and FE3.

Particular attention was paid to reducing quiick abrupt motions and developing them into controlled, fluid body motions. This not only improved the ride qualities but also made the vehicle handle better on the race track by more progressively managing normal tire loads when encountering bump disturbances such as apex curbing and when cornering through elevation changes. Keeping the rear tires "hooked up" when accelerating out of corners was also enhanced. A sufficient level of compression damping was developed to force the very stiff sidewalls of the EMT tires to envelop bumps. Digressive valving was utilized to keep from being over-controlled in high velocity events. Force-velocity curves comparing FE1 and FE3 damping are shown in Figure 7.

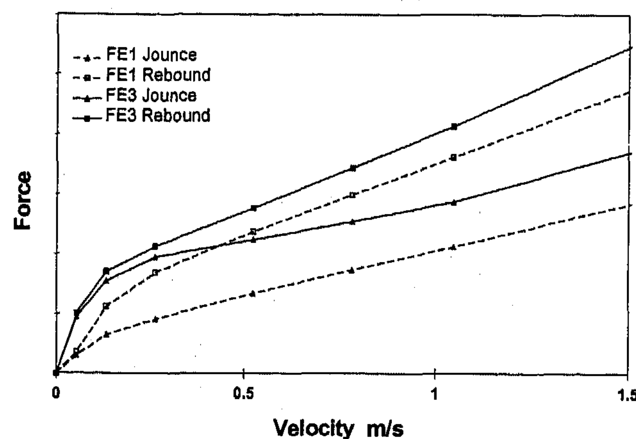


Figure 7. Typical FE1 and FE3 Shock Absorber Damping Characteristics.

REAL TIME DAMPING - An optional suspension for the C5 is the F45 or Real Time Damping (RTD - a system developed by Delphi Chassis of General Motors). RTD measures body and wheel motions, vehicle speed, lateral acceleration, steering wheel angle and velocity and electronically adjusts a continuously variable valve that changes the damping in each shock absorber independently. The system is extremely fast and is capable of controlling quick events such as chatter bumps.

The system will adjust several times during a single event. For instance: the body motion algorithm adjusts the damping to reduce or eliminate vertical motion of the sprung mass while driving through a negative swell. It does this by releasing rebound shock damping as the wheel falls into the swell so that the body will not be tugged down. The system then stiffens compression damping to hold the body up while the wheel travels the bottom of the swell. As the wheel rises up out of the swell, compression damping is released so that the body will not be pitched up. Finally, both rebound and compression valving are moderately stiffened to damp any further relative motion between the sprung and unsprung masses. This algorithm tightly controls heave, pitch and roll body motions. There are also algorithms that monitor lateral and fore-aft acceleration. Some of these anticipate body motions for

various road induced or driver commanded events (i.e. a severe lane change) and adjust damping accordingly.

The system also has a three position driver input switch with "Touring"," "Sport" and "Performance" modes. The "Touring" mode was developed to isolate the body from road contours as discussed in the previous paragraph. The "Performance" mode was developed to follow road contours as with highly damped passive shocks. A compromise middle ground was developed for the "Sport" mode.

Refining RTD involved traditional valving development, plus calibrating the many algorithms that control all the various ride and handling events.

STEERING - A single steering system is used for all suspension levels. C5 was designed from the beginning to use the Magnasteer ™ II variable effort power steering system (a product of Delphi Saginaw of General Motors).

This system was selected because it provided smooth, continuous tuning flexibility. It had the widest effort range of any variable effort system evaluated. It also provided tuning capability with vehicle lateral acceleration.

Reducing steering system friction was very important to meeting the steering system performance goals. This was accomplished two ways. First, the suspension was designed to minimize the ball joint loads and as a result minimize friction. Ball joint loads were minimized by reducing kingpin inclination, scrub radius, and spindle length [Ref 2]. Second, component friction was managed by setting friction specifications for all moving elements of the complete steering system, including ball joints, steering gear and steering column.

Steering ratio was determined by the Steering Sensitivity specification [Ref. 1], fore-aft rack packaging and steer arm length.

Several power steering valve profiles and torsion bar sizes were synthesized from the Steering Torque Gradient and Parking Effort requirements. The valve profiles were subjectively evaluated for on-center performance including linearity. During development, the valve profile was optimized first without the enhancements from Magnasteer™ II. This was followed by calibration of the variable assist curves.

Magnasteer™ II is an enhancement of Magnasteer™ I, but also calculates lateral acceleration from steering wheel angle and vehicle speed to adjust effort as a function of lateral acceleration. (Magnasteer™ I uses an electromagnet to manipulate the torsional rate of the power steering spool valve, and adjusts steering effort as a function of vehicle speed.) Magnasteer™ II allows the vehicle to be tuned for extremely smooth and linear steering feel with increasing lateral acceleration. The steer effort versus vehicle speed and lateral acceleration curves are shown in Figures 8 and 9.

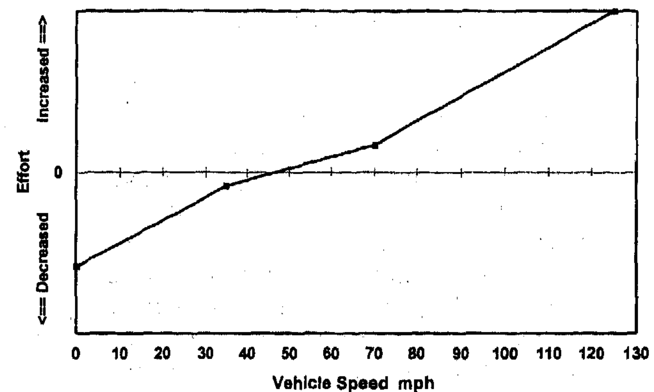


Figure 8. Magnasteer™ II Steer Effort Contribution vs. Vehicle Speed

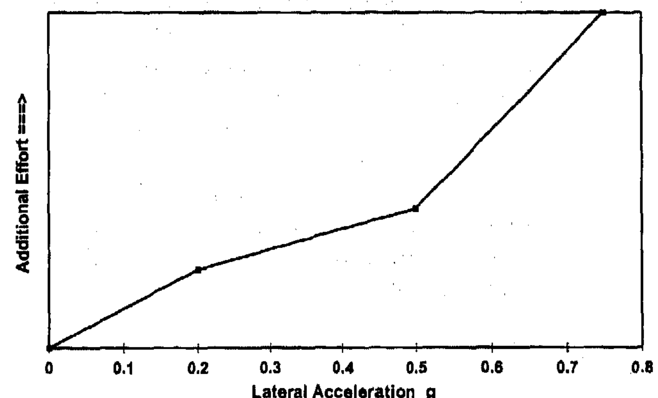


Figure 9. Magnasteer™ II Steer Effort Contribution vs. Lateral Acceleration.

ALIGNMENT - The C5 suspension geometry was uses more static negative camber and more positive caster than C4. The static alignment was evaluated at several points during the development process. It was finalized after the tire construction was fully developed, front camber, caster and toe and rear camber and toe were optimized for on-center steering performance, responsiveness, linearity, cornering power, tread wear etc. Consideration was made for negative camber adjustment capability for race track and gymkhana handling set-ups.

MEASUREMENTS - Two types of measurements were taken periodically to verify the design and to aid in the development process. Suspension kinematics, ride and roll rates, lateral and fore-aft compliances were measured at GM's Vehicle Handling Facility. Test results were compared to the target SDFs and used to validate the design [Ref. 2] and provide guidance for chassis development.

Total vehicle handling characteristics such as Steering Sensitivity, Lateral Acceleration Response Time, Yaw Velocity Damping, Linear Range Understeer, etc. [Ref. 1] were measured by GM's Vehicle Dynamics

Laboratory. Test data was compared to the Ride and Handling section of the Vehicle Technical Specification for validation and development guidance.

ASSESSMENT - The final activity of the development process was to close the loop in validating vehicle performance described in the Ride and Handling section of the Vehicle Technical Specification [Ref.1].

Objective measures and subjective evaluations were used to assess the vehicle performance. They confirmed that the production C5 met its objectives of high level handling characteristics and ride quality.

On-Center steering performance is more precise than C4. Handling is more nimble and linear, which translates into a very predictable vehicle. Limit handling is extremely forgiving with less tractive force steer and improved drop throttle characteristics. Maximum lateral acceleration is increased significantly.

Ride quality is enormously enhanced over the C4 and comparable to many sedans. This is a remarkable achievement considering the relatively stiff nature of the standard equipment EMT tires. Impact hardness is much better and structural memory shake is almost entirely eliminated. Steering wheel kick is negligible. Coarse road noise is noticeably less audible and body motions are less disturbing. There is an overall plushness not found in sports cars with the high level handling capabilities of the Corvette.

CONCLUSION

The 1997 Chevrolet Corvette's highly optimized yet simple suspensions, combined with other key vehicle subsystems and components (body structure, crossmembers, steering system and tires) produced significant improvements in handling, especially overall balance, integration, and predictability, and large gains in ride and isolation.

An exhaustive team effort produced the ride and handling related chassis components and systems introduced on the 1997 Chevrolet Corvette. Initially, vehicle level specifications based on customer input and guidance from recognized experts set the direction for chassis design. Computer simulations of vehicle and subsystem performance plus expert advice established design requirements. Early mule vehicle feedback provided valuable information on concept feasibility and complex interactions. The use of premium and innovative materials and components allowed design solutions to be executed in the hardware.

Development engineers then tuned the whole package into a pleasing, capable vehicle for the customer. The vehicle engineering loop was closed by validating vehicle performance with measured objective metrics that were compared to the Ride and Handling section of the Vehicle Technical Specification.

Above all, it was Management vision and leadership in partnership with Engineering that cultivated the people and resources necessary to produce the all-new suspension, steering, and tire designs unique to C5. The result was a no-compromise balance of ride and handling which extended and expanded upon the Corvette's unique heritage [Ref. 2].

ACKNOWLEDGMENTS

The development of the C5 suspension and steering system was the result of an exceptional team effort by a group of individuals too numerous to mention. However, the authors would be remiss if they failed to acknowledge the contributions of the following individuals: Paul Cavanaugh, Darin Dellinger, Brian Deutschel, Steve Fuja, John Heinricy, Jim Lloyd, Dave McLain, Jeff Roethlisberger, Joe Ryan, Hank Schmid, Rex Stump, Dick Topping, Kyle Tucker, and Dave Wickman.

REFERENCES

[1] Ryan, J. P., Fuja, S. P., Schmid, H. A.., "Objective Ride and Handling Goals for the 1997 Chevrolet Corvette", SAE 970091.

[2] Fuja, S. P., Schmid, H. A., Ryan, J. P., "Synthesis of Chassis Parameters for Ride and Handling on the 1997 Chevrolet Corvette", SAE 970097

[3] Schmid, H. A., Ryan, J. P., Fuja, S. P., "Design Synthesis of Suspension Architecture for the 1997 Chevrolet Corvette", SAE 970092.

[4] Longo, S. P., Moss, E. D., Deutschel, B. W., "The 1997 Chevrolet Corvette Structure Architecture Synthesis," SAE 970089.

[5] Brown, K., Juras, P., "The 1997 Chevrolet Corvette Suspension Crossmember," SAE 970370.

ABOUT THE AUTHOR

Mike Neal received his BSME degree from the University of Illinois in 1981. He has worked exclusively for General Motors since then in various areas of vehicle engineering. As a vehicle development engineer since 1986, he has been responsible for ride and handling development for several domestic and international rear wheel drive and front wheel drive programs. He has been the lead chassis development engineer for the 1997 Chevrolet Corvette. He also has designed, built, and raced formula cars, sportsracers, and production based cars since 1984.

970101

Physical Modeling of the Hysteretic Behaviour of Automotive Shock Absorbers

Stefaan W. Duym
Vrije Universiteit Brussel

Randy Stiens
LMS International

Gino V. Baron
Vrije Universiteit Brussel

Koenraad G. Reybrouck
Monroe Europe

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A physical model is proposed to account for the hysteretic effects of automotive shock absorbers. Besides the oil compressibility, the compressibility of a variable gas phase in the shape of bubbles are the root of these dynamic effects. Due to the pressure dependent solubility of gas in the liquid the gas bubbles appear or disappear thus changing the compressibility characteristics.

The model is used to identify the internal architecture of the damper, i.e. the valve parameters, and the hysteretic effects simultaneously. Some justifiable assumptions are introduced to estimate directly from damper force, piston displacement and velocity measurements.

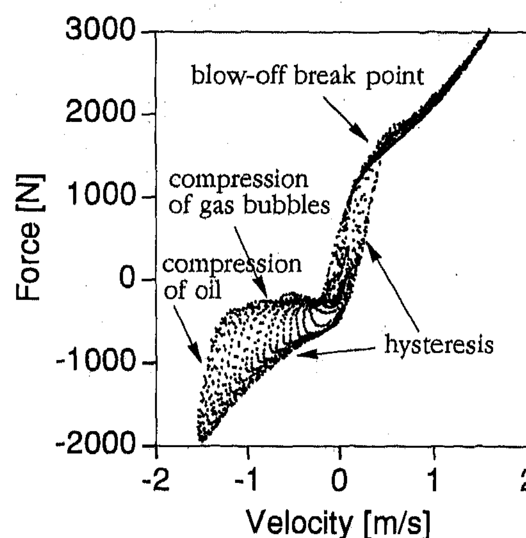
1 INTRODUCTION

The dampers are characterized by three facets. Firstly, when the characteristic diagram of a damper is observed, the force turns out to be a nonlinear function of velocity, due to the nonlinear pressure drop across the valve assemblies as a function of flow [1]. The two most important nonlinear effects are the bilinear character [2], that makes the damper force in the compression phase (negative damper velocities) smaller than in rebound phase (positive damper velocities), and the softening effect that appears from a certain break point, determined by the velocity, at which pressure-controlled blow-off valves open (Figure 1).

A second aspect is the oil temperature dependence of the damper force [3]. The force is determined by the oil viscosity, which, in turn, is strongly influenced by the oil temperature. Due to viscous losses, the oil temperature increases and consequently changes the damper characteristic. For short excitations (typically less than 1 minute), this effect is negligible.

Finally, the appearance of hysteresis loops, especially for higher excitation frequencies, is a third facet [4]. These are partly caused by the compressibility of the damper oil and to less extent by the compliance of the cylinder walls. These phenomena are well documented in the literature but the influence of gas bubbles in the oil is modelled inadequately.

Figure 1: Characteristic diagram



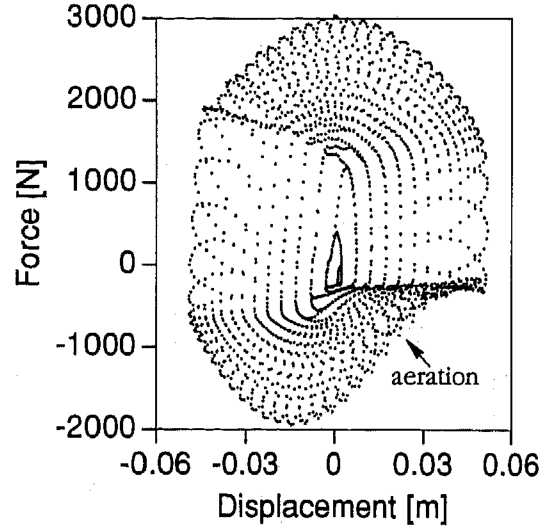
Lang [5] was the first to model the bubbles with some heuristic rules as an evaporation of damper oil, known as cavitation. However, Segel and Lang concluded that a more probable hypothesis is to assume that the bubbles are not formed from vapour but from nitrogen gas, emanating from the reserve chamber [6]. This effect

is known as frothing. Morman [7] elaborated their work and implemented the gas bubbles as a gas volume that was alternatively compressed and expanded according to the isentropic law.

This did not explain the sudden appearance or disappearance of the bubbles observed when experiments were performed on transparent dampers. This lacuna is partly studied in this paper. The two basic ideas are the transfer of bubbles between the three chambers and the pressure dependent solubility of the gas in the damper oil. When the pressure increases, the solubility increases and the bubbles disappear because the gas dissolves in the liquid and vice versa. An identification procedure that estimates this effect directly from force and velocity measurements is demonstrated.

The 'aeration' effects are also observed as asymmetries on the measured work diagram (Figure 2). The characteristic diagram (Figure 1) and work diagram were measured on a Mc-Pherson strut typically used on the front suspension of a standard Porsche 911. A modulated signal consisting of the sum of two sines with the same amplitude and close frequencies is used as excitation signal.

Figure 2: Work diagram



2 MODELING DAMPERS

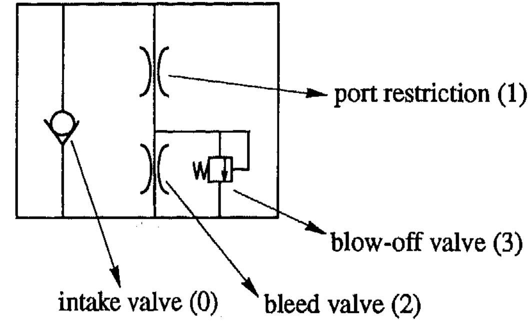
2.1 Valve modeling

2.1.1 The total valve characteristic

The dynamics of the damper force is governed by the time history of the internal pressure drops, which are caused by valve assemblies. In Figure 3 a valve assembly is shown schematically.

An intake valve (0) or check valve offers little resistance to flow in one direction but restricts flow in the opposite direction. The port restriction (1) is a pre-orifice placed in series with two valves: a bleed valve (2), typically realized with notches, and a blow-off valve (3). The bleed valve or initial orifice provides a highly restricted flow path between two chambers. The blow-off valve, also called control valve, is a spring-controlled valve that only opens at a specified pressure differential across the valve.

Figure 3: Valve assembly (schematic)



The bleed valve and the port restriction are passive elements that in the turbulent regime cause a pressure drop that is almost quadratic as a function of flow rate [1]:

Restriction characteristics:

$$\begin{cases} \Delta p_1 = K_1 Q_1^{\frac{7}{4}} & \text{(a)} \\ \Delta p_2 = K_2 Q_2^{\frac{7}{4}} & \text{(b)} \end{cases} \quad (1)$$

Control valves are more complex because they consist of moving elements. Lang worked out a nonlinear dynamical model that exhibits an extensive amount of hysteresis even when the valve inertia and damping are neglected due to the friction effects [5]. In this paper a simplified characteristic derived from the former is used [8]:

Static blow-off characteristic:

$$Q_3 = G(\Delta p_3 - \Delta p_0)\sqrt{\Delta p_3} \quad (2)$$

For $\Delta p_3 > \Delta p_0$, G is the blow-off conductivity factor. For $\Delta p_3 \leq \Delta p_0$, the blow-off valve remains closed and G is zero.

For the combination of the three different valves, the laws for parallel (3) and series (4) flow are used. For parallel flow, the pressure drops are the same but the respective flow rates are added to obtain the total flow. For series flow, the two respective flows are equal but the pressure drops are added to obtain the total pressure drop.

Parallel flow: $\begin{cases} \Delta p_2 = \Delta p_3 \\ Q_2 + Q_3 = Q_{tot} \end{cases}$ (3)

Series flow: $\begin{cases} \Delta p_{tot} = \Delta p_1 + \Delta p_2 \\ Q_1 = Q_{tot} \end{cases}$ (4)

The total valve characteristic is described by equations (5) and (6) where the flow rates Q_1 and Q_3 were eliminated by combining (1)-(4):

$$Q_{tot} = Q_2 + G\left(K_2 Q_2^{\frac{7}{4}} - \Delta p_0\right)\sqrt{K_2} Q_2^{\frac{7}{8}} = f_Q(Q_2) \quad (5)$$

$$\Delta p_{tot} = K_1 Q_{tot}^{\frac{7}{4}} + K_2 Q_2^{\frac{7}{4}} \quad (6)$$

For a pressure drop Δp_{tot} larger than the blow-off pressure Δp_0 this set of equations is implicit in both variables because no exact analytical solution exists. In order to recover Δp_{tot} when Q_{tot} is given, a lot of numerical algorithms can be used [9]. Nevertheless, an approximate analytical solution is presented to lower the computation time.

2.1.2 An explicit total valve model

For a pressure drop Δp_{tot} larger than the blow-off pressure Δp_0 the right hand side of equation (5) is expanded into a Taylor series around the blow-off flow rate Q_{20}:

$$Q_{tot} \approx \sum_{i=0}^{D} \frac{\partial^{(i)}}{\partial Q_2^{(i)}} f_Q(Q_2)\bigg|_{Q_2 = Q_{20}} (Q_2 - Q_{20})^i \quad (7)$$

$$\text{with } Q_{20} = \left(\frac{\Delta p_0}{K_2}\right)^{\frac{4}{7}} \quad (8)$$

A second order expansion (D=2) can easily be inversed and given Q_{tot}, Q_2 is calculated, viz:

$$Q_2 = Q_{20} + \frac{\sqrt{A_1^2 + 4A_2(Q_{tot} - Q_{20})} - A_1}{2A_2} \quad (9)$$

$$\text{with } \begin{cases} A_0 = f_Q(Q_{20}) = Q_{20} \\ A_1 = \frac{\partial}{\partial Q_2} f_Q(Q_2)\bigg|_{Q_2 = Q_{20}} = 1 + \frac{7\Delta p_0^{\frac{3}{2}} G}{4Q_{20}} \\ A_2 = \frac{\partial^2}{\partial Q_2^2} f_Q(Q_2)\bigg|_{Q_2 = Q_{20}} = \frac{35\Delta p_0^{\frac{3}{2}} G}{16 Q_{20}^2} \end{cases} \quad (10)$$

and finally Δp_{tot} is obtained from expression (6). The positive root (9) is chosen to obtain a positive real value according to the physical constraints. Higher order approximations can be used [8]. This explicit total valve characteristic will be used in the identification procedure.

2.1.3 An explicit inverse total valve model

For purpose of simulation, an inverse total valve model is needed where the flow rate Q_{tot} is calculated as a function of the total pressure drop Δp_{tot}. An approximate analytical solution is introduced similar to the one presented in section 2.1.2. The total pressure drop is written as a function of Δp_2 by combining equations (1)-(4):

$$\Delta p_{tot} = K_1\left[\left(\frac{\Delta p_2}{K_2}\right)^{\frac{4}{7}} + G(\Delta p_2 - \Delta p_0)\sqrt{\Delta p_2}\right]^{\frac{7}{4}} + \Delta p_2 = f_\Delta(\Delta p_2) \quad (11)$$

The right hand side of this equation is developed into a Taylor series around the blow-off pressure Δp_0 :

$$\Delta p_{tot} \approx \sum_{i=0}^{D} B_i (\Delta p_2 - \Delta p_0)^i \quad (12)$$

$$\text{with } B_i = \frac{\partial^{(i)}}{\partial \Delta p_2^{(i)}} f_\Delta(\Delta p_2)\bigg|_{\Delta p_2 = \Delta p_0} \quad (13)$$

A second order series is inversed to obtain the solution for Δp_2 as a function of Δp_{tot}:

$$\Delta p_2 = \Delta p_0 - \frac{B_1 - \sqrt{B_1^2 + 4B_2(\Delta p_{tot} - B_0)}}{2B_2} \text{ with} \quad (14)$$

$$\begin{cases} B_0 = \Delta p_0\left(1 + \frac{K_1}{K_2}\right) \\ B_1 = \left(1 + \frac{K_1}{K_2}\right) + \frac{7}{4} G K_1 \sqrt{\Delta p_0}\left(\frac{\Delta p_0}{K_2}\right)^{\frac{3}{7}} \\ B_2 = \frac{G K_1}{32\sqrt{\Delta p_0}}\left(\frac{\Delta p_0}{K_2}\right)^{\frac{3}{7}}\left[52 + 21 G K_2 \sqrt{\Delta p_0}\left(\frac{\Delta p_0}{K_2}\right)^{\frac{3}{7}}\right] \end{cases} \quad (15)$$

Once Δp_2 is calculated, Q_{tot} is obtained by combining equations (1) and (4):

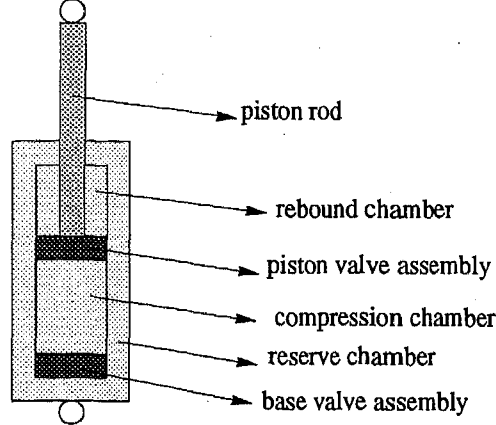
$$Q_{tot} = \left(\frac{\Delta p_{tot} - \Delta p_2}{K_1}\right)^{\frac{4}{7}} \quad (16)$$

2.2 Damper dynamics

The dual tube shock absorber consists of three chambers (Figure 4). In the internal tube, the portion of the cylinder above the piston is known as the rebound

chamber, and the portion of the cylinder below the piston is known as the compression chamber. The space between the outer and inner tube is called the reserve chamber. The reserve chamber is partly filled with nitrogen gas to compensate the volumetric changes, engendered by the inward/outward motion of the piston rod.

Figure 4: Dual tube shock absorber (schematic)



Every chamber with volume V is filled with an oil mass m_o, a bubble gas mass m_{gb} and within the reserve chamber an extra homogeneous gas volume V_{ge}. Using the density of oil ρ and the ideal gas law, the chamber volume is obtained, viz:

$$V = \frac{m_o}{\rho} + \frac{m_{gb}RT}{p} + V_{ge} \qquad (17)$$

The bubbles are compressed and expanded under isotherm conditions (T=cste) because due to the very small bubble diameter any heat produced inside the bubbles is transferred to the surrounding liquid instantaneously. However, the expansion and compression of the large homogeneous gas volume are better modelled under adiabatic conditions which are described by the isentropic law (with $\gamma = 1.4$):

$$V_{ge} = \sqrt[\gamma]{\frac{p_{initial}}{p}} V_{initial} = \frac{\xi}{\sqrt[\gamma]{p}} \qquad (18)$$

$$\text{with } \xi = p_{initial}^{1/\gamma} V_{initial} \qquad (19)$$

The compressibility of the oil makes the density dependent on the chamber pressure p. The relative change in volume is considered to be proportional to the applied pressure within a range up to a pressure of 10^7Pa:

$$\frac{V_0 - V}{V_0} = \alpha p \qquad (20)$$

Consequently, the actual density is calculated from:

$$\rho = \frac{\rho_0}{1 - \alpha p} \qquad (21)$$

For sake of convenience, the ratio r between gas bubble mass and oil mass is introduced by

$$r = \frac{m_{gb}}{m_o} \qquad (22)$$

Substitution of (18), (21) and (22) into (17), results in

$$V = \left[\frac{(1-\alpha p)}{\rho_0} + \frac{rRT}{p}\right] m_o + \frac{\xi}{\sqrt[\gamma]{p}} \qquad (23)$$

Derivation of (23) with respect to time, gives the rate of change in volume:

$$\dot{V} = Q - \left(\frac{\alpha}{\rho_0} + \frac{rRT}{p^2}\right)\dot{p} m_o - \frac{\dot{r}RT}{p} m_o - \frac{\xi \dot{p}}{\gamma \sqrt[\gamma]{p^{\gamma+1}}} \qquad (24)$$

Some dampers are designed through the use of baffles or membranes in such a way that no gas from the homogeneous gas phase is drawn into the inner cylinder, i.e. ξ is constant. Furthermore, the changes in chamber temperature T are assumed to be negligible. Q is defined as the net mass flow in the chamber:

$$Q = \left[\frac{(1-\alpha p)}{\rho_0} + \frac{rRT}{p}\right] \dot{m}_o \qquad (25)$$

The fundamental difference for the definition of the flow in comparison with the models of Lang [5] and Morman [7] is the fact that bubbles are also included in the flow and transferred between chambers. This assumption is sustained by Stokes' formula, which is explained later.

For the rebound chamber e.g., Q is the flow $Q|^r_{r \to c}$ between the compression chamber and the rebound chamber and is dependent on the two respective chamber pressures (see former section). Due to the effect of expansion the flow $Q|^c_{r \to c}$ that penetrates the compression chamber and originally drawn from the rebound chamber will be different from $Q|^r_{r \to c}$. If the conservation law of mass is expressed we obtain:

$$\dot{m}_o = \frac{Q|^r_{r \to c}}{\frac{1-\alpha p_r}{\rho_0} + \frac{rRT}{p_r}} = \frac{Q|^c_{r \to c}}{\frac{1-\alpha p_c}{\rho_0} + \frac{rRT}{p_c}} \qquad (26)$$

A similar relationship is found for the flow between compression chamber and reserve chamber:

$$\frac{Q|^s_{s \to c}}{\frac{1-\alpha p_s}{\rho_0} + \frac{rRT}{p_s}} = \frac{Q|^c_{s \to c}}{\frac{1-\alpha p_c}{\rho_0} + \frac{rRT}{p_c}} \qquad (27)$$

For sake of convention the flow calculated from the flow-pressure characteristics (section 2.1.3) will be on the side of the chamber from where the liquid originates.

The second term in (24) is due to the compressibility of the gas and oil mixture. It was shown that this term can

also account for the compliance of the cylinder walls [5], [10].

The third term is due to the solubility of the gas in the liquid, which is described by Henry's law [11]:

$$p = H\frac{m_{gs}}{V_{oil}} \tag{28}$$

where H is a function of the mixture and temperature. Since both are assumed to be given, H is further considered to be a constant. m_{gs} is the mass of gas in solution. The solved gas mass is calculated from (28) via substitution of the oil volume with the respective density and oil mass as follows

$$m_{gs} = \frac{m_o p}{\rho_0 H} \tag{29}$$

The fraction r can thus be obtained from the total gas mass m_{gt} and the gas mass in solution m_{gs}

$$r = \frac{m_{gt}-m_{gs}}{m_o} = r_t - \frac{p}{H\rho_0} \tag{30}$$

The total fraction r_t is assumed to be constant because this fraction can only change with a convection stream of bubbles relative to the fluid. Experiments with transparent absorbers have shown that the bubbles have a negligible velocity relative to the fluid. Stokes' formula [11] gives a rough estimate of the relative velocity of the bubbles in the oil:

$$v = \frac{(\rho_{oil}-\rho_{gas})\,d^2 g}{18\mu} \tag{31}$$

This formula emanates from the balance between the Archimedes force, due to the difference in densities, and the viscous friction force. If typical values of ρ_{oil}= 935kg/m^3 (density of oil), ρ_{gas}= 1.3 kg/m^3 (density of gas), d = 1e-4m (diameter of gas bubble), g = 9.81 m/s^2 (acceleration of gravity), μ = 0.037kg/ms (dynamic oil viscosity) are used for bubbles whose diameter is less than 1e-4m the relative speed of the bubbles is approximately 0.001m/s. Compared with the flow speed of the oil (above 0.01m/s), this is, indeed, negligible.

From (30), it is obvious that above a critical pressure p_k all bubbles disappear (r=0), which means that all gas remains in the solution, with:

$$p_k = r_t H\rho_0 \tag{32}$$

By substituting (30) in (23) and solving for m_o we get

$$m_o = \frac{V-\frac{\xi}{\gamma\sqrt{p}}}{\frac{(1-\alpha p)}{\rho_0}+\frac{r_t RT}{p}-\frac{RT}{H\rho_0}} \tag{33}$$

Substitution of (30) in (24), leads to

$$\dot{V} = Q-\left(\frac{\alpha}{\rho_0}+\frac{r_t RT}{p^2}\right)\dot{p}m_o-\frac{\xi\dot{p}}{\gamma\gamma\sqrt{p^{\gamma+1}}} \tag{34}$$

Substitution of (33) into (34) yields:

$$\dot{V} = Q-\frac{\left(V-\frac{\xi}{\gamma\sqrt{p}}\right)\left(\frac{\alpha}{\rho_0}+\frac{r_t RT}{p^2}\right)\dot{p}}{\frac{(1-\alpha p)}{\rho_0}+\frac{r_t RT}{p}-\frac{RT}{H\rho_0}}-\frac{\xi\dot{p}}{\gamma\gamma\sqrt{p^{\gamma+1}}} \tag{35}$$

The chamber volumes can also be obtained from the respective geometric relations for the rebound chamber volume:

$$V_r = (L-x)A_r \tag{36}$$

and the compression chamber volume:

$$V_c = (L+x)A_c \tag{37}$$

The volumes are dependent on the displacement x of the piston from the mid position where L is the half length of the inner cylinder. The cross area A_r of the rebound chamber is smaller than the cross area A_c of the compression chamber due to the presence of the piston rod with cross area A_d :

$$A_r = A_c - A_d \tag{38}$$

The volume V_s of the reserve chamber is a known constant value. The dynamics of the global system is described by a set of three differential equations that show how the pressures in the three individual chambers vary as a function of pressure, displacement and velocity. Combination of expressions (35)-(37) yields:

Rebound chamber: $(p_r < p_k)$

$$\dot{p}_r = \frac{[Q|^r_{c\to r}+\dot{x}A_r]}{(L-x)A_r}\frac{\left[\frac{(1-\alpha p_r)}{\rho_0}+\frac{r_t RT}{p_r}-\frac{RT}{H\rho_0}\right]}{\left(\frac{\alpha}{\rho_0}+\frac{r_t RT}{p_r^2}\right)} \tag{39}$$

Compression chamber: $(p_c < p_k)$

$$\dot{p}_c = \frac{[(Q|^c_{s\to c}-Q|^c_{c\to r})-\dot{x}A_c]}{(L+x)A_c}\frac{\left[\frac{(1-\alpha p_c)}{\rho_0}+\frac{r_t RT}{p_c}-\frac{RT}{H\rho_0}\right]}{\left(\frac{\alpha}{\rho_0}+\frac{r_t RT}{p_c^2}\right)} \tag{40}$$

Reserve Chamber:

$$\dot{p}_s = -\frac{Q|^s_{s\to c}}{\frac{\left(\frac{\alpha}{\rho_0}+\frac{r_t RT}{p_s^2}\right)\left(V_s-\frac{\xi}{\gamma\sqrt{p_s}}\right)}{\frac{(1-\alpha p_s)}{\rho_0}+\frac{r_t RT}{p_s}-\frac{RT}{H\rho_0}}+\frac{\xi}{\gamma\gamma\sqrt{p_s^{\gamma+1}}}} \tag{41}$$

After a number of strokes the bubbles and air in solution are homogeneously distributed in the three chambers therefore it is justified that the ratio r_t is the same in all chambers. After a sufficiently long rest period, the bubbles rise to the top of the rebound chamber and a large gas bubble is formed. If a new excitation is engaged with a slow rebound stroke, the gas will be evacuated through the highly restrictive space between the piston rod and rod guide into the reserve chamber. Before experiments are carried out, it is common to reset the damper with a slow sinusoidal displacement, which is referred to as the 'priming' sequence.

Except for the reserve chamber, ξ is zero. Equations (39)-(41) are formed for pressures below the critical pressure p_k. The reserve chamber is normally designed in such a way that higher pressures do not appear but this is not true for the other two chambers. For pressures higher than the critical pressure, equations (39)-(40) reduce to (42) and (43).

Rebound chamber: $(p_r \geq p_k)$

$$\dot{p}_r = \frac{[Q|^r_{c\rightarrow r} + \dot{x}A_r]}{(L-x)A_r}\frac{(1-\alpha p_r)}{\alpha} \tag{42}$$

Compression chamber: $(p_c \geq p_k)$

$$\dot{p}_c = \frac{[(Q|^c_{s\rightarrow c} - Q|^c_{c\rightarrow r}) - \dot{x}A_c]}{(L+x)A_c}\frac{(1-\alpha p_c)}{\alpha} \tag{43}$$

3 IDENTIFICATION

3.1 Assumptions

When the internal pressures are measured directly, it is possible to use the set of equations in the former section. When the identification is to be performed on force data, the problem turns out to be non-trivial. A possible approach is based on the use of parameter-sensitivity functions [12]. This method finds the parameters with a gradient method where the gradients are calculated from a related set of differential equations. In the present paper, in order to simplify the identification procedure, some assumptions are used as follows.

Firstly, it occurs that the pressure in the reserve chamber changes very little, compared with the two other. The homogeneous gas phase in the reserve chamber disables any large pressure variability and it is justified to assume that the pressure p_s remains constant. This pressure is determined from the static damper force F_0 at rest [8].

The force acting on the piston rod is calculated as:

$$F = A_r p_r - A_c p_c \tag{44}$$

At rest, the pressures in all chambers are the same and equal to the reserve chamber pressure. Substitution of $p_r = p_c = p_s$ into (44) leads to an expression for p_s:

$$p_s = \frac{F_0}{(A_r - A_c)} = -\frac{F_0}{A_d} \tag{45}$$

Since A_r is smaller than A_c, the static force is negative and it will slowly push the piston rod out of the damper body if no counter force is applied.

Secondly, during a compression stroke, fluid is forced to flow up from the compression chamber to the rebound chamber through the rebound intake valve in the piston head. During a rebound stroke, fluid from the reserve chamber is drawn through the compression intake valve in the base assembly to the compression chamber. These intake valves offer little resistance to flow but restrict flow in the opposite direction when the pressure differential is reversed. For convenience the respective pressure drops are set to zero for the appropriate phases, so that we get:

$$\begin{cases} p_c = p_s & \text{for} \quad F \geq F_0 \quad \text{(rebound)} \\ p_r = p_c & \text{for} \quad F \leq F_0 \quad \text{(compression)} \end{cases} \tag{46}$$

3.2 Simplified set of equations

The assumptions cited in section 3.1 reduce the number of independent pressure variables from three to one. The rebound chamber pressure p_r is the remaining independent state variable. The assumptions also leave undetermined $Q_{c\rightarrow r}$ during the compression stroke and $Q_{s\rightarrow c}$ during the rebound stroke. The set of three equations is thus rearranged so that these terms do not appear. For $(p < p_k)$, expressions (47) and (48) are used, instead, for $(p \geq p_k)$, expressions (49) and (50) are used.

Rebound phase: identical to (39)

$$\dot{p}_r = \frac{(Q|^r_{c\rightarrow r} + \dot{x}A_r)}{(L-x)A_r}\frac{\left(\frac{(1-\alpha p_r)}{\rho_0} + \frac{r_t RT}{p_r} - \frac{RT}{H\rho_0}\right)}{\left(\frac{\alpha}{\rho_0} + \frac{r_t RT}{p_r^{\,2}}\right)} \tag{47}$$

Compression phase: combining (39) and (40)

$$\dot{p}_r = \frac{-(Q|^c_{c\rightarrow s} + \dot{x}A_d)}{L(A_r + A_c) + xA_d}\frac{\left(\frac{(1-\alpha p_r)}{\rho_0} + \frac{r_t RT}{p_r} - \frac{RT}{H\rho_0}\right)}{\left(\frac{\alpha}{\rho_0} + \frac{r_t RT}{p_r^{\,2}}\right)} \tag{48}$$

Rebound phase: identical to (42)

$$\dot{p}_r = \frac{(Q|^r_{c\rightarrow r} + \dot{x}A_r)}{(L-x)A_r}\frac{(1-\alpha p_r)}{\alpha} \tag{49}$$

Compression phase: combining (42) and (43)

$$\dot{p}_r = \frac{-(Q|^c_{c\to s} + \dot{x}A_d)}{L(A_r + A_c) + xA_d}\frac{(1-\alpha p_r)}{\alpha} \qquad (50)$$

3.3 Identification procedure

The assumptions enable us to calculate the rebound pressure p_r directly from the measured damper force F:

$$p_r = \frac{F + A_c p_s}{A_r} \text{ for rebound} \qquad (51)$$

$$p_r = \frac{F}{A_r - A_c} \text{ for compression} \qquad (52)$$

This pressure is also defined as the measured rebound pressure $p_r|_{meas}$. The first derivative $\dot{p}_r$ can be obtained in time domain as well as in frequency domain since periodic excitations are applied.

Given the motion of the damper $(x, \dot{x})$ and the internal pressure and its first derivative $(p_r, \dot{p}_r)$, it is possible to get (a) the flow rate $Q|^r_{c\to r}$ between compression and rebound chamber during the rebound stroke for $(p_r < p_k)$ from (47) and (32):

$$Q|^r_{c\to r} = \frac{\left(\frac{\alpha}{\rho_0} + \frac{r_t RT}{p_r^2}\right)(L-x)A_r\dot{p}_r}{\frac{(1-\alpha p_r)}{\rho_0} + r_t RT\left(\frac{1}{p_r} - \frac{1}{p_k}\right)} - \dot{x}A_r \qquad (53)$$

and for $(p_r \geq p_k)$ from (49):

$$Q|^r_{c\to r} = \frac{\alpha(L-x)A_r\dot{p}_r}{(1-\alpha p_r)} - \dot{x}A_r \qquad (54)$$

and (b) the flow rate $Q|^c_{c\to s}$ between compression and reserve chamber during the compression stroke for $(p_r < p_k)$ from (48) and (32):

$$Q|^c_{c\to s} = -\frac{\left(\frac{\alpha}{\rho_0} + \frac{r_t RT}{p_r^2}\right)(L(A_r + A_c) + xA_d)\dot{p}_r}{\frac{(1-\alpha p_r)}{\rho_0} + r_t RT\left(\frac{1}{p_r} - \frac{1}{p_k}\right)} - \dot{x}A_d \qquad (55)$$

and for $(p_r \geq p_k)$ from (50):

$$Q|^c_{c\to s} = -\frac{\alpha(L(A_r + A_c) + xA_d)\dot{p}_r}{(1-\alpha p_r)} - \dot{x}A_d \qquad (56)$$

These flow rates lead us to the respective pressure drops in equations (6) and (9), and consequently to the modelled rebound pressure $p_r|_{mod}$. The rebound pressure is calculated as (a) the sum of p_s and the pressure drop over the base valve assembly during compression, and (b) the sum of p_s and the pressure drop over the piston valve assembly during rebound.

Each valve assembly is characterized by four independent parameters: K_1, K_2, G and Δp_0. The global dynamics are characterized by three parameters: α, r_t and p_k. The geometric parameters A_r, A_c, V_s and L, and the initial oil density ρ_0 are assumed to be known or given. An identification procedure estimates the three dynamic parameters and the two times four valve parameters in a least squares sense. The estimates are obtained from minimization of a cost function defined as

$$Z = \sum_{i=1}^{N} \frac{(p_r(t_i)|_{meas} - p_r(t_i)|_{mod})^2}{N} \qquad (57)$$

where the pressures are sampled at regular time instants $t_i = i/f_{sample}$ with N the number of samples and f_{sample} the sample frequency.

3.4 Estimate evaluation

The square root of Z is the Root Mean Square (RMS) residue between the measured and modelled values. A similar residue is calculated for the damper force as the square root of a cost function Z':

$$Z' = \sum_{i=1}^{N} \frac{(F(t_i)|_{meas} - F(t_i)|_{mod})^2}{N} \qquad (58)$$

The modelled damper force is obtained from inversion of equations (51) and (52), viz

$$F|_{mod} = A_r p_r|_{mod} - A_c p_s \text{ for rebound} \qquad (59)$$

$$F|_{mod} = -A_d p_r|_{mod} \text{ for compression} \qquad (60)$$

The residue originates from model errors, measurement noise and numerical errors. The variance σ^2 or level of measurement noise is estimated from M periods provided that the temperature effects are not important:

$$\hat{\sigma}^2 = \sum_{i=1}^{N}\sum_{j=0}^{M-1} \frac{\left(F(t_i + j\tau)|_{meas} - \overline{F(t_i)}|_{meas}\right)^2}{N(M-1)} \qquad (61)$$

$$\text{with } \overline{F(t_i)}|_{meas} = \sum_{k=0}^{M-1} \frac{F(t_i + k\tau)|_{meas}}{M} \qquad (62)$$

and $\tau = N/f_{sample}$ the excitation period. In the case that the uncertainty bound for the noise is equal to the model error the RMS residue will be equal to $\sqrt{2}\sigma$ in the absence of important numerical errors. The presence of numerical problems is detected by observing the condition number of matrix calculations during optimization.

For the final evaluation of the model the residue will be compared to $\sqrt{2}\sigma$. If the residue is much smaller than this value the identification will suffer from overmodeling which means that the model is probably modeling a part of the noise spectrum. Some model parameters are not estimated accurately due to masking by measurement noise. It is recommended to perform a new set of more accurate measurements (less corrupted with noise) or to remove the deficient parameters because they do not contribute to the model in a relevant way for the given experiments.

On the other hand if the residue is much larger than $\sqrt{2}\sigma$ this means that the model is probably too simple to accurately describe the measurement results. A more complex model is then recommended to overcome the deficient fit.

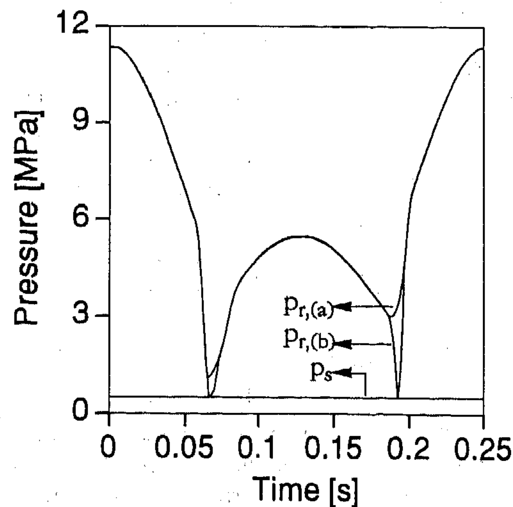
4 RESULTS

4.1 Simulation of internal pressures

In order to verify the effect of the assumptions, a damper is simulated (a) with the full set of equations (39)-(43) and (b) with the simplified set of equations (47)-(50). The set of ordinary differential equations is solved using fourth and fifth order Runge-Kutta formulas [13]. A sinusoidal excitation x is used with an amplitude of 5 cm and a frequency of 4Hz. The set of parameters is defined in Table 1 (appendix). The simulation results are shown in Figure 5.

If we compare the rebound pressure for (a) the full and (b) the simplified set we observe the largest discrepancy when the velocity of the piston approaches zero, i.e. at the time the damper switches from rebound to compression stroke or vice versa.

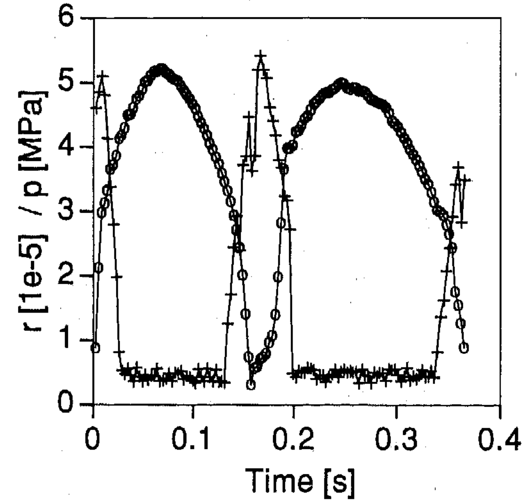
Figure 5: Simulated rebound pressures



4.2 Experimental gas fraction, r

Using an inverse calculation in [14] an approximation of the time history of the fraction r is obtained [(+) in (Figure 6)].

Figure 6: Experimental bubble gas fraction r(t)



Firstly, a similar model where the fraction r was assumed to be constant (i.e. H=∞) similar to Morman's model was fitted to measurements obtained from a beating signal. The modelled damper force did not explain certain effects that will be shown in a later section. Then it was assumed that all model parameters were exact and the fraction r was calculated to make the model match the experiments. A plot of the rebound pressure (o) and the resulting fraction are given in Figure 6. The fraction remains at a low level (around 5e-6) for a long portion of the excitation time but when the pressure decreases (close to zero piston velocity) the mass of gas bubbles increases and vice versa. This phenomenon is explained by Henry's law as will be shown in the next section. However the fraction r from the inverse calculation does not reach zero probably because the parameters were obtained from a fit where a constant time-averaged value for r was obtained. This probably introduced a bias on all parameters leading to the offset for the calculated fraction.

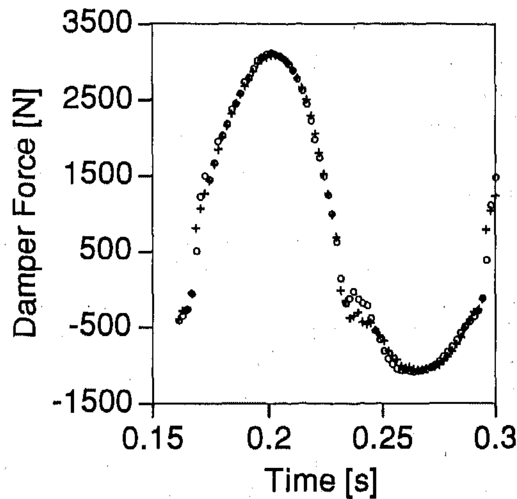
Furthermore it is also experienced from Figure 6 that the critical pressure at which the fraction reaches its minimum is different from the pressure at which the fraction rises from that minimum. For both rebound and compression strokes bubbles start appearing around 3MPa while the fraction only reaches its minimum around an effective pressure of 4MPa. This delay is probably due to the solving of bubbles that does not take place immediately as assumed by Henry's law due to the finite diameter of bubbles.

4.3 Identification results

A shock absorber was excited with a beating signal with an approximate excitation period of 4 seconds. Then measurements of damper force, piston displacement and velocity were processed to fit the parameters of the given model. The same test set-up as presented in [1] was used. A second model was fitted where the fraction r was assumed to be constant. This is equivalent to fitting the first model with p_k having a fixed infinite value ($p_k=\infty$). The cost function was then minimized using a Levenberg-Marquardt algorithm [9].

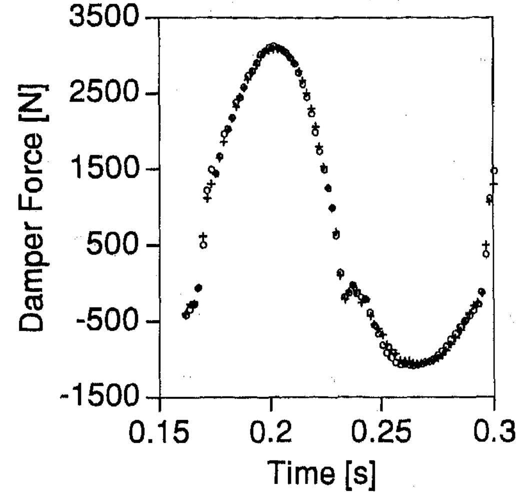
Detailed results of the measured force (o) and the force reconstructed from the ($p_k=\infty$)-model (+) are shown in Figure 7. The model fits satisfactorily when the force reaches its maximum and minimum values; but when the force gets close to zero a larger discrepancy is observed, especially around t=0.235s. This is not the case when the full model ($p_k<\infty$) is fitted (Figure 8).

Figure 7: Fit of model (r=cste) to measurements



In order to give an objective quality measure on both fits the standard deviation σ of the measurement noise is calculated according to (61) from 4 consecutive periods. It is assumed that temperature effects are negligible over this short period (4 x 4s = 16s). It was shown in [14] that the force experiences a linear trend of about 40N over this time. The estimate of σ equals **52N/sample**. The RMS residues for both models are respectively (a) for the (r=cste)-model **99N/sample** and (b) for the full model **82N/sample**. The improvement is not very important but this is due to the minor importance of this second order effect of the pressure dependent solubility described by Henry's law. The residue of the full model is close to the limit $\sqrt{2}\sigma$ **=73N/sample** which means that a good fit has been performed according to the discussion in section 3.4.

Figure 8: Fit of model (r≠cste) to measurements



The fitted and fixed parameters of the full model are given in Table 2 (appendix)

4.4 Discussion

The compressibility of the mixture of gas and fluid is determined by the factor

$$\text{compressibility} = \frac{\alpha}{\rho_0} + \frac{r_t RT}{p_r^2} \qquad (63)$$

Dependent on the rebound pressure, one of the two terms will be dominant compared to the other. Using the fitted parameters in Table 2 it is possible to determine the pressure at which both terms are equal:

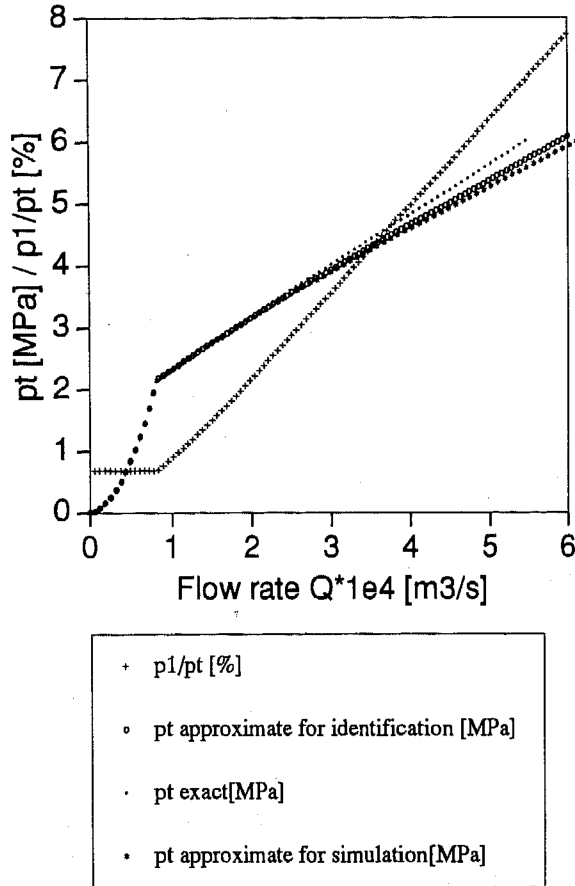
$$p_r = \sqrt{\frac{r_t RT\rho_0}{\alpha}} = 4.37MPa \qquad (64)$$

This is a rather high value due to mixing of oil and gas phase caused by the specific geometry and tuning of the tested strut. Below this pressure, it is unjustified not to take the influence of gas bubbles into account. Models [2], [3] and [10] that only use compressibility of fluid and possibly the elasticity of cylinder walls to account for hysteretic effects would induce important predictive errors for this absorber under test. However some dampers are found where the gas phase is separated from the fluid by a membrane [4] or a floating piston which is typical for a mono-tube shock absorber [1]. If care was taken to remove all dissolved air or bubbles in the fluid before manufacture of the damper, e.g. by heating the oil, these models will work. Models [15] and [16] that do not take oil compressibility into account have no realistic predictive power concerning damper hysteresis.

It is clear from Table 2 that the pressure drop over the leak restriction is much more important than the drop over the port restriction. A difference of two orders of magnitude for the restrictions of the rebound valve assembly as well as the compression valve assembly

suggests a possibility to neglect the port restriction.

Figure 9: Total valve characteristic of the piston



In Figure 9 the total valve characteristic of the identified piston assembly is shown with the tokens (.), (o) and (•), respectively, for the exact equations (5) and (6), the approximate equations (6) and (9) used for identification and the approximate equations (14) and (16) used for simulation. Above the blow-off pressure a quasi linear pressure drop as a function of flow rate is observed. The ratio of the port pressure drop to the total pressure drop $\Delta p_1/\Delta p_{tot}$ is presented with (+). Below the blow-off flow rate the ratio is indeed very small (less than 1%) but the ratio increases almost linearly above the blow-off break point. Consequently, for higher pressure drops neglecting the port restriction will induce larger simulation errors.

The relative error induced by the approximate valve equations is around 5% for a pressure drop of 4MPa and 12% for a drop of 6MPa. The pressure drop for a relative error of 10% lies about 5.3MPa. Both approximations give an underestimate of the actual pressure drop.

For the damper temperature, T=323K was used which is possibly not the real temperature because dampers easily experience an increase of 50K from the initial temperature [2]. Nevertheless, the identification can handle this problem because the two factors T and r_t appear coupled [equations (53) and (55)] and the latter is estimated. For instance, when the assumed temperature, used in the identification procedure, turns out to be lower than the real temperature the estimator will compensate this by a larger value of r_t and vice versa.

5 FUTURE DEVELOPMENTS

5.1 Valve modeling

A more complex model for the valve assemblies may improve the predictive power of the full model although it was argued in [7] that the valve inertia and linear damping used in [5] are almost undetectable. However it has been observed that the sharp blow-off edge (Figure 9) is often rounded [8], possibly because the blow-off valve starts leaking before the blow-off pressure Δp_0. An empirical transition coefficient was used in [1] to account for this effect. The blow-off valve in [5] is assumed to translate up and down during compression and rebound respectively. In reality a modern blow-off valve consists of a disk spring clamped on the inner side and opening by bending.

5.2 Friction

McPherson struts are dampers that allow side loading. When a large side load is applied to the strut, large values of friction are detected between the piston and the inner tube and between the rod guide and the damper rod. The rod guide and damper rod are highly finished because the largest values of friction appear between them due to the lever effect. For low velocities even sticktion (sticking friction) can appear. An elaborated model has been synthesized by Armstrong-Hélouvry [17] to account for Coulomb friction, Stribeck friction, frictional memory, dwell time, etc.

Nevertheless, when no side loading is applied, a simple model can accurately describe friction where the damper force is calculated viz:

$$F = A_r p_r - A_c p_c + F_c sign(\dot{x}) \qquad (65)$$

F_c is the Coulomb friction and typical values within the range of 0-100N have been recorded.

5.3 Bubble dynamics

Henry's law describes the effect of bubbles present in the fluid very well. Experiments with transparent dampers show that the bubbles are present in the form of a diffusion which means that the diameter of the bubbles is very small (typically less than 0.1mm). This is probably due to the large strokes that push the fluid

with bubbles through the narrow openings of the valve. However, for a small stroke larger bubbles are expected and this changes the value of the critical pressure. A delay for solving bubbles was also experienced in section 4.2. A model that accounts for any size of bubbles can possibly improve simulation results.

5.4 Temperature effects

The next most challenging aspect of damper modeling is certainly the modeling of temperature effects. A change in temperature affects the valve parameters due to a change in viscosity. The temperature dependent part of the valve parameters has been identified in [1]. If the inside temperature is measured, it should be possible to account for these effects but when the temperature is not measured directly, the problem becomes non-trivial and a predictive model for the oil temperature describing a heat balance must be used.

The heating source within the absorber is easily identified with the product of velocity and force. However a lumped heat capacitance approach is questionable and furthermore the problem of heat evacuation remains. Heat is evacuated by convection which is dependent on cooling, humidity and isolation of the damper. Finally for the identification of a model that includes temperature effects a more complex and longer excitation signal must be used.

5.5 Identification

It was shown in section 4.1 that the assumptions used for convenience of identification give rise to a discrepancy in particular when the piston velocity approaches zero. It is also at that time that the bubbles will induce important hysteresis as described by the law of Henry. The question as to how both phenomena correlate should be investigated by using the full set of three differential equations.

Secondly, the velocity and displacement are measured and thus corrupted with measurement noise. The use of the output error approach can cause bias errors on the parameters. An Error-in-Variables (EV) approach [18] where errors on all variables are accounted for may solve this problem. An EV estimator will be helpful especially around zero velocity where the model is more sensitive to a shift in velocity.

6 CONCLUSION

A model has been presented for the nonlinear and hysteretic behaviour of shock absorbers. The model has been identified using some convenient and justifiable assumptions. It has been shown that the hysteresis is caused by oil compressibility and more importantly by compressibility of present gas bubbles. For higher pressures the solubility of nitrogen in oil increases and the bubbles disappear. This effect seems to be important and cannot be neglected without introducing considerable errors. The presented model can easily be used either for identification or for simulation. The parameters are directly related to the internal architecture of the damper.

Acknowledgements

This research is sponsored by a grant from the Flemish Institute for the Improvement of the Scientific and Technological Research in Industry (IWT), the Flemish government (GOA-IMMI$_2$) and the Belgian government as a part of the Belgian program on InterUniversity Poles of Attraction (IUAP-17) initiated by the Belgian State, Prime Minister's Office, Science Policy Programming.

The authors wish to express their gratitude to ir. Xavier Lauwerys for his comments, ing. Hans Geukens and ing. Steven Juvens for the technical support.

References

[1] **Reybrouck K**, A Non Linear Parametric Model of an Automotive Shock Absorber, *SAE International Congress*, Detroit, Michigan, February 28-March 3, 79-86, 1994

[2] **Surace C, K Worden and GR Tomlinson**, An Improved Nonlinear Model for an Automotive Shock Absorber, *Nonlinear Dynamics* 3, 413-429, 1992

[3] **Surace C, D Storer and GR Tomlinson**, Characterising an Automotive Shock Absorber and the Dependency on the Temperature, Proceedings of the *10th International Modal Analysis Conference*, San Diego, 1317-1326, 1992

[4] **Audenino AL and G Belingardi**, Modelling the dynamic behaviour of a motorcycle damper, *Proceedings of the Institution of Mechanical Engineers, part D: Journal of Automobile Engineering* 209, 249-262, 1995

[5] **Lang HH**, A Study of the Characteristics of Automotive Hydraulic Dampers at High Stroking Frequencies, PhD Dissertation, The University of Michigan, 1977

[6] **Segel L and HH Lang**, The Mechanics of Automotive Hydraulic Dampers at High Stroking Frequencies, *Vehicle System Dynamics* 10, 82-85, 1981

[7] **Morman KN**, A Modelling and Identification Procedure for the Analysis and Simulation of Hydraulic Shock Absorber Performance, ASME *Winter Annual Meeting*, New Orleans, Louisiana, December 9-14, 1984

[8] **Duym S, R Stiens and K Reybrouck**, Fast Parametric and Nonparametric Identification of Shock Absorbers, Proceedings of the *21st International Seminar on Modal Analysis*, Noise and Vibration Engineering Conference, Leuven, Belgium, 1157-1169, 1996

[9] **Fletcher R**, *Practical Methods of Optimization*, John Wiley and Sons, 1980

[10] **Lang R and R Sonnenburg**, A Detailled Shock Absorber Model for Full Vehicle Ride Simulation, presented at the *10th European ADAMS Users' Conference*, Frankfurt, Germany, November 14-15, 1995

[11] **Coulson JM, JF Richardson, JR Backhurst and JH Harker**, *Chemical Engineering Vol 2*, 4th Edition, Pergamon Press Oxford, 1991

[12] **Eykhoff P**, *System Identification*, John Wiley & Sons, 1974

[13] **Forsythe GE, MA Malcolm and CB Moler**, *Computer Methods for Mathematical Computations*, Prentice-Hall, 1977

[14] **Stiens R**, *Parametrische Identificatie van Schokdempers*, Thesis burgerlijk werktuigkundig-electrotechnisch ingenieur, Vrije Universiteit Brussel, Mei, 1996 [in Dutch]

[15] **Su H, S Rakheja and TS Sankar**, Vibration and Shock Isolation Performance of a Pressure-limited Hydraulic Damper, *Mechanical Systems and Signal Processing* 3, 71-86, 1989

[16] **Wallashek J**, Dynamics of Non-Linear Automobile Shock-Absorbers, *International Journal of Non-Linear Mechanics* 25, 299-308, 1990

[17] **Armstrong-Helouvry B, P Dupont and C Canudas De Wit**, A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction, *Automatica* 30, 1083-1138, 1994

[18] **Schoukens J and R Pintelon**, *Identification of Linear Systems: a Practical Guideline to Accurate Modelling*, Pergamon Press, London, 1991

List of symbols

Δp : pressure drop [Pa]

Q : flow rate [m^3/s]

K : pressure loss factor [Pa $s^{-1.75}$ $m^{-5.25}$]

G : blow-off valve conductivity [m^6 $N^{-1.5}$ s^{-1}]

Δp_0 : blow-off pressure [Pa]

D : expansion order

V : chamber volume [m^3]

V_0 : reference chamber volume (at 0.1MPa) [m^3]

α : oil compressibility [Pa^{-1}]

p : chamber pressure [Pa]

ρ : oil density [kg/m^3]

ρ_0 : reference oil density (at 0.1MPa) [kg/m^3]

m_o : chamber oil mass [kg]

m_{gb} : chamber bubble gas mass [kg]

R : gas constant for air [J kg^{-1} K^{-1}]

T : temperature [K]

r : bubble gas fraction []

H : constant of Henry [m^2 s^{-2}]

V_{oil} : oil volume [m^3]

m_{gs} : solved chamber gas mass [kg]

m_{gt} : total chamber gas mass [kg]

r_t : total gas fraction []

ξ : adiabatic constant (reserve chamber) [Pa $m^{4.2}$]

ρ_{gas} : gas density [kg/m^3]

d : bubble diameter [m]

v : bubble velocity in fluid [m/s]

g : acceleration of gravity [m/s^2]

μ : dynamic oil viscosity [kg m^{-1} s^{-1}]

p_k : critical or saturation pressure [Pa]

L : mid length of inner cylinder [m]

x : position of piston with respect to mid position [m]

A : area [m^2]

F : damper force [N]

F_0 : offset force [N]

Z : cost function [Pa^2]

Z' : alternative cost function [N^2]

N : number of samples per period []

M : number of periods []

f_{sample} : sample frequency [Hz]

t : time [s]

t_i : sample instant [s]

τ : excitation period [s]

σ : standard deviation of measurement noise [N]

List of indices

(0) intake valve
(1) port restriction
(2) bleed valve or leak restriction
(3) blow-off valve
(tot) total valve
(r) rebound (chamber)
(c) compression (chamber)
(s) reserve chamber

(d) piston rod
(meas) measured
(mod) modelled

Appendix

Table 1: Simulation parameters

valve parameters :	
$K_{1,r}$ =2.78 10^{12} Pa $s^{1.75}$ $m^{-5.25}$	
$K_{2,r}$ =1.52 10^{14} Pa $s^{1.75}$ $m^{-5.25}$	
G_r =1.07 10^{-14} m^6 $N^{-1.5}$ s^{-1}	
$K_{1,c}$ =4.78 10^{12} Pa $s^{1.75}$ $m^{-5.25}$	
$K_{2,c}$ =1.28 10^{15} Pa $s^{1.75}$ $m^{-5.25}$	
G_c =8.55 10^{-15} m^6 $N^{-1.5}$ s^{-1}	
$\Delta p_{0,r}$ = 5.27 MPa	$\Delta p_{0,c}$ = 2.46 MPa
dynamic parameters :	
α = 2 10^{-12} Pa^{-1}	r_t = 10^{-5}
ρ_0 = 935 kg/m^3	ξ = 0.90 Pa $m^{4.2}$
geometric parameters :	
A_r = 3.85 10^{-4} m^2	A_c = 4.80 10^{-4} m^2
L = 0.11 m	V_s = 1.15 10^{-4} m^3
fixed parameters :	
T = 323 K	H = 10^8 m^2 s^{-2}
R = 297 J kg^{-1} K^{-1}	p_s = 0.5 MPa

Table 2: Identification parameters

valve parameters :	
$K_{1,r}$ =2.04 10^{11} Pa $s^{1.75}$ $m^{-5.25}$	
$K_{2,r}$ =2.99 10^{13} Pa $s^{1.75}$ $m^{-5.25}$	
G_r =5.95 10^{-14} m^6 $N^{-1.5}$ s^{-1}	
$K_{1,c}$ = 5.81 10^{12} Pa $s^{1.75}$ $m^{-5.25}$	
$K_{2,c}$ = 2.33 10^{15} Pa $s^{1.75}$ $m^{-5.25}$	
G_c = 1.95 10^{-14} m^6 $N^{-1.5}$ s^{-1}	
$\Delta p_{0,r}$ = 2.15 MPa	$\Delta p_{0,c}$ = 2.10 MPa

Table 2: Identification parameters

dynamic parameters :	
α = 58 10^{-12} Pa^{-1}	r_t = 1.27 10^{-5}
ρ_0 = 935 kg/m^3	p_k = 5.7 MPa
geometric parameters :	
A_r = 3.85 10^{-4} m^2	A_c = 4.80 10^{-4} m^2
L = 0.11 m	
fixed parameters :	
T = 323 K	N = 2048
R = 288.5 J kg^{-1} K^{-1}	f_{sample} = 526.32 Hz

Addresses of the authors

Stefaan W Duym
Vrije Universiteit Brussel
Department of Electrical Engineering
Pleinlaan 2, 1050 Brussels, Belgium
e-mail: sduym@vnet3.vub.ac.be
fax: +32 2 629 28 50

Randy Stiens
LMS International
Interleuvenlaan 68, 3001 Leuven, Belgium
e-mail: rst@lms-hq.mhs.compuserve.com
fax: +32 16 38 43 50

Gino V Baron
Vrije Universiteit Brussel
Department of Chemical Engineering
Pleinlaan 2, 1050 Brussels, Belgium
e-mail: gvbaron@vnet3.vub.ac.be
fax: +32 2 629 32 48

Koenraad G Reybrouck
Monroe Europe, Division Research & Systems
Industriezone Schurhovenveld 1420
3800 St-Truiden, Belgium
e-mail: koen.reybrouck@monroe.be
fax: +32 11 70 33 04

970103

Effect of Steering-Housing Rubber Bushings on the Handling Responses of a Vehicle

Jungho Park and Parviz E. Nikravesh
The University of Arizona

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

An investigation of the effect of rack-housing rubber bushings on the handling characteristics of a vehicle is presented using a sophisticated three-dimensional vehicle model based on multibody dynamic analysis method. Previous research on this problem has been limited to a transfer function model for a simplified one- or two-dimensional steering subsystem. This paper uses a multibody modeling approach to find the effects of the steering-system compliance on the complete vehicle system. Sample simulations for circular cornering and pulse steering show that the steering-system compliance is the source of the frequency peak in the yaw rate to hand-wheel angle response function.

INTRODUCTION

The rotation of the steering hand-wheel provides the rotation of the road-wheel around the kingpin axis (steering axis). Conversely, tire forces and moments could rotate the steering hand-wheel, which is called the reversibility of the steering system. However, the reduction ratio of a steering gear produces some amount of irreversibility. In general, the steering system is reversible for light vehicles and semi-reversible for heavy vehicles. The reversibility allows jolts from irregularities in the road surface to be transmitted to the steering hand-wheel. In order to diminish this problem, a suitable compliance is introduced to the steering system. The means of providing such compliance are the rubber-elastic connections and the steering-housing rubber bushings. However, it is generally known that the steering compliance degrades the handling performance of a vehicle. The steering system not only poses a ride-quality problem due to the hand-wheel vibration, but also influences the lateral stability of a vehicle. Therefore, the steering-system dynamics have long been a subject for investigation, using mathematical models for steering systems.

Slibar and Paslay [1], Fiala [2], and Bohm [3] have found that the steering-system compliance influences the directional stability of the vehicle system. Campbell [4] pointed out that the compliance or free play in a steering system could lead to the phenomenon known as "shimmy", where the front wheels flutter in and out with a relatively small magnitude. Shock load in the tire lateral force excites the shimmy. Another excitation source is a single bump or a pothole in the road which causes angular deflection of the wheel axis, which in turn results in steering oscillations due to the gyroscopic torque of the front wheels. The latter case is especially evident in the beam-axle front suspension becaue the left and right wheels roll together.

Lanchester [5] and Bastow et al. [6-7] also discussed a steering-vibration problem which is variously described as front wheel wobble, wheel flap or veer vibration. They described the veer vibration problem as the interaction between the front end mass connected to the kingpin pivots, the lateral compliance inherent in the steering system, and the self-aligning torque of the front tires. Ignoring the yaw motion of a car, they constructed a model for the front end of a car as shown in Figure 1. The wheels pivot about points P which are the points where the effective caster applies. It should be noted that the distance between points C and P is different from the lever-arm length of the steering arm. The effective mass M may be smaller than the front end mass of the car. The lateral compliance is lumped as the two springs K between the effective mass and the kingpin pivots.

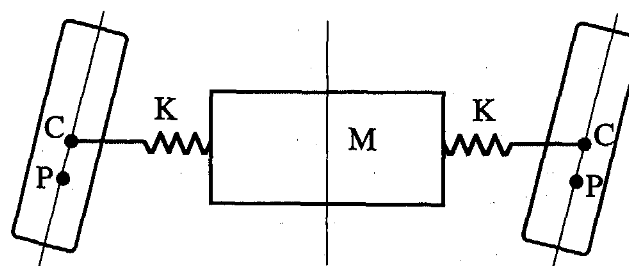


Figure 1. Front end model of a car for veer vibration

Segel [8] proposed a quasi-linear representation of a steering system which is a one-dimensional mechanism having two degrees of freedom with Coulomb friction. Figure 2 depicts the schematic of Segel's steering-system model. He incorporated the steering system model in a linear vehicle model with three degrees of freedom. Using the combined model, he studied the influence of the steering system parameters on the vibration modes of the free-control vehicle. Mabrouka et al. [9] used Segel's mathematical model to investigate the effect of tire lateral flexibility on the dynamic behavior of the steering system.

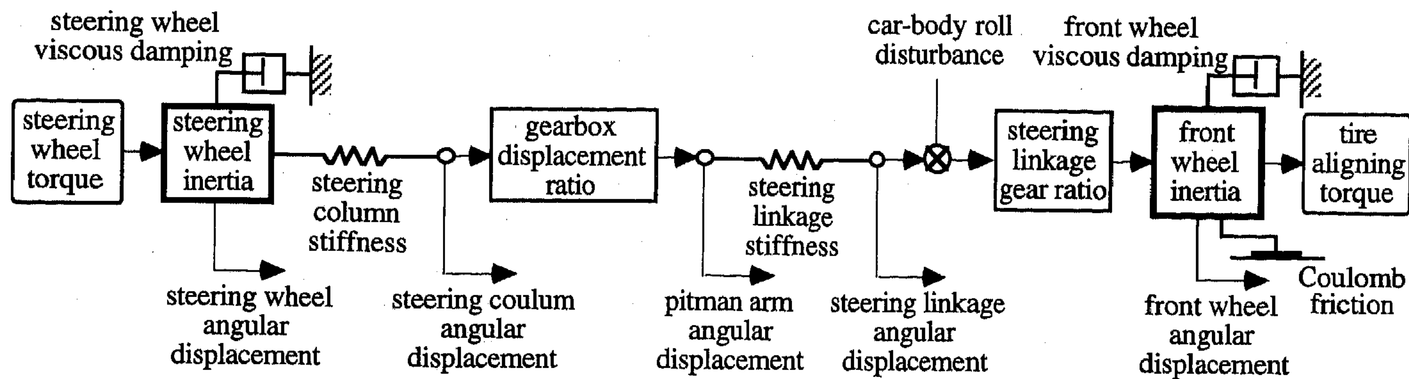


Figure 2. Segel's discrete model for the Pitman-arm steering system

Gillespie [10-11] suggested a lumped parameter model for a steering subsystem. Figure 3 depicts the schematic of his model. The model is a two-dimensional representation of a steering mechanism including the composite stiffness, K_{ss}, between the steering gearbox and the road wheels. He developed the modeling equations based on the geometry of the mechanism and the relationship between the forces and displacements. He suggested possible improvements to the model such as the modeling of the steering gearbox as the effective steering ratio with the appropriate input/output torque relationships and the addition of Ackerman steer angle relationships between the road wheels. MacAdam et al. [12] also used a similar steering-subsystem model and combined it to a vehicle model.

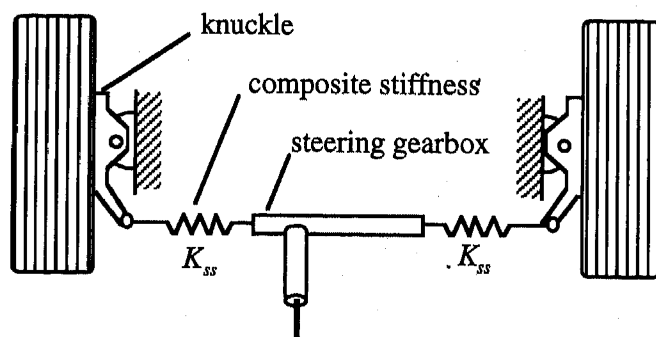


Figure 3. Lumped parameter model for the steering system

Ellis [13] introduced both one and two degrees of freedom models for the steering system. His one degree of freedom model is a single spring-mass system where all of the elasticity of the steering mechanism is concentrated into the single spring. His two degrees of freedom model is the extension of the one degree of freedom model by adding the yaw acceleration of the vehicle to the angular acceleration of the front wheels. He incorporated the steering-system equations into the vehicle equation. He showed through various simulations that the compliance of the steering mechanism produced an under-steer tendency to the vehicle for steady-state cornering situation and affected the transient response of the vehicle.

The aforementioned steering-vibration models are basically one- or two-dimensional representations of a steering subsystem, which neglect the gyroscopic effect of the rotating wheels. However, as Campbell [4] pointed out, the gyroscopic effect of front wheels not only induces an excitation torque but also provides an additional inertia effect to the vibration. Other factors that are ignored are the effects of the caster angle, the trail (the caster offset), the kingpin offset, and the geometric effect of the steering and suspension mechanisms. In order to consider these effects, the steering-system model should be three-dimensional.

This paper presents a sophisticated three-dimensional multibody representation of the total vehicle, where the steering-system model is combined with the vehicle model. The vehicle model has a rack-and-pinion steering mechanism, where the rack housing (steering housing) is mounted to the car body by rubber bushings providing lateral compliance of the rack housing with respect to the car body. Using the developed model, an investigation is performed to assess the effect of the stiffness of the rack-housing rubber bushings on the handling characteristics of the vehicle. Sample simulations for a steady-state circular cornering and a pulse steering maneuver [14-15] are performed using three different stiffness levels of the rack-housing rubber bushing. Another simulation without rack-housing rubber bushings (rigid-mounting case) is also performed for the purpose of comparison. The simulation results show that the steering-system compliance is the source of the frequency peak in the yaw rate to hand-wheel angle response function.

MULTIBODY REPRESENTATION OF STEERING SYSTEM

The available studies on the steering compliance have been focused on the analysis of the shimmy and veer vibration of the steering system using simplified one- or two-dimensional models for the steering subsystem, whereas this paper presents a sophisticated three-dimensional multibody representation of the rack-and-pinion steering system as a part of the complete vehicle model.

Figure 4 shows a typical rack-and-pinion steering mechanism. The rack-housing is mounted on the car body by the rubber bushings such that the rack-housing has some lateral compliance. The tie rods link the rack-bar and the knuckles. The connection between them are rod ends (ball joints) whose compliance is, in general, negligible compared to that of the rack-housing bushing.

Figure 5 presents a multibody representation of the rack-and-pinion steering system. In order to allow a relative lateral motion of the rack-housing, a translational joint connects the car body and the rack-housing. A translational spring-damper

element connected between the car body and the rack-housing represents the stiffness and damping of the rack-housing rubber bushing. Considering the nonlinearity of rubber bushings, the spring and damping characteristics of the spring-damper element can generally be expressed as a function of rubber-bushing parameters such as deflection (d), deflection speed ($\dot{d}$), and actuation frequency (ω):

$$K = f(d, \dot{d}, \omega)$$
$$C = g(d, \dot{d}, \omega) \quad (1)$$

where K is the spring coefficient, C is damping coefficient, and f and g are polynomial functions. These kinds of nonlinear characteristics can be easily implemented in user-supplied subroutines of a general-purpose multibody dynamic analysis program [18]. The rack-bar is connected to the rack-housing by another translational joint allowing its relative lateral shift. In addition to the translational joint, a steering constraint defined between the rack-housing and the rack-bar determines the relative shift of the rack-bar. The tie-rod with two rod ends is modeled as a massless spherical-spherical constraint [17] because its inertia is relatively small and can be transferred to the inertia of the rack-bar.

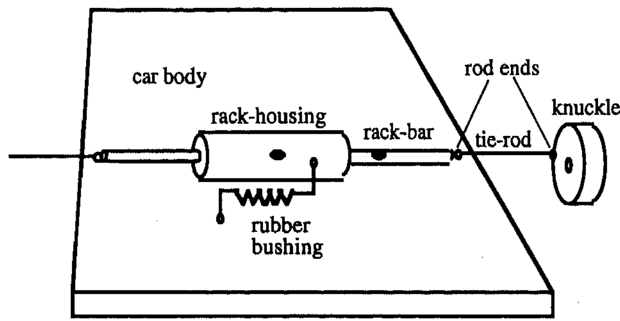


Figure 4. Typical rack-and-pinion steering system

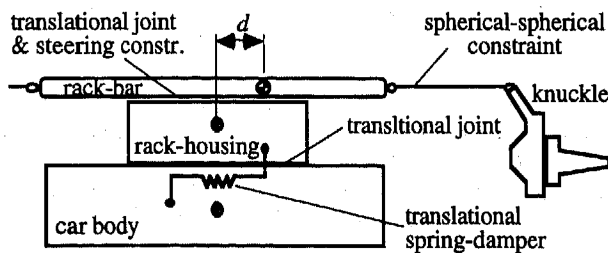


Figure 5. Multibody representation of rack-and-pinion steering system

In order to provide steering angles to the front wheels, some relative shift of the rack-bar with respect to the rack-housing should be mathematically enforced in the equations of motion of the vehicle. In the case of the joint coordinate formulation [18] used in this study, such enforcement can be easily done by introducing a simple constraint for the joint coordinate corresponding to the relative rack-bar shift. Let θ_{rb} be the joint coordinate for the relative rack bar-shift. Then, the simple constraint equation is

$$\Phi \equiv \theta_{rb} - \theta(t) = 0 \quad (2)$$

where $\theta(t)$ is the user defined time-history for the relative rack-bar shift. The first time derivative of Equation (2) yields the velocity constraint equation which contains the Jacobian, $\mathbf{C} \equiv \partial\Phi / \partial\theta$, where θ is the vector of joint coordinates. The second time derivative of Equation (2) yields the acceleration-constraint equation.

Some of the commercially available multibody dynamics programs use a set of so-called absolute coordinates to formulate the equations of motion. In the case of the absolute coordinate formulation, assume that the center of gravity (CG) of the rack-bar is off-center when it is in neutral steering position. Let d be the distance from the CG of the rack-housing to the CG of the rack-bar, and let l be the target value of the distance, i.e., the steering command in the rack-bar position (Figure 5). Then, the kinematic steering constraint for the relative location of the rack-bar can be expressed as a scalar equation:

$$\Phi \equiv d^2 - l^2 = 0 \quad (3)$$

This steering constraint is also called the displacement constraint equation. The square of the distance d can be computed using the absolute locations of the rack-housing and the rack-bar as:

$$d^2 = |\bar{r}_j - \bar{r}_i|^2$$
$$= (r_j(x) - r_i(x))^2 + (r_j(y) - r_i(y))^2 + (r_j(z) - r_i(z))^2 \quad (4)$$

where $\bar{r}_i = (r_i(x) \quad r_i(y) \quad r_i(z))$
$\bar{r}_j = (r_j(x) \quad r_j(y) \quad r_j(z))$

The first time derivative of Equation (3) yields the velocity constraint equation which contains the Jacobian, $\mathbf{D} \equiv \partial\Phi / \partial \boldsymbol{q}$, where $\mathbf{q}$ is the vector of absolute coordinates. The second time derivative of Equation (3) yields the acceleration-constraint equation.

In the joint coordinate formulation, it can be shown [16-18] that the equations of motion for a multibody system not containing any closed kinematic loops can be described as a set of ordinary differential equations as:

$$\boldsymbol{M}\ddot{\theta} = \boldsymbol{f} \quad (5)$$

where $\boldsymbol{M}$ is the generalized inertia matrix, $\ddot{\theta}$ is the vector of joint accelerations (which contains $\ddot{\theta}_{rb}$), and $\boldsymbol{f}$ is the vector of generalized forces. However, when closed kinematic loops are present in a multibody system, Equation (5) finds a slightly different form as:

$$\begin{bmatrix} \boldsymbol{M} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ddot{\theta} \\ \nu \end{bmatrix} = \begin{bmatrix} \boldsymbol{f} \\ \varepsilon \end{bmatrix} \quad (6)$$

where **C** represents the Jacobian of all of the constraints (closed loop constraints and the steering angle constraint), ν is a vector of Lagrange multipliers, and ε contains the terms in the acceleration constraints which do not contain accelerations (quadratic velocity and time dependent terms).

In the case of the absolute coordinate formulation, the equations of motion for a multibody system are written as [17]:

$$\begin{bmatrix} \mathbf{M} & \mathbf{D}^T \\ \mathbf{D} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{q}} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{g} \\ \gamma \end{bmatrix} \qquad (7)$$

where $\mathbf{M}$ is the mass matrix, , $\mathbf{g}$ is the vector of external forces and moments, $\mathbf{D}$ is the Jacobian matrix, λ is a vector of Lagrange multipliers, and λ is the vector of right-hand-side terms in the acceleration constraints (quadratic velocity and time dependent terms). It should be noted that the acceleration constraints contain all of the joint constraints and the steering constraint.

For the simulation case of no rack-housing bushing (rigid mounting), additional simple constraint may be introduced between the rack-housing and the car body. The simple constraint is a deletion of a specific relative degree of freedom in the equations of motion. Therefore, the simple constraint can be easily implemented in the joint-coordinate-based equations of motion. If a simple constraint is applied to the translational degree of freedom of the rack-housing with respect to the car body, the rack-housing would behave just as if it were fixed to the car body (Figure 6).

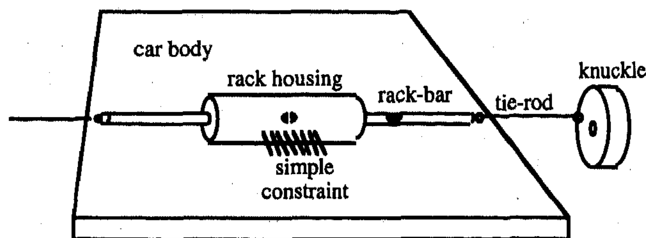


Figure 6. Rack-and-pinion steering without bushing effect

The steering-system model mentioned above is combined with the vehicle model described in Appendix A forming a complete vehicle system. Figure 7 shows how the driving simulation using the complete vehicle model is performed. In the figure, the simulation data file contains the time history of a user-defined open-loop rack-bar shift command which is used in the dynamic analysis program as an input to the kinematic rack-bar shift constraint.

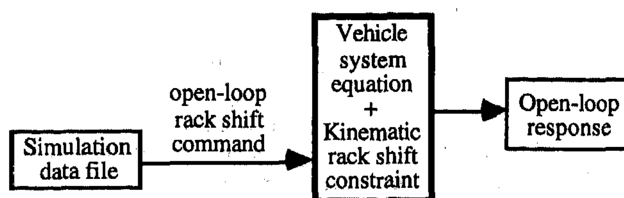


Figure 7. Control of rack-and-pinion steering mechanism

SIMULATION RESULTS AND DISCUSSION

Computer simulations using a sophisticated three-dimensional vehicle model and a general purpose multibody dynamic analysis program are presented for a steady state circular maneuver and a pulse steering maneuver. The purpose of the steady state circular maneuver is to obtain information about the influence of rack-housing bushing on the steady state handling characteristics of a vehicle. The pulse steering maneuver reveals how the flexibility of tires affects the motion of a vehicle in transient state. Simulation results for the pulse steering maneuver is analyzed in both time- and frequency-domain. Time-domain evaluations were conducted based on the maximum values of rubber bushing deflection, vehicle yaw rate, roll rate, and lateral acceleration. Frequency-domain evaluations were conducted using frequency response functions of yaw rate, roll rate, lateral acceleration, and tire cornering powers, that were obtained by the fast Fourier transformation (FFT) of the time-domain response. For comparative study, the rack-housing bushings are assumed to have linear spring and damping characteristics. In addition, the damping coefficient of the bushing is assumed to be 1% of its stiffness value. However, nonlinear characteristics can be easily modeled in a multibody dynamic analysis program if the nonlinear data is available.

3D VEHICLE MODEL - To investigate the effect of the steering-system compliance through computer simulations, a sophisticated three-dimensional vehicle model which accurately represents a mid-size sedan is constructed as presented in Appendix A. In brief, the vehicle model is a rigid multibody system composed of thirty-one rigid (non-deformable) bodies and thirty-six kinematic joints or constraints. The vehicle model includes all rubber-flexibility effects such as rack-housing compliance, suspension compliance, and tire lateral flexibility. The vehicle model has a total of twenty two degrees of freedom: 6 for the car body, 4 for the wheel travel, 4 for the suspension longitudinal compliance, and 4 for the tire lateral flexibility.

STEADY STATE CIRCULAR TEST - In the steady-state circular test, an open-loop steering command (input) is provided as a fixed rack-bar shift of 6.6 mm to the right with respect to the rack housing, which corresponds to about 1.4 degrees left turn of the steering road-wheel. Initially, the steering command is set to zero, and then gradually it rises up to 6.6 mm which is held constant so that the motion of the vehicle model reaches a steady state. The simulation is performed for 60 and 80 km/h cases. A standard rack-housing bushing with the stiffness of 951.2 kN/m and the damping of 9.512 kN-sec/m is used for both simulation cases. Steady-state values for 60 km/h steady-state circular test are listed in Tables 1, and those for 80 km/h are listed in Tables 2.

The simulation results show noticeable lateral shift of the rack-housing: 1.549 mm (23.4% of 6.6 mm, the relative rack-bar shift) for 60 km/h case; 2.091 mm (31.7% of 6.6 mm) for 80 km/h case. Thus, the steering compliance reduces the absolute rack-bar shift and the net steering angles, resulting in under-steer tendency in the vehicle. However, the deformations of the suspension-rubber bushings are less than 0.25 mm for both 60 and 80 km/h cases. This implies that the suspension longitudinal compliance plays a negligible role in the steady-state circular maneuver. The longitudinal compliance is known to be effective on a rough terrain.

Table 1. Steering and body data at 60 km/h

Items	Without Lat. Compliance	With Lateral Compliance
Rack-housing shift (mm)	1.427	1.549
Avg. steering angle (deg)	2.404	2.338
Yaw rate (deg/sec)	13.72	13.36
Lateral accel. (m/s^2)	3.989	3.886
Roll angle (deg)	1.377	1.014
Path Radius (m)	69.54	71.45

Table 2. Steering and body data at 80 km/h

Items	Without Lat. Compliance	With Lateral Compliance
Rack-housing shift (mm)	1.870	2.091
Avg steering angle (deg)	2.171	2.055
Yaw rate (deg/sec)	14.50	13.76
Lateral accel. (m/s^2)	5.620	5.339
Roll angle (deg)	1.942	1.334
Path Radius (m)	87.73	92.40

PULSE STEERING TEST - The frequency-domain analysis using a pulse-steering input is called the "pulse steering technique" [19]. The open-loop steering inputs that have been used for the frequency-domain analysis are pulse input, sinusoidal sweep input, discrete sinusoidal input, and random sweep input. Heydinger et al. [20] pointed out that the pulse steering technique is generally faster and cheaper than the alternative test methods without sacrificing accuracy. They also found that the frequency-response results are relatively insensitive to the pulse shape and size, provided that the pulse input have sufficient power to excite the vehicle system. The open-loop pulse steering input used in this paper is a sinusoidal pulse with the magnitude of 3 mm in relative rack-bar shift and the duration time of 0.4 second as shown in Figure 9 (a).

For frequency-domain analysis, fast Fourier transformation (FFT) of time domain data of selected cornering parameters is performed to obtain frequency response functions (FRF), and then complex-cornering powers are computed based on a simplified bicycle model of the vehicle (refer to Appendix B). Figure 8 presents the procedure of frequency-domain analysis for computing the frequency-response functions for yaw rate to hand-wheel angle (YR/MA), lateral acceleration to hand-wheel angle (LA/MA), and roll rate to hand-wheel angle (RR/MA); and for generating complex-cornering powers for front- and rear-wheel sets.

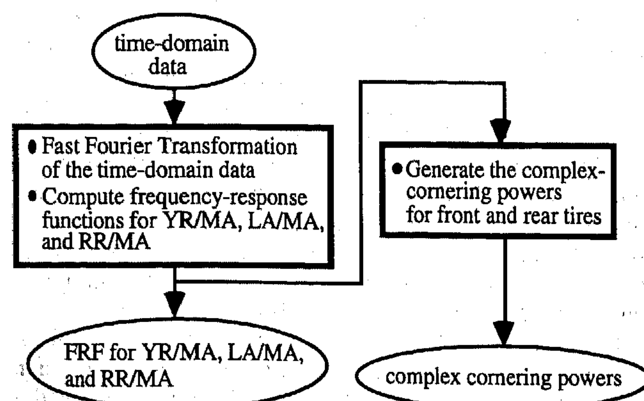


Figure 8. Schematic of frequency-domain analysis procedure

Simulations are performed for four different levels of rack-housing rubber bushings as listed in Table 3. The rigid mounting does not have any elasticity such that the rack-housing is rigidly mounted on the car body. The time-domain plots are presented in Figure 9, and the frequency-domain plots are presented in Figure 10.

Table 3. Simulation cases for pulse steering test

Type of bushing	Stiffness (kN/m)	Damping (kN-sec/m)	% of Standard
Soft	571	5.71	60%
Standard	951.2	9.512	100%
Hard	1,332	13.32	140%
Rigid	infinite	infinite	infinite

Time-domain simulation results show that the peak deformation of rack-housing bushing increases as the bushing gets softer (Figure 9 (c)). The magnitude of deformation is noticeable even though the front-wheel trail is only 16 mm. The trail is the lever arm for cornering forces which are produced by 3.08 degrees of caster angle. The peak deformation is 1.1 mm for the hard-bushing case (37% of 3 mm max. rack-bar shift), 1.26 mm for the standard-bushing case (42%), and 1.6 mm for the soft-bushing case (53%). This steering-subsystem compliance not only reduces road-wheel steering angles but also induces transient fluctuation of steering angles (Figure 9 (b)). In the figure, the peak steering angle is 1.26 degrees for the hard-bushing case (10.6% reduction compared to 1.41 mm of the rigid-bushing case), 1.21 degrees for the standard-bushing case (14.2% reduction), and 1.11 degrees for the soft-bushing case (21.3% reduction). The reduced steering angle directly affects the tire forces (Figure 9 (d)), the lateral acceleration (Figure 9 (f)), the yaw rate (Figure 9 (g)), and the roll rate (Figure 9 (h)), resulting in the under-steer tendency to the vehicle (Figure 9 (j)). It noteworthy that the case of soft rack-housing bushing shows prolonged vibration of rack-housing (Figure 9 (c)), which affects not only the handling characteristics (Figure 9 (f)) but also the ride quality of a vehicle (Figure 9 (g) & (h)).

Frequency-domain results show that rack-housing bushing is the source of the peak frequency in yaw rate to hand-wheel angle response function (Figure 10 (a)): It is clearly seen in the figure that the rigid bushing case does not have a peak frequency. As the stiffness of the rack-housing bushing decreases, the peak frequency gradually moves to higher frequencies. Figure 10 (e) presents roll rate to hand-wheel angle response function. As in the case of the yaw rate to hand-wheel angle response function, the peak frequency of the roll rate to hand-wheel angle response increases from 0.65 Hz (rigid bushing case) to 1 Hz (soft bushing case). For all frequency response functions except the one for rear cornering power, low-frequency (less than 5 Hz) gains decrease as rack-housing bushing becomes softer. This implies that steering-subsystem compliance tends to reduce the handling-responsiveness of the vehicle.

SUMMARY AND CONCLUSIONS

A multibody approach to the investigation of the relationship between the steering-system compliance and the dynamic characteristics of a vehicle has been presented. Three different stiffness/damping levels of rack-housing rubber bushing along with a rigid mounting are used for comparative

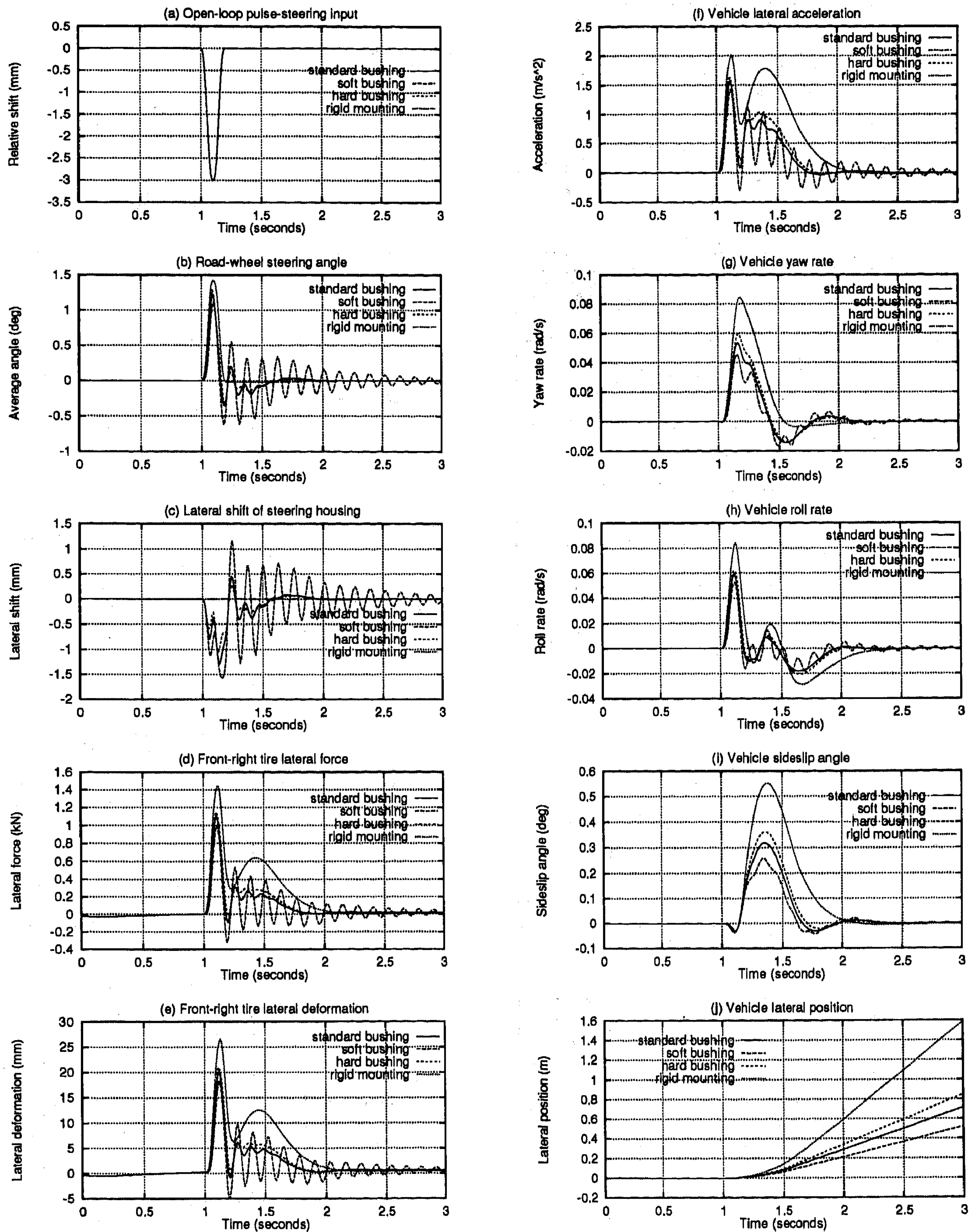


Fig 9. Time-domain responses for the pulse steering test

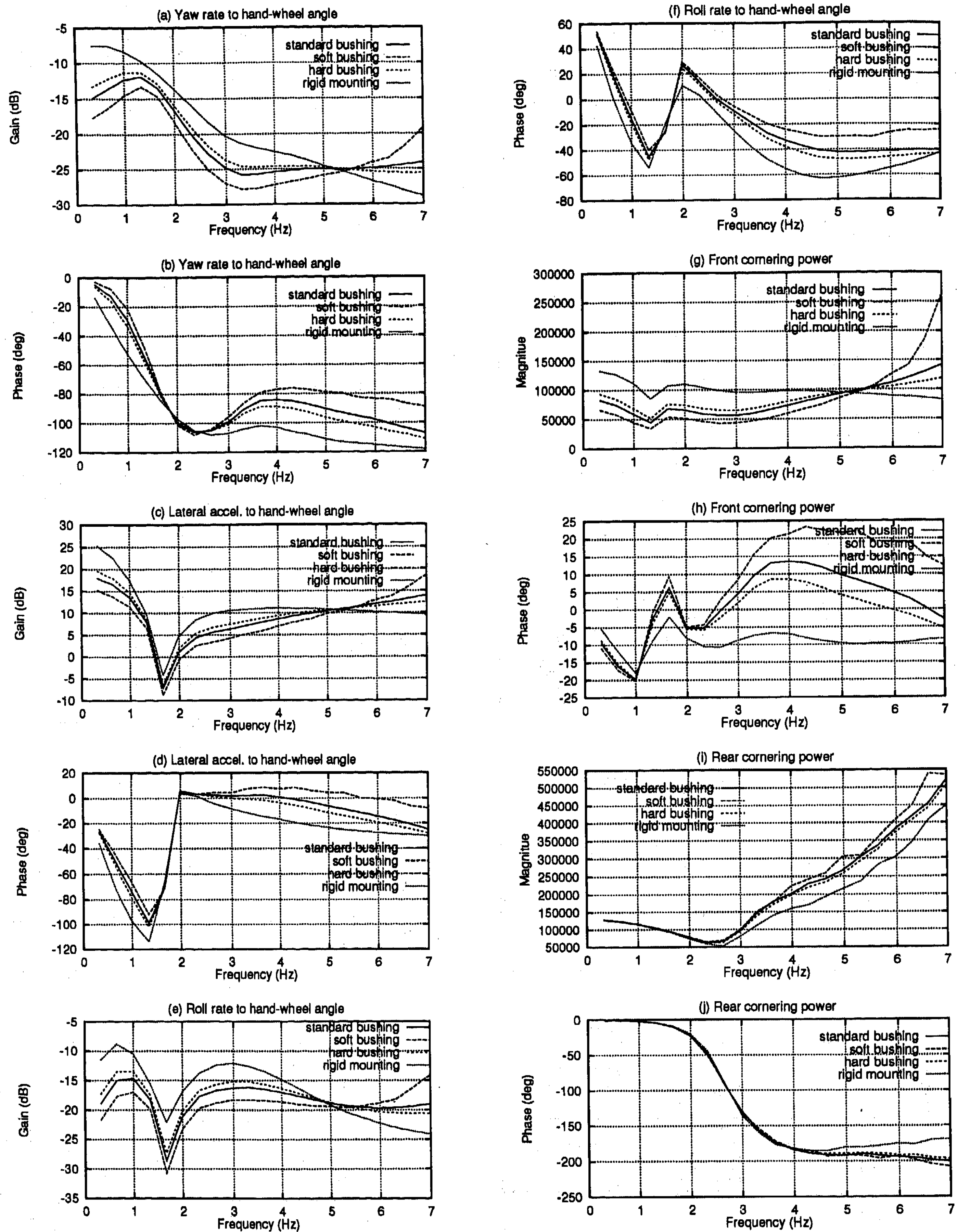


Fig 10. Frequency-domain responses for the pulse steering test

study. Computer simulations are performed for a steady state circular maneuver and a pulse steering maneuver. It has been shown that a sophisticated three-dimensional multibody model of the vehicle including the steering system is an effective tool for the investigation of total vehicle handling characteristics. Key points of the findings may be summarized as follows:

- Rack-housing rubber bushings are intended to isolate the steering hand-wheel from road-induced vibrations. However, they change the dynamic performance of a vehicle by creating an under-steering tendency and a less responsive vehicle.
- The time-domain simulation results show that the steering-system compliance reduces the handling sensitivity of the vehicle and that a too much compliance could induce steering-system vibration that lasts for several seconds.
- Comparison of the frequency-domain simulation results with and without the rack-housing rubber bushing show that the steering compliance is the source of the peak frequency in the yaw rate to hand-wheel angle response function.
- The frequency-domain simulation results also show that the increase in the compliance of the steering system reduces the peak frequencies and the magnitudes of not only the yaw rate but also the roll rate response functions.
- The simulation results indicate that longitudinal suspension compliance has negligible effect on the pulse steering response.

ACKNOWLEDGMENT

This work was supported in part by the Toyota Motor Corp., Higashifuji Technical Center.

REFERENCES

1. Slibar, A. and Paslay, P. R., "Behavior of Vehicles Subjected to Wind Gusts," *Ingenieur Archiv*, Vol. 28, 1959.
2. Fiala, E., "Zur Fahrdynamik des Strassenfahrzeuges unter Berucksicktingung der Lenkungselastizitat," *Automobiltechnische Zeitschrift*, Vol. 62, No. 3, pp. 71, March 1960.
3. Bohm, F., "Uber die Fahrtrichtungsstabilitat und die Seitenwindempfindlichkeit des Kraftwagens bei Geradeaus-Fahrt," *Automobiltechnische Zeitschrift*, Vol. 63, No. 5 and 10, pp. 128, 339, March 1961.
4. Campbell, C., *Automobile Suspensions*, Chapman and Hall, London, 1981.
5. Lanchester, Dr. F. W., "Automobile Steering Gear Problems and Mechanisms," *Proc. Inst. Auto. Engrs.*, Vol. XXII (1927-28).
6. Bastow, D., "Steering Problems and Layout," *Proc. Inst. Auto. Engrs.*, Vol. XXXII (1937-38).
7. Bastow, D. and Howard, G. P., *Car Suspension and Handling*, 3rd edition, Pentech Press, London, 1993.
8. Segel, L., "On the Lateral Stability and Control of the Automobile as Influenced by the Dynamics of the Steering System," ASME Paper 65-WA/MD-2, 1965.
9. Mabrouka, H., Chen, H. F., Maalej, A. Y., and Guenther, D. A., "Effect of Lateral Tire Flexibility on the Steering System Dynamic Behavior," SAE Paper 910239, 1991.
10. Gillespie, T. D., "Front Brake Interactions with Heavy Vehicle Steering and Handling during Braking," SAE Paper 760025, 1976.
11. Gillespie, T. D., *Fundamentals of Vehicle Dynamics*, Society of Automotive Engineers, Inc., Warrendale, 1992.
12. MacAdam. C. C., et al., "A Computerized Model for Simulating the Braking and Steering Dynamics of Trucks, Tractor-Semitrailers, Doubles, and Triples Combinations-User's Manual, Phase 4," Highway Safety Research Institute, University of Michigan, Report No. UM-HSRI-80-58, September 1980.
13. Ellis, J. R., *Road Vehicle Dynamics*, John R. Ellis Inc., Arkon, 1988.
14. Vedamuthu, S., and Law, E. H., "An Investigation of the Pulse Steer Method for Determining Automobile Handling Qualities," SAE Paper 930829, 1993.
15. Heydinger, G. J, Grygier, P. A., and Lee S., "Pulse Steering Techniques Applied to Vehicle Handling Dynamics," SAE paper 930828, 1993.
16. Kim, S. S., and Vanderploeg, M. J,. "A General and Efficient Method for Dynamic Analysis of Mechanical Systems Using Velocity Transformations," *ASME J. Mech., Trans., and Auto. in Design*, Vol. 108, No. 2, June, pp. 176-182, 1986.
17. Nikravesh, P. E., *Computer-Aided Analysis of Mechanical Systems*, Prentice Hall, 1988.
18. Nikravesh, P. E., and Gwanghun Gim, "Systematic Construction of the Equations of Motion for Multibody Systems Containing Closed Kinematic Loops," Paper No. 89-DAC-58, *ASME Design Automation Conference*, Montreal, Canada, September 17-20, 1988.
19. Park, J., and Nikravesh, P. E., "A Look-Ahead Driver Model for Autonomous Cruising on Highways," SAE Paper 961686 (also available in SP-1190), *Future Transportation Technology Conference*, Vancouver, Canada, August 5-8, 1996.
20. Bakker, E., Nyborg, L., and Pacjeka, H. B., "Tyre Modeling for Use in Vehicle Dynamics Studies," *SAE paper*, No. 870421, 1987.
21. Park, J., and Nikravesh, P. E., "A Multibody Approach to Modeling Tire Longitudinal/Lateral Flexibility," in progress, to appear in the 1997 SAE International Congress and Exposition, February 1997.

APPENDIX A: 3D VEHICLE MODEL

The vehicle model used in this paper is an enhanced version of the model used in [19]. This sophisticated three dimensional rigid-multibody vehicle model closely represents a typical mid-size sedan. A schematic of the model with body numbers and joint types is shown in Figure A-1. The multibody modeling of this system with the joint coordinate formulation [16-18] assumes the car body (body 1) to be the only floating base-body in the system. Other bodies form branches extending out from the base-body. The vehicle model

has super-strut front suspensions and quadra-link rear suspensions, both of which are variants of a McPherson strut suspension.

The front wheels are steered by a rack-and-pinion steering mechanism, where the shift of the rack-bar relative to the rack-housing determines the steering angles of the front wheels through the tie-rods. The rack-housing is kinematically joined to the chassis by a translational joint and a rubber-bushing element. A kinematic steering constraint defined between the rack-bar and rack-housing governs the relative shift of the rack-bar. Suspension rubber bushings are modeled by using supplementary bodies (24-31), joints (29T-36T), and translational spring-damper elements of which characteristics are based on a test data but are simplified to be linear.

Any accurate vehicle ride simulation would require a realistic tire model. For this study, the Magic Formula tire model [20] is used. In this tire model, the interacting forces and moments between the tire and the road surface are determined based on empirical formulas as a function of tire orientation, velocity, and radial stiffness. The tire coefficients in the formulas are for the 195/60R13 tire. The tire model is defined on the ring bodies 16 through 19. The tire lateral flexibility is modeled based on the "ring body" concept [21]. Spring-damper elements that represent the tire lateral stiffness and damping are defined between the ring bodies and the cylinder bodies (bodies 10 through 13). The vehicle model represents a front wheel drive car. Therefore, torque of the two front wheels are feedback controlled to maintain a constant vehicle speed. Speed-dependent rolling resistance is applied to all wheels.

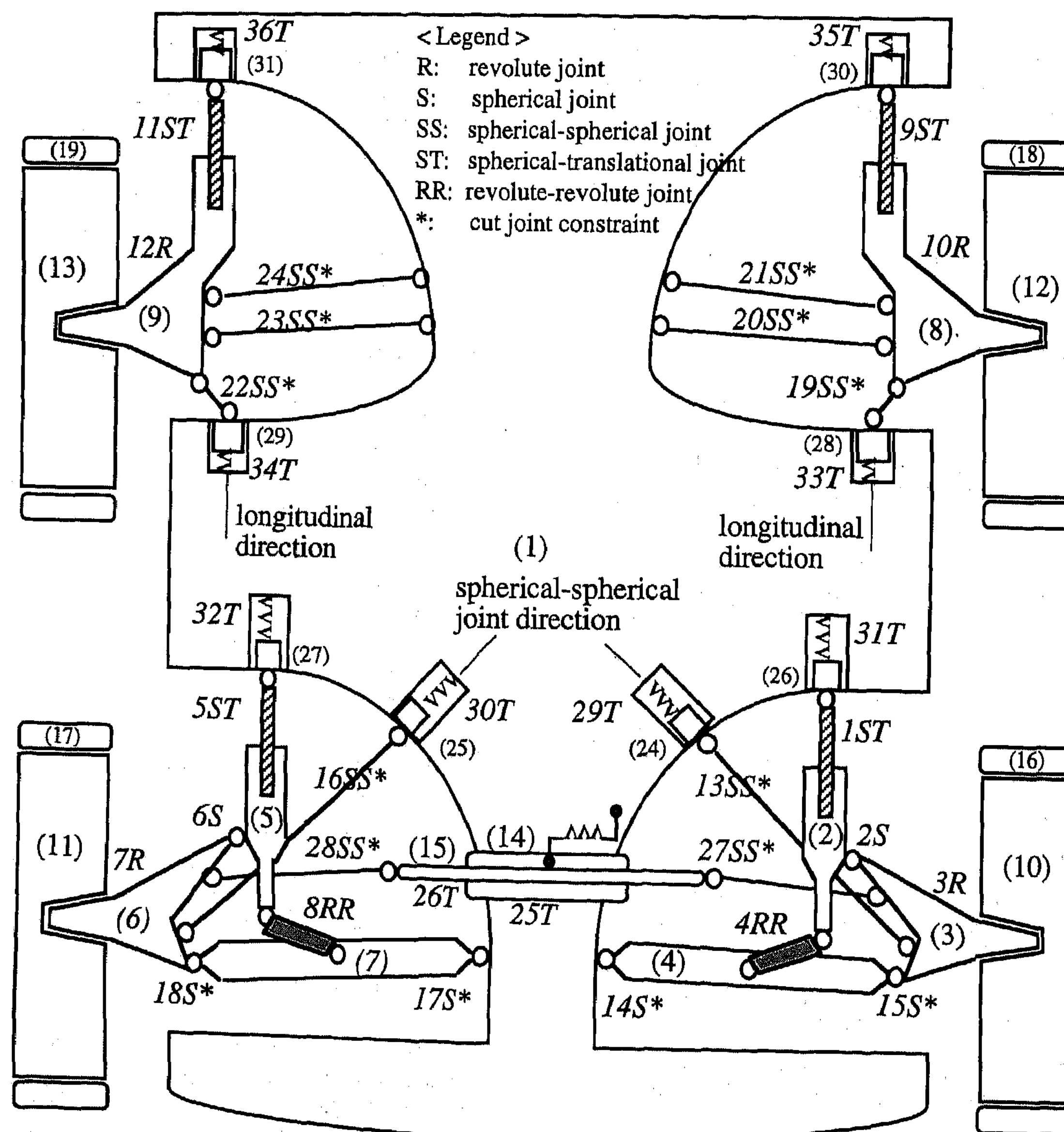


Figure A-1. Schematic of the vehicle model

APPENDIX B: COMPLEX CORNERING POWER

Campbell [4] defined cornering power as cornering force divided by the slip angle required to generate the force. The cornering power is the most important performance parameter of tires. According to Campbell, the factors influencing the cornering power are wheel load, tire profile (aspect ratio), camber angle, tire longitudinal force, and rim width of wheel. Normally, the cornering power designates static cornering power since it is computed from steady-state cornering parameters. A simple extension of the static cornering power to the transient region is called the "dynamic cornering power" which is naturally a function of time. The dynamic cornering power transformed into frequency-domain is called the "complex cornering power".

In order to obtain complex cornering power, let's assume that a planar motion of a vehicle can be simplified as a bicycle model as shown in Figure B-1. In the figure, variables are defined as below:

m : sprung mass of the vehicle (kg)
I : chassis yaw moment of inertia ($kg-m^2$)
l : wheel base (m)
a : distance from chassis CG to front axle (m)
b : distance from chassis CG to rear axle (m)
V : velocity vector of chassis CG (m/sec)
u : forward velocity component (m/sec)
v : lateral velocity component (m/sec)
δ : direct steering angle (radian)
ξ : indirect steering angle due to rolling, compliance, chassis deformation, etc.
β : course angle of the vehicle (radian)
w : yaw rate of the vehicle (rad/sec)
F : cornering force at tire (N)
C : cornering power (N/deg)

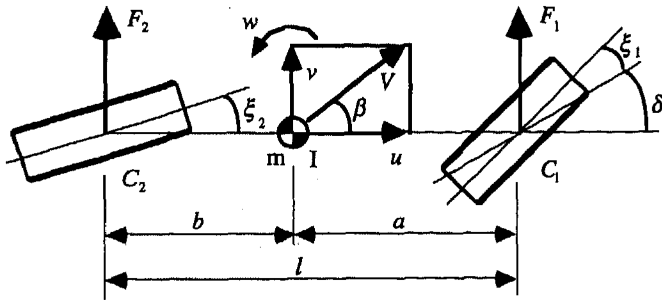


Figure B-1. Bicycle model of a vehicle

The indirect steering angle is a steering angle due to any indirect steering effects such as roll steer, compliance steer, steering due to chassis deformation, etc., with the exception of the direct steering angle command The influence of the indirect steering angle can be equivalently included into the cornering power, producing equivalent cornering power C^*. Then the tire slip angle, α, should also be changed to the pseudo slip angle, α^*. Figure B-2 presents the equivalent model. The relationship between the two slip angles is

$$\alpha_i^* = \alpha_i - \xi_i$$

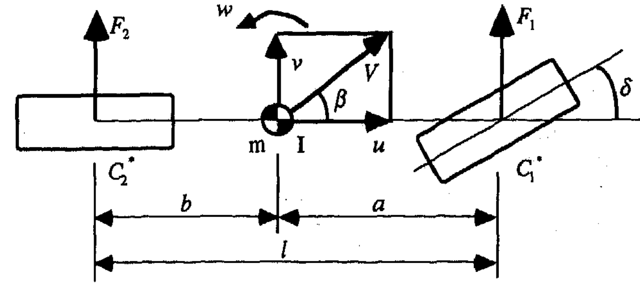


Figure B-2. Equivalent bicycle model of a vehicle

Since the cornering force should be consistent in each model, the cornering forces can be expressed as

$$-F_i = C_i\alpha_i = C_i^*\alpha_i^* \qquad \text{(B.1)}$$

Then, the equivalent cornering power can be defined as

$$C_i^* = C_i\alpha_i/\alpha_i^* = C_i\alpha_i/(\alpha_i - \xi_i)$$

or

$$C_i^* = C_i(\alpha_i^* - \xi_i)/\alpha_i^* = C_i(1 + \xi_i/\alpha_i^*)$$

Once the equivalent cornering power is obtained, the indirect steering angle is given by

$$\xi_i = \alpha_i(1 - C_i/C_i^*) \qquad \text{(B.2)}$$

The above discussion which is based on steady-state time-domain can be expanded to the transient frequency-domain by applying Laplace transformation. In the frequency-domain, each variable should be a complex number, and we can use the term 'complex' implying an expansion from steady-state time-domain to transient frequency-domain.

Now let's consider the equation of motion for the equivalent model in Figure B-2.

$$m \cdot u(\beta(s) \cdot s + w(s)) = F_1(s) + F_2(s) \qquad \text{(B.3)}$$
$$I \cdot w(s) \cdot s = a \cdot F_1(s) - b \cdot F_2(s) \qquad \text{(B.4)}$$
$$F_i(s) = C_i^*(s) \cdot \alpha_i^*(s) \qquad \text{(B.5)}$$
$$\alpha_1^*(s) = \beta(s) + a \cdot w(s)/u - \delta(s) \qquad \text{(B.6)}$$
$$\alpha_2^*(s) = \beta(s) - b \cdot w(s)/u \qquad \text{(B.7)}$$

where s is a complex number in Laplace transformation, and it is expressed as

$$s = -j\omega = -j2\pi f$$

where $j = \sqrt{-1}$. Let's suppose following time domain data are available from multibody dynamic simulation of a vehicle:

MA(t) : hand wheel (steering wheel) angle
YR(t) : yaw rate = w(t)
LA (t) : lateral acceleration of the vehicle

Then , through FFT of these time domain data and through some complex domain operations, transfer functions of YR(s)/MA(s) and LA(s)/MA(s) can be obtained. Therefore, we can use following normalized terms for yaw rate and lateral acceleration in Laplace domain:

$$w(s) = YR(s)/MA(s) \quad (B.8)$$
$$la(s) = LA(s)/MA(s) \quad (B.9)$$

In addition, steering angle in Laplace domain can be expressed using a steering gear ratio and a frequency domain steering model. If we assume that the steering gear ratio, GR, is constant and that the steering model can be represented as a simple first order delay of

$$st(s) = \frac{1}{1+\tau s} \quad (B.10)$$

where τ is a time constant, then the steering angle can be represented as

$$\delta(s) = GR \cdot st(s) = \frac{GR}{1+\tau s} \quad (B.11)$$

From equation (B.3), the lateral acceleration in frequency domain is

$$la(s) = u \cdot (\beta(s) \cdot s + w(s)) \quad (B.12)$$

From this equation, he frequency-domain course angle of the vehicle can be expressed as

$$\beta(s) = (la(s)/u - w(s))/s \quad (B.13)$$

Then, the pseudo slip angles of tire can be calculated by

$$\begin{aligned} \alpha_1^*(s) &= \beta(s) + a \cdot w(s)/u - \delta(s) \\ &= \beta(s) + a \cdot w(s)/u - GR/(1+\tau s) \\ \alpha_2^*(s) &= \beta(s) - b \cdot w(s)/u \end{aligned} \quad (B.14)$$

Cornering forces are obtained from equations (B.3) and (B.4) as

$$\begin{aligned} F_1(s) &= (b \cdot m \cdot la(s) + I \cdot w(s) \cdot s)/(a+b) \\ F_2(s) &= (a \cdot m \cdot la(s) - I \cdot w(s) \cdot s)/(a+b) \end{aligned} \quad (B.15)$$

Then, the complex (dynamic) equivalent cornering power is given from equation (B.1) by

$$\begin{aligned} C_1^* &= -F_1(s)/\alpha_1^*(s) \\ C_2^* &= -F_2(s)/\alpha_2^*(s) \end{aligned} \quad (B.16)$$

970104

Elastokinematic Analysis and Optimization of Suspension Compliance Characteristics

Ju Seok Kang, Jung Rak Yun, and Jang Moo Lee
Seoul National Univ.

Tae Oh Tak
Kangwon National Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Compliance elements such as bushings of a suspension system play a crucial role in determining the ride and handling characteristics of the vehicle. In this research, a general procedure for the optimum design of compliance elements to meet various design targets is proposed. Based on the assumption that the displacements of elastokinematic behavior of a suspension system under external forces are very small, linearized elastokinematic equations in terms of infinitesimal displacements and joint reaction forces are derived. Directly differentiating the linear elastokinematic equations with respect to design variables associated with bushing stiffness, sensitivity equations are obtained. The design process for determining the bushing stiffness using sensitivity analysis and optimization technique is demonstrated.

INTRODUCTION

Bushings are widely used in a vehicle suspension system to improve ride characteristics. The elastic behavior of bushings, however, inherently causes change of wheel attitude due to the forces applied to the tire, which has significant effects on handling characteristics. Thus, when designing a suspension system, the effects of bushings should be considered carefully in view of handling as well as ride characteristics. Compliance characteristics of suspension systems can be accounted for in the light of the changes of static design factors such as toe angle, camber angle and, longitudinal and lateral deflections of the wheel due to the forces and moments applied to the tire-road contact patch. The effects of these characteristics on handling were investigated by Bundorf[1],[2]. The compliance effects on ride have been also taken into consideration to design a suspension system[3].

Compliance characteristics of a suspension system can be measured through the Suspension Parameter Measuring Devise(SPMD)[4]. However in the early developing stage of a suspension, there should be an analytical means for calculating compliance characteristics. Simple calculation by 2-dimensional analysis can be adopted, but its accuracy is open to doubt. Recently, more accurate computation of compliance characteristics has been tried by elastokinematic analysis techniques for multibody systems. These methods find the solutions by simultaneously solving the kinematic constraint equations and force equilibrium equations. However, the drawback of these approaches is that the solution process is very sensitive to the initial estimates of the solutions. To resolve these problem, the method of minimizing the potential energy of the system has been proposed[5].

The magnitude of wheel attitude change due to external forces at the tire contact point is nearly proportional to the magnitude of force, thus this behavior can be considered to be linear. So, in this study, based on the assumption that the displacements of suspension systems under the external forces are very small, a linear form of the elastokinematic equations in terms of infinitesimal displacements and joint reactions forces is derived. The equations can be applied to any form of suspension once the types of kinematic joints and bushings are identified. The validity of the method can be proved through the comparison of the results from any established elastokinematic analysis software such as ADAMS or DADS. By directly differentiating the linear elastokinematic equations with respect to design variables associated with bushing stiffness, sensitivity equations can be obtained. Then, using the results of sensitivity analysis, an optimization process to meet the various compliance targets is to be performed.

DERIVATION OF LINEAR ELASTOKINEMATIC EQUATIONS

To obtain linear equations for elastokinematic behavior of a suspension system, it is assumed that the displacements of the suspension system supported by kinematic joints and elastic elements are small under the action of external forces. Let the generalized coordinates for a suspension element be the small displacements of the Cartesian coordinates, as shown in Fig.1

$$\Delta \mathbf{q}_i = \{\Delta \mathbf{r}, \Delta \boldsymbol{\pi}\}_i^T = \{\Delta x, \Delta y, \Delta z, \Delta \pi_x, \Delta \pi_y, \Delta \pi_z\}_i^T \quad (1)$$

where, $\Delta \mathbf{r}_i$ and $\Delta \boldsymbol{\pi}_i$ are respectively the vectors for small translational and rotational displacements in the inertial reference system.

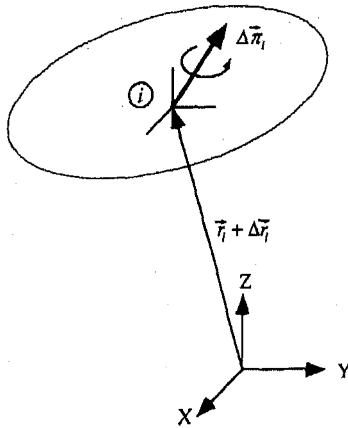


Fig. 1 Generalized coordinates

Assume that the suspension system under the action of external forces is in equilibrium state. The kinematic joints of the suspension can be represented as some algebraic constraint equations, where the general form can be expressed as;

$$\boldsymbol{\Phi}(\mathbf{q}) = \mathbf{0} \quad (2)$$

For small displacement $\Delta \mathbf{q}$, the constraint equation must satisfy the following condition

$$\Delta \boldsymbol{\Phi} = \boldsymbol{\Phi}_{\mathbf{q}} \Delta \mathbf{q} = \mathbf{0} \quad (3)$$

where $\boldsymbol{\Phi}_{\mathbf{q}}$ is Jacobian matrix of kinematic constraints.

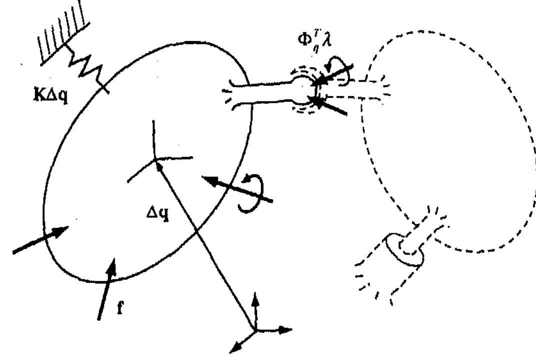


Fig.2 Force-displacement diagram between the bodies

The displacements of a body under external forces can be assumed to be small if the elastic stiffness is big enough to exert sufficient reaction forces. This assumption is very plausible for vehicle suspension systems where, in general, very large bushing stiffness are used. The sum of external forces acting on the body is equal to the sum of elastic reaction forces of the bushings and constraint reaction forces of the joints as shown in Fig. 2. This equilibrium condition can be written as

$$\boldsymbol{\Phi}_{\mathbf{q}}^T \boldsymbol{\lambda} + \mathbf{K}\Delta \mathbf{q} = -\mathbf{f} \quad (4)$$

where $\mathbf{K}\Delta \mathbf{q}$ is the elastic reaction forces due to small displacements exerted by elastic elements such as springs or bushing elements, $\mathbf{f}$ is the external force, $\boldsymbol{\lambda}$ is Lagrange multiplier and, $\boldsymbol{\Phi}_{\mathbf{q}}^T \boldsymbol{\lambda}$ is the term associated with constraint reaction forces.

Eqs. (3) and (4) can be combined to a matrix form.

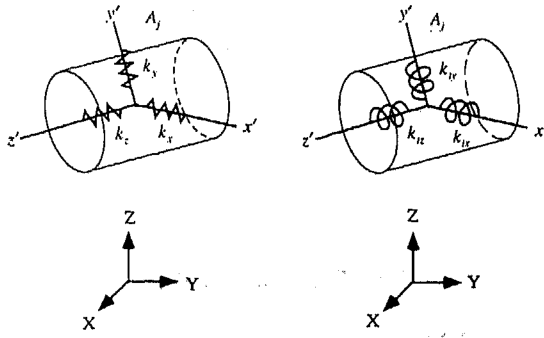
$$\begin{bmatrix} \mathbf{K} & \boldsymbol{\Phi}_{\mathbf{q}}^T \\ \boldsymbol{\Phi}_{\mathbf{q}} & 0 \end{bmatrix} \begin{Bmatrix} \Delta \mathbf{q} \\ \boldsymbol{\lambda} \end{Bmatrix} = \begin{Bmatrix} -\mathbf{f} \\ \mathbf{0} \end{Bmatrix} \quad (5)$$

Eq. (5) is the governing equations for elastokinematic behavior of a suspension system that undergoes a small displacement due to the external forces. The left hand side of Eq. (5) is symmetric if stiffness matrix K is symmetric, which will be demonstrated later. To solve the equations, stiffness matrix and constraint Jacobians should be determined.

The elastic elements of a suspension system such as bushings and springs can be specified as 3 translational springs and 3 rotational springs, as shown in Fig. 3, in the local coordinate system. Stiffness matrices in the local coordinate system can be expressed as

$$\mathbf{K}_T{}' = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{bmatrix} \quad (6)$$

$$\mathbf{K}_R{}' = \begin{bmatrix} k_{tx} & 0 & 0 \\ 0 & k_{ty} & 0 \\ 0 & 0 & k_{tz} \end{bmatrix} \quad (7)$$



(a) Translational spring element (b) Rotational spring element

Fig. 3 Modeling of bushing as spring elements

Stiffness matrices defined in terms of the local coordinate system can be transformed into stiffness matrices in the global coordinate system by multiplying by the transformation matrix.

$$\mathbf{K}_T = \mathbf{A}_j \mathbf{K}_T{}' \mathbf{A}_j{}^T \quad (8)$$

$$\mathbf{K}_R = \mathbf{A}_j \mathbf{K}_R{}' \mathbf{A}_j{}^T \quad (9)$$

where $\mathbf{A}_j$ is the rotational transformation matrix transforming from the local coordinate system to the global coordinate system. Since stiffness matrices are defined in the global coordinate system as Eqs. (8) and (9), elastic reaction forces are to be derived in the global coordinate system.

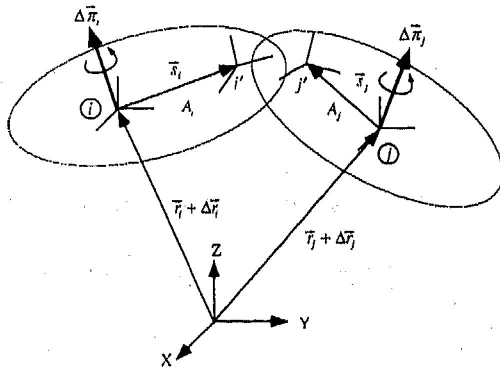


Fig. 4 Relative displpacements between body i and body j

In Fig. 4, the local coordinate system attached to each body has the same orientation as the global coordinate system. $\mathbf{s}_i$ and $\mathbf{s}_j$ are respectively position vectors from the origins of the local coordinate systems of body i and j to the local coordinates for elastic elements, i' and j' body i and body j. $\mathbf{A}_i$ and $\mathbf{A}_j$ are transformation matrices from the local coordinate systems of elastic elements to the global coordinate system of body i and body j respectively. Small displacements at i' and j' can be expressed in terms of generalized coordinates as

$$\Delta\mathbf{r}_i{}' = \Delta\mathbf{r}_i + \Delta\tilde{\boldsymbol{\pi}}_i \mathbf{s}_i = \Delta\mathbf{r}_i - \tilde{\mathbf{s}}_i \Delta\boldsymbol{\pi}_i \quad (10)$$

$$\Delta\mathbf{r}_j{}' = \Delta\mathbf{r}_j - \tilde{\mathbf{s}}_j \Delta\boldsymbol{\pi}_j \quad (11)$$

where the tilde operator '~' represents a skew-symmetric matrix.

Forces of the elastic elements exerted on the body i can be computed by premultiplying stiffness matrix $\mathbf{K}_T$ of Eq. (8) to the relative displacement $\Delta\mathbf{r}_i{}' - \Delta\mathbf{r}_j{}'$.

$$\mathbf{F}_i = -\mathbf{K}_T(\Delta\mathbf{r}_i - \Delta\mathbf{r}_j - \tilde{\mathbf{s}}_i \Delta\boldsymbol{\pi}_i + \tilde{\mathbf{s}}_j \Delta\boldsymbol{\pi}_j) \quad (12)$$

The total moment exerted on body i is the sum of moments associated with force $\mathbf{F}_i$ and pure moments generated by relative rotation between bodies.

$$\mathbf{T}_i = \tilde{\mathbf{s}}_i \mathbf{F}_i - \mathbf{K}_R({}_i\Delta\boldsymbol{\pi}_i - \Delta\boldsymbol{\pi}_j) \quad (13)$$

Similarly, force and moment exerted on body j can be expressed as

$$\mathbf{F}_j = -\mathbf{F}_i \quad (14)$$

$$\mathbf{T}_j = \tilde{\mathbf{s}}_j \mathbf{F}_j + \mathbf{K}_R({}_i\Delta\boldsymbol{\pi}_i - \Delta\boldsymbol{\pi}_j) \quad (15)$$

Arranging Eqs. (12)~(15), the relationship between force and displacement can be given as

$$\begin{Bmatrix} \mathbf{F}_i \\ \mathbf{T}_i \\ \mathbf{F}_j \\ \mathbf{T}_j \end{Bmatrix} = \left[\begin{array}{cc|cc} -\mathbf{K}_T & \mathbf{K}_T\tilde{\mathbf{s}}_i & \mathbf{K}_T & -\mathbf{K}_T\tilde{\mathbf{s}}_j \\ -\tilde{\mathbf{s}}_i\mathbf{K}_T & \tilde{\mathbf{s}}_i\mathbf{K}_T\tilde{\mathbf{s}}_i - \mathbf{K}_R & \tilde{\mathbf{s}}_i\mathbf{K}_T & -\tilde{\mathbf{s}}_i\mathbf{K}_T\tilde{\mathbf{s}}_j + \mathbf{K}_R \\ \hline \mathbf{K}_T & -\mathbf{K}_T\tilde{\mathbf{s}}_i & -\mathbf{K}_T & \mathbf{K}_T\tilde{\mathbf{s}}_j \\ \tilde{\mathbf{s}}_j\mathbf{K}_T & -\tilde{\mathbf{s}}_j\mathbf{K}_T\tilde{\mathbf{s}}_i + \mathbf{K}_R & -\tilde{\mathbf{s}}_i\mathbf{K}_T & \tilde{\mathbf{s}}_j\mathbf{K}_T\tilde{\mathbf{s}}_j - \mathbf{K}_R \end{array}\right] \begin{Bmatrix} \Delta\mathbf{r}_i \\ \Delta\boldsymbol{\pi}_i \\ \Delta\mathbf{r}_j \\ \Delta\boldsymbol{\pi}_j \end{Bmatrix}$$

$$= \begin{bmatrix} \mathbf{K}_{ii} & \mathbf{K}_{ij} \\ \mathbf{K}_{ji} & \mathbf{K}_{jj} \end{bmatrix} \begin{Bmatrix} \Delta\mathbf{q}_i \\ \Delta\mathbf{q}_j \end{Bmatrix} \quad (16)$$

$$= [\mathbf{K}]\{\Delta\mathbf{q}\}_{i,j}$$

Since the stiffness matrix K in Eq. (16) is symmetric, it is evident that the coefficient matrix of Eq. (5) is also symmetric.

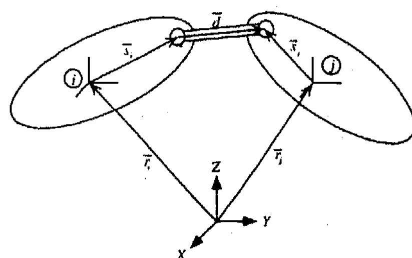


Fig.5 Displacement Constraint

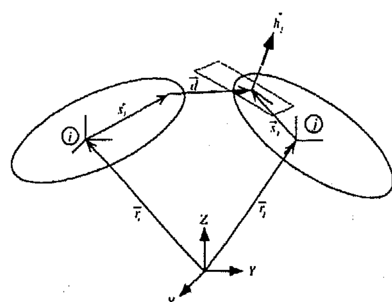


Fig. 6 Inplane constraint

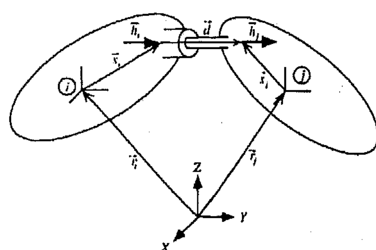


Fig. 7 Cylindrical constraint

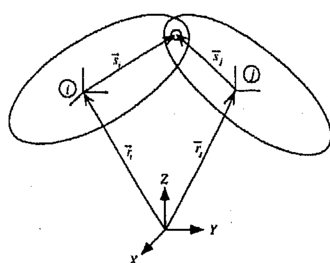


Fig. 8 Spherical constraint

Typical kinematic joints of vehicle suspension systems are shown in Fig. 5 ~ Fig. 8. Constraint Jacobians can be obtained by differentiating kinematic constraint equations. Constraint Jacobians for these joints are summarized in Table 1.

Table 1. Constraint Jacobians

Constraints	$\left[\mathbf{\Phi}_{\mathbf{q}i} \,\vdots\, \mathbf{\Phi}_{\mathbf{q}j}\right]$
Displacement	$\left[-2\mathbf{d}^T \quad 2\mathbf{d}^T\tilde{\mathbf{s}}_i \,\vdots\, 2\mathbf{d}^T \quad -2\mathbf{d}^T\tilde{\mathbf{s}}_j\right]$
Inplane	$\left[-\mathbf{h}_j^T \quad \mathbf{h}_j^T\tilde{\mathbf{s}}_i \,\vdots\, \mathbf{h}_j^T \quad -\mathbf{h}_j^T\tilde{\mathbf{s}}_j - \mathbf{d}^T\tilde{\mathbf{h}}_j\right]$
Cylindrical	$\begin{bmatrix} \mathbf{0} & \tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_i & \vdots & \mathbf{0} & -\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_j \\ -\tilde{\mathbf{h}}_i & \tilde{\mathbf{h}}_i\tilde{\mathbf{s}}_i + \tilde{\mathbf{d}}\bar{\mathbf{h}}_i & \vdots & \tilde{\mathbf{h}}_i & -\tilde{\mathbf{h}}_i\tilde{\mathbf{s}}_j \end{bmatrix}$
Spherical	$\left[-\mathbf{I} \quad \tilde{\mathbf{s}}_i \,\vdots\, \mathbf{I} \quad -\tilde{\mathbf{s}}_j\right]$

CALCULATION OF COMPLIANCE

A McPherson strut suspension system is shown in Fig. 9. Body 0 represents the ground, body 1 represents the wheel assembly, body 2 represents the strut, and body 3 represents the lower control arm. Each body is interconnected by elastic elements or kinematic joints. The coordinate where elastic elements or kinematic joints are defined is denoted as i^k, where i is the number of the body and superscript k is the number of the elastic element or kinematic joint.

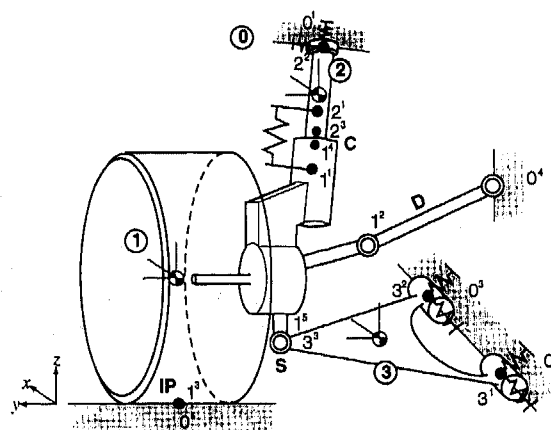


Fig. 9 Modeling of a McPherson strut suspension

Table 2. Elastic elements

index	coordinate (i - j)
1	$1^1 - 2^1$
2	$2^2 - 0^1$
3	$3^1 - 0^2$
4	$3^2 - 0^3$

Table 3. Constraint elements

index	coordinate (i - j)	type
1	$1^2 - 0^4$	D
2	$1^3 - 0^5$	IP
3	$1^4 - 2^3$	C
4	$1^5 - 3^3$	S

Connection by elastic elements is summerized at Table 2. Body 1 is connected to body 2 with a linear spring. Body 2 and body 0 are connected with a strut mount bushing. Body 3 is connected to body 0 at two points with bushings.

Kinematic constraints are summerized in Table 3. A displacement constraint is imposed on body 1 to simulate bump and rebound motion of the wheel assembly. Body 1 is connected to body 0 with an inplane constraint. The constraint between body 1 and body 2 is a cylindrical joint. Body 1 and body 3 are connected by a spherical joint.

Applying Eq. (5) elastokinematic equation of the system can be written as

$$\left[\begin{array}{ccc:c} \mathbf{K}_{ii}^1 & \mathbf{K}_{ij}^1 & 0 & \\ \mathbf{K}_{ji}^1 & \mathbf{K}_{jj}^1+\mathbf{K}_{ii}^2 & 0 & \mathbf{\Phi}_q^T \\ 0 & 0 & \mathbf{K}_{ii}^3+\mathbf{K}_{ii}^4 & \\ \hdashline \mathbf{\Phi}_{qi}^1 & 0 & 0 & \\ \mathbf{\Phi}_{qi}^2 & 0 & 0 & \mathbf{0} \\ \mathbf{\Phi}_{qi}^3 & \mathbf{\Phi}_{qj}^3 & 0 & \\ \mathbf{\Phi}_{qi}^4 & 0 & \mathbf{\Phi}_{qj}^4 & \end{array}\right] \begin{Bmatrix} \Delta\mathbf{q}_1 \\ \Delta\mathbf{q}_2 \\ \Delta\mathbf{q}_3 \\ \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \boldsymbol{\lambda}_3 \\ \boldsymbol{\lambda}_4 \end{Bmatrix} = \begin{Bmatrix} -\mathbf{f}_1 \\ -\mathbf{f}_2 \\ -\mathbf{f}_3 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \quad (17)$$

where $\Delta\mathbf{q}_1$, $\Delta\mathbf{q}_2$ and $\Delta\mathbf{q}_3$ are generalized coordinates of body 1, body 2 and body 3, respectively. Body 0 of the ground has no degree of freedom. Equation (17) can be solved by any linear system solving method.

Table 4. Displacements of wheel due to lateral force

F_{1y} = 1000 N

$\Delta\mathbf{q}_1$	Linear Analysis	ADAMS
Δx (mm)	0.3065	0.3084
Δy (mm)	0.1579	0.1583
Δz (mm)	0.0000	-0.0003
$\Delta\pi_x$ (deg)	0.0409	0.0406
$\Delta\pi_y$ (deg)	-0.0191	-0.0189
$\Delta\pi_z$ (deg)	-0.0470	-0.0473

Compliance of the suspension system can be calculated under the lateral or longitudinal forces at the tire contact point. When the lateral force F_{1y} = 1000 N is applied, displacements of the wheel are computed. The results of $\Delta\mathbf{q}_1$ are shown in Table 4. Results are very consistent with those of ADAMS.

SENSITIVITY ANALYSIS

Sensitivity equations can be obtained by differentiating Eq. (5) with respect to a design variable associated with compliance characteristics. Representing a design variable by the symbol b, sensitivity equations can be written as

$$\begin{bmatrix} \mathbf{K} & \mathbf{\Phi}_q^T \\ \mathbf{\Phi}_q & 0 \end{bmatrix} \begin{Bmatrix} \Delta\mathbf{q}_b \\ \boldsymbol{\lambda}_b \end{Bmatrix} = \begin{Bmatrix} -\mathbf{K}_b\Delta\mathbf{q} \\ \mathbf{0} \end{Bmatrix} \quad (18)$$

where subscript b means partial differentiation. Since the coefficient matrix of Eq. (18) is the same as that of Eq. (5), only $\mathbf{K}_b$ needs to be calculated. $\mathbf{K}_b$ can be computed by directly differentiating the stiffness matrix of Eq. (16) as

$$\mathbf{K}_b = \begin{bmatrix} -\frac{\partial \mathbf{K}_T}{\partial b} & \frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_i & \frac{\partial \mathbf{K}_T}{\partial b} & -\frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_j \\ -\tilde{\mathbf{s}}_i\frac{\partial \mathbf{K}_T}{\partial b} & \tilde{\mathbf{s}}_i\frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_i-\frac{\partial \mathbf{K}_R}{\partial b} & \tilde{\mathbf{s}}_i\frac{\partial \mathbf{K}_T}{\partial b} & -\tilde{\mathbf{s}}_i\frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_j+\frac{\partial \mathbf{K}_R}{\partial b} \\ \frac{\partial \mathbf{K}_T}{\partial b} & -\frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_i & -\frac{\partial \mathbf{K}_T}{\partial b} & \frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_j \\ \tilde{\mathbf{s}}_j\frac{\partial \mathbf{K}_T}{\partial b} & -\tilde{\mathbf{s}}_j\frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_i+\frac{\partial \mathbf{K}_R}{\partial b} & -\tilde{\mathbf{s}}_i\frac{\partial \mathbf{K}_T}{\partial b} & \tilde{\mathbf{s}}_j\frac{\partial \mathbf{K}_T}{\partial b}\tilde{\mathbf{s}}_j-\frac{\partial \mathbf{K}_R}{\partial b} \end{bmatrix} \quad (19)$$

where $\frac{\partial \mathbf{K}_T}{\partial b}$, $\frac{\partial \mathbf{K}_R}{\partial b}$ are calculated from Eqs. (8) and (9) as

$$\frac{\partial \mathbf{K}_T}{\partial b} = \mathbf{A}_j \frac{\partial \mathbf{K}_T'}{\partial b} \mathbf{A}_j^T \quad (20)$$

$$\frac{\partial \mathbf{K}_R}{\partial b} = \mathbf{A}_j \frac{\partial \mathbf{K}_R'}{\partial b} \mathbf{A}_j^T \quad (21)$$

Entries of matrices $\frac{\partial \mathbf{K}_T'}{\partial b}$ and $\frac{\partial \mathbf{K}_R'}{\partial b}$ are either 0 or 1 according to the choice of the design variable. For example, if b = K_x in Eq. (6), $\frac{\partial \mathbf{K}_T'}{\partial b}$ becomes

$$\frac{\partial \mathbf{K}_T'}{\partial b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (22)$$

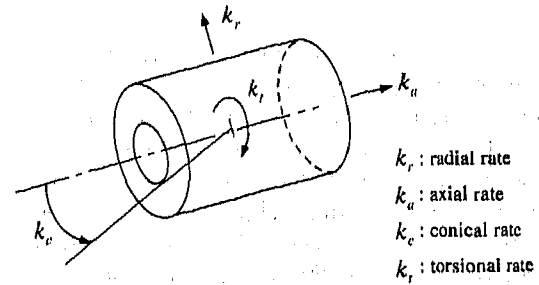


Fig. 10 Spring rates of cylindrical type bushing

Bushing elements of the McPherson strut suspension system as shown in Fig. 9 are strut mount bushing, front lower arm bushing, and rear lower arm

bushing. In this study, stiffnesses of these three bushings are selected as design variables. Since the stiffness of bushings in the radial direction is generally constant for a cylindrical type bushing, four design variables are to be considered as represented in Fig. 10.

There can be many performance variables associated with compliance characteristics of suspension system. Here, toe change G_1, camber change G_2, and lateral deflection G_3 due to the lateral force are considered. Also toe change G_4, and longitudinal deflection G_5 due to the longitudinal force are chosen as performance variables. Sensitivities for these performance targets with respect to each design variable are shown at Table 5.

Table 5. Sensitivities of compliance characteristics

Design Variables \ Sensitivity	toe change by lateral force G_1	camber change by lateral force G_2	lat. disp. change by lateral force G_3	toe change by long. force G_4	long. disp. change by long. force G_5
k_{r2}	-0.0303	-0.0214	0.0455	0.0011	0.0412
k_{a2}	0.0000	0.0000	0.0000	0.0000	0.0000
k_{c2}	0.0000	0.0000	0.0000	0.0000	0.0000
k_{t2}	0.0000	0.0000	0.0000	0.0001	-0.0012
k_{r3}	0.0129	-0.0031	-0.0326	0.0131	-0.0449
k_{a3}	-0.0001	0.0000	0.0000	-0.0376	-0.7593
k_{c3}	0.0000	0.0000	0.0000	0.0000	0.0000
k_{t3}	0.0000	0.0000	0.0000	0.0000	0.0000
k_{r4}	0.0001	-0.0004	-0.0036	0.0010	-0.2963
k_{a4}	-0.0001	0.0000	0.0000	-0.0376	-0.7593
k_{c4}	0.0000	0.0000	0.0000	0.0000	0.0000
k_{t4}	0.0000	0.0000	0.0000	0.0000	0.0000

As shown in Table 5, sensitivities of rotational stiffness of the bushings are small enough to be neglected. Radial rate k_{r2} of strut mount has significant effects on all performance variables, however, effects of axial rate k_{a2} is very small. k_{r2} of strut mount can be considered as the main design variable which affects lateral compliance characteristics. The design variables most sensitive to longitudinal compliance are k_{a3} and k_{a4}, which are axial rates of front and rear lower arm bushings, respectively.

OPTIMIZATION

Optimization of bushing characteristics is to be performed taking five performance variables, G_1, ..., G_5, into consideration. Gradients needed to perform an optimization process are provided by the derived sensitivity equations. The goal attainment method is adopted to solve the multi-objective optimization problem. This method seeks the solution by converting vector type multi-objective function $\mathbf{G(x)}$ to a single objective function using a scalar variable γ. This optimization problem can be stated as

$$\begin{aligned} &\textit{Minimize} \quad \gamma \\ &\textit{subject to} \quad \mathbf{G(x)} - \mathbf{W}\gamma \le \mathbf{G}^* \end{aligned} \qquad (23)$$

where $\mathbf{G}^*$ is a set of design goals, and $\mathbf{W}$ is the weighting factor, which is proportional to design goal $\mathbf{G}^*$. The results of optimization using the sequential quadratic programming are shown in Table 6.

Table 6. Optimization results

Performance /parameter	initial state	optimal state (goal, G*)
G_1	-0.0470	0.0055 (0.01)
G_2	0.0409	0.0145 (0.01)
G_3	0.1579	0.0055 (0.01)
G_4	-0.0404	0.0145 (0.01)
G_5	1.6709	0.7460 (1.00)
k_{r2}	1000	1870
k_{a2}	600	604
k_{c2}	1800000	1800000
k_{t2}	0	0
k_{r3}	6000	42565
k_{a3}	500	806
k_{c3}	200000	200001
k_{t3}	100000	100000
k_{r4}	2000	64444
k_{a4}	600	906
k_{c4}	300000	300001
k_{t4}	50000	50000

Values of design goal $\mathbf{G}^*$ are all set to 0.01 except for G_5^*=1.0. The design goals may be chosen differently by practical considerations. Optimal compliance characteristics are achieved near the design goal. Stiffness parameters, which have greater contribution to compliance characteristics as shown in Table 4, have changed more than other parameters. In Fig. 11 and Fig. 12, iteration history of G_1, G_5 of performance variables are shown.

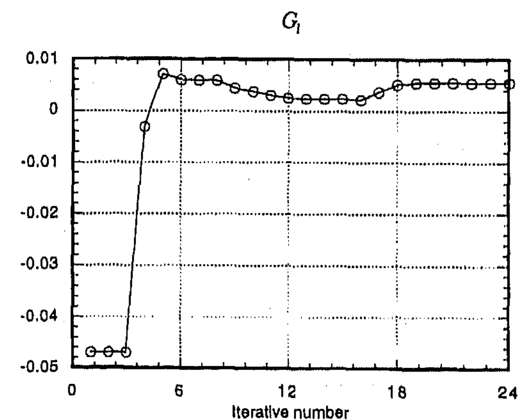


Fig. 11 Iteration history of G_1

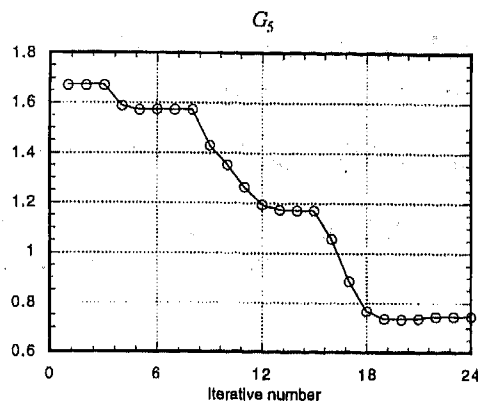


Fig. 12 Iteration history of G_5

CONCLUSION

Compliance characteristics of suspension systems have direct influence on ride and handling of a vehicle. In this study, linear elastokinematic equations have been derived to compute compliance characteristics of suspension systems. It is shown that these equations can be applied to any suspension system and are verified through the comparison with ADAMS results. Elastokinematic equations presented in this paper turned out to be not only very simple to implement, but also in close agreement with the more complex solution offered by ADAMS.

By directly differentiating the linear elastokinematic equations with respect to design variable associated with bushing stiffness, sensitivity equations are obtained. Then, sensitivity analysis with respect to stiffness of bushing elements was performed. Based on the sensitivity analysis, optimization of bushing characteristics is carried out, and the optimal solution subject to the various design targets are obtained.

REFERENCES

1. R.T. Bundorf, "The Influence of Vehicle Design Parameters on Characteristic Speed and Understeer",SAE Paper NO. 670078, 1967

2. R.T. Bundorf, "The Cornering Compliance Concept for Description of Vehicle Directional Control Properties", SAE Paper NO. 760713, 1976

3. Y. Tsukuda, Y. Tsubota, H. Tonomura, H. Noguchi, "Development of a New Multi-Link Rear Suspension", SAE Paper NO. 881774, 1988

4. J.R. Ellis, S.C. Burns, W.R. Garrott, S.C. Bell, " The Design of a Suspension Parameter Measurement Device", SAE Paper NO. 870576, 1987

5. P.E. Nikravesh, M. Srinivasan, "Generalized Co-Ordinate Partitioning in Static Equilibrium Analysis of Large-Scale Mechanical Systems", Int. J. Numerical Eng. , Vol. 21, 1985

6. K. Tachikawa, H. Sakai, "A Multi-Objective Optimization Method by Sequential Linear Programming", SAE Paper NO. 880887, 1988

7. P.M. Riede, R.L. Leffert, W.A. Cobb, " Typical Vehicle Parameters for Dynamics Studies Revised for the 1980's", SAE Paper NO. 840561, 1984

8. G. Gallas, B. Renzo, "New Products Developed to Improve the Isolation Qualities of Automobile Suspension Systems: "Torsidyne" and "Atadyne"", SAE paper NO. 851651, 1985

9. A.B. Zdunek, M. Bercovier, "Numerical Evaluation of Finite Element Methods for Rubber Parts", SAE Paper NO. 860817, 1986

10. "Vehicle Dynamics Terminology", SAEJ670e, Society of Automotive Engineers, Warrendale, PA

11. P.E. Nikravesh, Computer-Aided Analysis of Mechanical Systems, Prentice-Hall, Inc., 1988

12. ADAMS User's Manual, Mechanical Dynamics, Inc., Ann Arbor MI, 1985

970106

An Investigation of the Feasibility of Refrigerant Charge Loss Detection Using Low Cost Measurements

C. D. Collins, N. R. Miller, and W. E. Dunn
University of Illinois at Urbana-Champaign

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The feasibility of automatically detecting refrigerant charge loss in mobile air conditioning (MAC) systems by analyzing inexpensive dynamic measurements was studied. An indicator of the refrigerant inventory of the evaporator was developed. This measure, termed Time to Temperature Turning (TTT), is based on dynamic measurement of the evaporator outlet refrigerant temperature, and correlates strongly with charge level. TTT correlated well with clutch cycling behavior, a metric which is employed in current shop diagnostic practice to indicate refrigerant charge loss.

Laboratory data were generated from a factorial experiment design on the following factors: condenser air inlet temperature, condenser air flow rate, evaporator air inlet temperature, compressor speed, and refrigerant charge. Experiments to date were conducted with a dry evaporator.

INTRODUCTION

This paper reports work to date on microprocessor based refrigerant charge loss detection at the Air Conditioning and Refrigeration Center (ACRC) of the University of Illinois. The data in this report were obtained from late model original equipment manufacturer (OEM) MAC systems employing fixed displacement compressors, orifice tube expansion devices, and clutch cycling capacity control. These systems were installed and tested in the ACRC Mobile Air Conditioning (MAC) Project test facility.

MOTIVATION -. Refrigerant leakage from MAC systems contributes to environmental pollution and leads to reduced cooling system performance. Compressor lubrication in MAC refrigeration systems is provided by oil that circulates with the refrigerant. Therefore, inadequate refrigerant circulation due to low refrigerant charge may cause reduced compressor lubrication. Reduced compressor lubrication can lead to premature mechanical failure of the compressor. Under certain operating conditions, in systems employing a suction pressure controlled compressor clutch, refrigerant charge loss may induce frequent clutch cycling. Excessive clutch cycling may cause premature clutch failure and may adversely affect driveability, fuel economy and engine idling control. Early detection of charge loss would be highly desirable. The goals of the MAC project include addressing the clear need for the development of a reliable method for detecting refrigerant charge loss in MAC systems

Early detection and diagnosis of refrigerant charge loss would permit time for automatic or manual corrective actions to be taken prior to significant deterioration in performance or catastrophic system failure. A commercial diagnostic system should possess the following desirable properties: (1) be capable of accurately diagnosing a reduction in refrigerant charge, (2) require a minimal number of sensors, (3) be simple to install and reliable in operation, and (4) be low cost.

The benefits of a microprocessor based refrigerant charge leakage detection system include (1) programmability and (2) performance data analysis and record-keeping. The use of microprocessor technology in luxury and mid-size vehicles is already widespread, and is likely to grow in the future. Microprocessor based engine control systems, automatic brake systems and environmental control systems currently log and use data such as engine speed, vehicle speed, exterior air temperature, and interior air temperature. One of the goals of this project was to utilize data already available on the vehicle instrumentation bus. Clearly, such refrigerant charge diagnostic systems would first be introduced in models which already employ microprocessors.

BASIS FOR STUDY - The problem of detecting and diagnosing a loss in refrigerant charge is twofold. First, a departure from some normal, or expected, operating condition must be determined. Second, the cause of this abnormal operating state must be uniquely determined. Refrigeration systems pose a difficult problem because there are a considerable number of factors which influence system performance, and changes in these factors may produce similar system

responses. Diagnostic schemes for MAC systems are particularly difficult to develop because mobile systems rarely operate in steady state.

Both theory and practice show that the mass of refrigerant present in a refrigeration system affects the system's performance. An example from the open literature is a 1988 experimental study which investigated the effect of varying refrigerant charge in a clutch cycling, expansion valve MAC system [1]. Steady-state experiments were performed which documented the effect of reducing refrigerant charge on several key refrigerant side and air side system parameters. Significant changes in the system parameters were observed. We extended this experimental principle to studies of refrigerant system dynamic performance.

Recent literature documents several efforts made to analyze stationary air conditioning and refrigeration system performance using multiple temperature and pressure measurements [2,3,4]. The measurements are processed using one of several computationally intensive techniques (e.g. neural networks, statistical pattern recognition, or expert systems). A refrigerant monitoring device based on mechanical measurement of the quality of the refrigerant at the condenser outlet was reported in [5]. We tried to synthesize these approaches by searching for a robust, non-invasive measure of system performance requiring minimal computational overhead.

EXPERIMENT

APPROACH - The problem of refrigerant charge loss detection was considered for fixed displacement, cycling clutch, orifice tube air conditioning systems operating under a specific, limited set of operating conditions. That is, we considered the problem as one of detecting the change in refrigerant system performance, due to refrigerant charge loss, over a limited range of operating conditions which are encountered frequently in vehicle operation. This concept is an extension of current shop diagnostic practice, in which MAC systems are tested under specific operating conditions. In the shop, the existence and extent of refrigerant charge loss is determined by comparing the system response at specified conditions to reference data provided in the form of plots of condenser inlet temperature versus clutch cycling period.

TEST FACILITY - The facility used in this study is located in the Mechanical Engineering Laboratory at the University of Illinois. The facility permits the detailed study of the refrigerant side and air side performance of a modified, late model OEM MAC system under tightly controlled environmental conditions.

The refrigeration system consists of factory standard components from the air conditioning system of the 1994 Ford Crown Victoria: a fixed displacement compressor, an orifice tube expansion device, a condenser, an accumulator and an evaporator. The heat exchangers are installed in separate wind tunnels in the test facility. A copper piping system connects all the system components in a closed loop and permits the instrumentation of the refrigeration system. Refrigerant temperature and pressure are measured upstream and downstream of each major component, and refrigerant mass flow rate is measured in the vapor and liquid lines. The facility control system permits the control of the compressor speed and clutch operation.

The condenser and evaporator are housed in independent air flow loops. The facility control system allows the air flow rate through, and inlet air temperature of, each heat exchanger to be tightly controlled. This paper reports testing which was performed with a dry evaporator coil. The addition of a humidity control system to the evaporator air loop during the fall of 1996 now allows the control of evaporator air loop relative humidity. The air loops are equipped with air temperature, relative humidity and air flow rate instrumentation.

EXPERIMENT DESIGN - Diagnosis of refrigerant charge loss requires that the effect on system performance of a change in refrigerant charge (mass), m_{ref}, be evaluated with respect to the effects of varying other parameters which affect the refrigeration system operating state. These parameters include condenser air inlet temperature (T_{cai}), condenser air volume flow rate (Q_{cai}), evaporator air inlet temperature (T_{eai}), evaporator air inlet relative humidity (Rh_{eai}), evaporator air volume flow rate (Q_{eai}), and compressor speed (N_{comp}).

The baseline charge of the system was a crucial parameter we had to determine experimentally. The system performance with the initial charge would serve as the baseline to which we would compare system performance with reduced charge. We determined the initial charge for our modified system by using an industry standard charge optimization procedure. The charge we determined for the test facility differed from the charge specified by the manufacturer for the factory system. We expected this difference, and it is attributable to the refrigeration system volume added by the piping sections which support installing the system in the test facility. We also used an orifice tube diameter that was one standard size smaller (.052" vs. .057") than that specified for the factory system. We expected that these differences between the experimental system and the factory system would affect the steady state operating points of the system and the baseline system charge. However, we believe that the *dynamic* effects we observed in our study of charge reduction will be preserved. This hypothesis requires further experimental study.

We eliminated evaporator air flow rate from the study by fixing this parameter at its maximum design value. Evaporator air flow rate does have a significant effect on performance; however, we believe it is reasonable to assume that a typical mobile air conditioning system is operated in the high blower mode frequently enough to provide sufficient opportunity for an automatic diagnostic system to collect data. It would also be a simple matter to extend the technique demonstrated in this paper to include evaporator air flow rate (or equivalently, evaporator blower speed) as an additional parameter.

Evaporator air inlet relative humidity is known to have a major impact on system dynamic performance. At the time of this study, we could not control this parameter, so the study was conducted with a dry coil. The results presented herein are valid in dry climates. Further, we intend to study the effects of inlet humidity in the near future.

We limited the study to small (~ ±10%) changes in each variable of interest with respect to a nominal operating state. Our goal was to determine if relatively small losses of refrigerant charge could be detected based on small, repeated variations of system operating parameters such as would occur during normal highway driving. We defined the nominal operating state as follows: N_{comp}, 2800 rpm; Q_{cai}, 1800 cfm; T_{cai}, 110 °F; T_{eai}, 70°F; and m_{ref}, 2.85 lbm. Q_{eai} was set at 330 cfm for all tests. We based our setpoints on typical highway driving data provided by our industrial sponsors. The high pressure (clutch *on*) and low pressure (clutch *off*) cycling setpoints were set to 42 psig and 25 psig, respectively.

A 2 level, 5 factor, fractional factorial experiment design was employed to investigate the system performance. T_{cai}, Q_{cai}, T_{eai}, N_{comp}, and m_{ref} were varied, and we measured the corresponding responses of the total clutch cycling period, the clutch duty cycle, and Time to Temperature Turning (TTT). TTT is a temperature-based, indirect indication of the evaporator refrigerant inventory, and will be fully explained later. The experiment design is presented in Tables 1 and 2.

RESULTS

We found that m_{ref} is the dominant factor governing clutch cycling period, over the range of variables we studied. Tables 3 and 4 present the results of the designed experiment described in Tables 1 and 2.

Table 1. Key for Coded Experiment Design Notation

Tag	Factor		+	-
1	Condenser Air Flow	[cfm]	1800	1600
2	Condenser Air Inlet Temperature	[°F]	110	100
3	Evaporator Air Inlet Temperature	[°F]	70	65
4	Refrigerant Charge	[lbm]	2.85	2.55
5	Compressor Speed	[rpm]	2800	2500

Table 2. Coded Experiment Design

	Factor				
Run	1	2	3	4	5
1	-	-	-	-	+
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	+
5	-	-	+	-	-
6	+	-	+	-	+
7	-	+	+	-	+
8	+	+	+	-	-
9	-	-	-	+	-
10	+	-	-	+	+
11	-	+	-	+	+
12	+	+	-	+	-
13	-	-	+	+	+
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	+

Boldface numbers in the column headings in Table 3 and in the body of Table 4 correspond to the factor tag numbers defined in Table 1.

Table 3 records the factor levels, the number of complete clutch cycles (N_c), the mean clutch cycling period (T_c), and the standard deviation (s) of the mean clutch cycling period. Table 4 reports the main effects and two factor interactions. Each experimental run lasted for 600 seconds.

We can take advantage of the experiment design and draw some preliminary conclusions from the data in Table 3. Notice that the factor levels in each of runs 1 through 8 differ from those in runs 9 through 16 in only two factors: m_{ref} and N_{comp}. Any significant difference in the response observed between, say, run 2 and run 10, was due to (1) the change in m_{ref}, or (2) the change in N_{comp}, or (3) the combined effect of both changes. Similarly, each pair of neighboring runs, starting with runs 1 and 2, differ only in N_{comp} and Q_{cai}. Table 3 clearly shows that the effect of changing of N_{comp} and Q_{cai} in tandem is small, while the effect of changing m_{ref} and N_{comp} in tandem is, in general, large. This indicates that the main effect of m_{ref} may be quite significant. Table 4 presents the calculated main effects and two-factor interactions. The calculated mean period and the standard deviation of the mean are provided at the beginning of Table 4.

The most significant main effect is due to m_{ref} (**4**), followed by T_{eai} (**3**) and T_{cai} (**2**). In addition, the heat exchanger inlet temperatures exhibit a strong 2 factor interaction (**23**). Over the range of variables studied, N_{comp} and Q_{cai} did not significantly affect clutch cycling period. The main effect due to m_{ref} is not strictly "significant" according to the rules governing interpretation of statistics computed in a designed experiment. This typically requires the main effect (or any interaction) to exceed the standard deviation by at least a factor of 2. However, the data clearly show the relative magnitude of the effects that each controlled variable (factors) had on the clutch cycling period.

Further analysis of the data showed that the *on* portion of the clutch duty cycle was significantly more sensitive than the *off* portion to changes in the refrigerant charge level. In fact, the *off* portion of the duty cycle was relatively insensitive to changes in the refrigerant charge level. The *off* cycle data are reported in Table 5.

During each *on* cycle, notice that the slope of the evaporator refrigerant outlet temperature (T_{ero}) exhibits a clear change. We associated this feature of the outlet temperature with the initial pulldown and termed it the Time to Temperature Turning (TTT). We calculated mean TTT for all runs in all experiments, and evaluated the effect of different factors on TTT. The results are presented in Table 6.

We noticed that T_{ero} behaved in a predictable way in response to the transient imposed by the cycling clutch. At the beginning of each *on* cycle, T_{ero} starts to drop. This effect is due to the saturation pressure drop associated with the reduced refrigerant density in the low pressure side piping caused by the pumping action of the compressor. At some point after the compressor comes on, the slope of T_{ero} levels off, or turns completely. This indicates that the refrigerant exiting the evaporator is superheated. We believe that this behavior is associated with the redistribution of

Table 4. Main Effects and Two Factor Interactions

Mean period = 33.1 sec, s = 11.62 sec

Factor	Effect
1 =	-0.07 sec
2 =	8.38 sec
3 =	6.36 sec
4 =	14.1 sec
5 =	-0.07 sec
12 =	-0.56 sec
13 =	-0.20 sec
14 =	1.42 sec
15 =	5.03 sec
23 =	10.1 sec
24 =	6.32 sec
25 =	-2.77 sec
34 =	3.16 sec
35 =	5.29 sec
45 =	0.96 sec

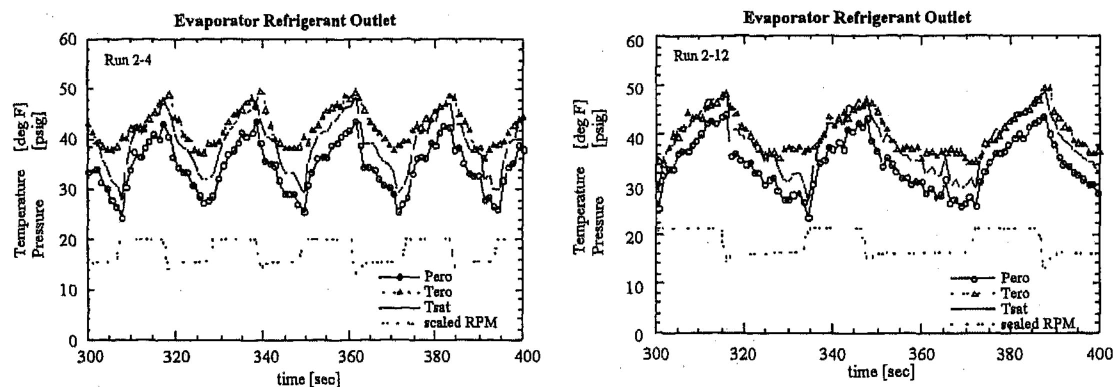
Table 5. Analysis of Off-Cycle Time for all runs

	Mean [sec]	Std Dev [sec]
Grand Avg	10.4	2.2
Charge (+)	11.6	2.1
Charge (-)	9.4	1.7
Tcai (+)	10.0	2.1
Tcai (-)	10.6	2.2
Teai (+)	9.2	1.8
Teai (-)	11.2	2.0
Charge (+)/Teai (-)	12.1	2.1

Table 3. Experiment Results

Run	1	2	3	4	5	N_c	T_c [sec]	s [sec]
1	-	-	-	-	+	22	28.0	3.31
2	+	-	-	-	-	25	24.2	0.95
3	-	+	-	-	-	25	24.4	3.18
4	+	+	-	-	+	30	21.1	1.57
5	-	-	+	-	-	25	24.8	1.27
6	+	-	+	-	+	27	23.0	1.31
7	-	+	+	-	+	20	29.9	4.97
8	+	+	+	-	-	17	32.8	1.40
9	-	-	-	+	-	20	31.2	4.76
10	+	-	-	+	+	15	39.4	0.86
11	-	+	-	+	+	17	34.6	2.50
12	+	+	-	+	-	17	36.1	2.35
13	-	-	+	+	+	20	30.4	3.34
14	+	-	+	+	-	20	29.9	2.47
15	-	+	+	+	-	9	61.4	40.72
16	+	+	+	+	+	6	57.7	41.99

Figure 1. Evaporator Outlet Refrigerant Transient Temperature and Pressure



refrigerant which takes place each time the compressor is engaged and disengaged from the refrigerant system. We observed that this behavior, which we termed Time to Temperature Turning, is strongly correlated to refrigerant charge. Figure 1 illustrates this concept. The right hand and left hand plots of Figure 1 differ only in the amount of refrigerant charge present in the system. The right hand plot presents data collected at normal charge. (Note that the scaled RPM signal falls when the compressor clutch is engaged.)

Table 6 shows that TTT is strongly influenced by refrigerant charge, and that the difference is statistically significant at the .01 level, if we use the low charge level as a reference. That is, the mean TTT at high charge is different from the mean TTT at low charge by more than 3 standard deviations of the mean TTT at low charge.

DISCUSSION

Refer to Table 4, and notice again that the largest main effect is due to factor 4, *mref*. To be rigorously statistically significant at, for example, the .05 level, a main effect or interaction should differ from the mean response by greater than twice the sample standard deviations of the mean response. Main effect 4 is approximately 1.2 times the sample standard deviation of the mean response. From a qualitative standpoint, however, the level of refrigerant charge clearly affects system response. The analysis of factorial designs makes assumptions about the smoothness and linearity of the response function. These assumptions are not likely to be applicable to the clutch cycling period. The arithmetical statistical analysis must be supplemented with prior physical knowledge and engineering reasoning. Table 4 clearly illustrates the strong effect that refrigerant charge has on clutch cycling period and the variance of the clutch cycling period. The data imply that the clutch cycling period of this air conditioning system is a reliable indicator of the relative amount of refrigerant charge in the system.

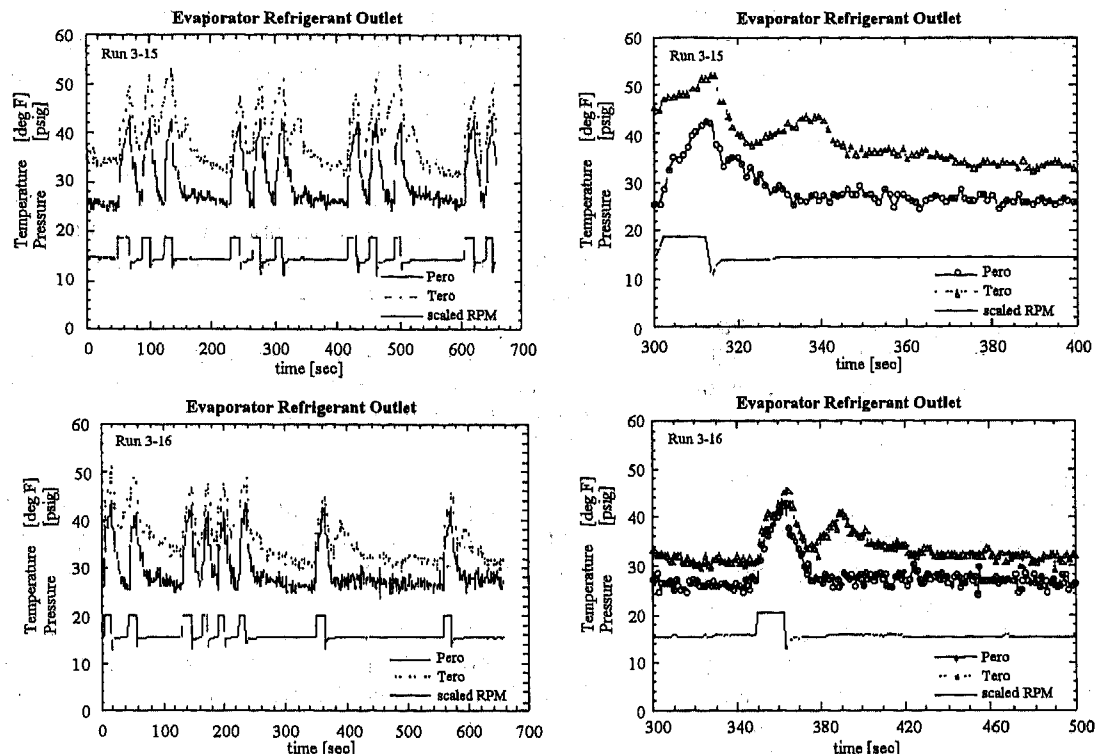
Now consider the main effect of factors 2 and 3, *Tcai* and *Teai*. Both factors exhibit noticeable main effects, and together exhibit the largest two-factor interaction. When a significant interaction exists, no positive statement can be made about the significance of the corresponding main effects. The subject factors must be considered jointly. The physical interpretation of the **23** interaction is that the higher inlet temperatures raise the steady state condensing and evaporating pressure of the refrigerant. In addition, *Tcai* has a strong effect on the subcooling of the liquid refrigerant. As *Tcai* increases, condensing pressure and temperature increase. As condensing pressure increases, the latent heat of vaporization of R134a decreases, causing the subcooling of the refrigerant to increase for a given amount of heat rejected at the condenser. Raising either condenser or evaporator air temperature independently causes the evaporating pressure to rise; raising both temperatures causes the evaporating pressure to rise dramatically. As the evaporating pressure rises, it takes longer for the evaporator outlet pressure to "pull down" to the low pressure clutch cycling setpoint.

Table 6. Time to Temperature Turning (TTT) at Evaporator Outlet

	Mean [sec]	Std Dev [sec]
Grand avg	7.8	1.9
Charge(+)	9.2	1.9
Charge(-)	6.5	0.8
Tcai(+)	7.5	1.8
Tcai(-)	7.9	2.0
Teai(+)	7.6	1.6
Teai(-)	7.9	2.1

Refer now to Figure 1. Compare the behavior of evaporator temperature and pressure, between runs 2-4 and 2-12, during each *on* cycle. In both cases, the

Figure 2. Time history of evaporator outlet refrigerant temperature and pressure.



refrigerant temperature and pressure initially drop rapidly as the compressor evacuates the low pressure side of the refrigerant loop. In the case of run 2-4, there is insufficient refrigerant inventory in the high pressure side of the loop to reflood the evaporator and stabilize the outlet pressure. The outlet pressure drops rapidly to the low pressure setpoint and the compressor clutch trips. In run 2-12, there is sufficient refrigerant inventory in the high pressure side, at the end of the *off* cycle, to reflood the evaporator. The compressor repressurizes the condenser, and the mass flow rate of refrigerant through the orifice tube increases. After the initial evacuation, the evaporator refloods and reduces the rate of change of the decrease in evaporator outlet pressure.

During each *on* cycle, the slope of the evaporator outlet temperature signal exhibits a clear change. We explain this behavior by considering the relationship between the measured refrigerant temperature and the saturation temperature of the refrigerant. TTT is very closely related to the point at which the refrigerant state at the evaporator outlet changes from a high-quality, two-phase mixture to a single-phase, slightly superheated vapor. This is a manifestation of the refrigerant inventory available to reflood the evaporator at the end of the *off* cycle. At lower charges, the evaporator does not reflood as quickly as at higher charges, and the dry out point migrates upstream in the evaporator.

We have now described two complementary methods of measuring features of the clutch cycling transient that we can use to make diagnostic inferences about the state of refrigerant charge in the system: clutch cycling period and TTT.

Earlier, we observed that the combined effect of air side inlet temperatures was significant. The data of runs 15 and 16, in which both air inlet temperatures and the charge were set to their high level, clearly demonstrate this interaction. Refer again to Table 3 and notice that the clutch cycling period is almost twice as long in these runs as for any other run. The period, however, was highly irregular, as we illustrate in Figure 2 (see Run 3-15 and Run 3-16).

The performance of the evaporator in runs 15 and 16 exhibits characteristics of both the low charge and high charge performance of other air inlet temperature combinations. There are isolated groups of fairly regular cycling of 20-30 seconds duration. There are also long periods during which the evaporator refrigerant outlet pressure remains steady at just above the low pressure setpoint of 25 psig. At higher T_{cai}, the enthalpy of the refrigerant in the liquid line is higher than it is at lower T_{cai}. We infer that the enthalpy of the two-phase refrigerant at the inlet to the evaporator will be higher, and the evaporating pressure will be higher. At higher T_{eai}, it takes longer for the evaporator to "pull down" to a low temperature and pressure. At higher charges, the evaporator becomes an entirely two-phase device. The greater refrigerant inventory in the evaporator, coupled with higher condensing and evaporating pressures driven by higher air inlet temperatures, causes the evaporator outlet pressure to remain above the low pressure cycling setpoint. The clutch is probably tripped intermittently due to random fluctuations in the evaporator outlet pressure when the

pressure is close to the setpoint. These fluctuations are associated with the random fluctuation in the location of the dryout point in the evaporator. This fluctuation of the dryout point has been observed before, and was reported by Gruhle and Isermann [6] and others.

CONCLUSION

For the range of factors studied, we found that refrigerant charge exerted the strongest influence on the total clutch cycling period. Further, we found that the *on* portion of the compressor duty cycle and the Time to Temperature Turning (TTT) were highly sensitive to the amount of refrigerant charge in the system, and to the combined effect of heat exchanger air inlet temperatures. We found that the *off* portion of the compressor duty cycle was less affected by the amount of charge. We suggest that the clutch cycling period (or, alternatively, the *on* cycle time) and TTT together constitute a basis for the detection of refrigerant charge loss in mobile air conditioning systems. We found that the evaporator air inlet temperature and the condenser air inlet temperature exert a strong combined effect on the total clutch cycling period, and we suggest that these temperatures would have to be sensed, as well.

We found that, over the range of factors studied, compressor speed and condenser air flow rate had virtually no effect on the measured responses. We plan to verify this remark by performing experiments from which unconfounded second order interactions can be computed, and which include the effect of variations in evaporator air inlet relative humidity.

FUTURE WORK - This research opens a number of questions. The data imply that refrigerant charge loss can be separated and identified uniquely from among a number of factors which affect refrigerant system performance. This conclusion needs to be further bolstered by experiments conducted over both a wider range of operating conditions, and within a more detailed testing scheme, for example a complete factorial design.

The TTT data were generated by hand. An automated method to generate this data is necessary. One such method might employ a microprocessor which stores a number of temperature readings and calculates the slope, and rate of change of the slope, of the temperature signal on a running basis.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Mike Whitchurch, Eric Wandell, Joe Kavanagh, Tara Hemami, the ACRC, and Ms. Valerie Allan and Mr. David Zietlow of Ford Motor Company.

REFERENCES

[1] M. Tanino, T. Takano, and H. Sakamoto, "Damage Modes and Protective System for an Automobile Cooler Cycle," SAE Transactions, vol. 96, pp. 1.107-1.116, 1988.

[2] T. M. Rossi and J. E. Braun, "Using Thermodynamic Impact for Detecting Refrigerant Leak in Vapor Compression Equipment," presented at Proceedings of the American Control Conference, Seattle, Washington, 1995.

[3] H. T. Grimmelius, J. K. Woud, and G. Bee "On-line failure diagnosis for compression refrigeration plants," Int. Journal of Refrigeration, vol. 18, pp. 31-41, 1995.

[4] P. S. Fasolo and D. E. Seborg, "Monitoring and Fault Detection for an HVAC Control System," HVAC&R Research, vol. 1, pp. 177-193, 1995.

[5] H. Inatsu, H. Matsuo, et. al., "Developmen of Refrigerant Monitoring System for Automotive Air-Conditioning System," SAE Transactions, vol. 101, pp. 29-37, 1992.

[6] W.-D. Gruhle and R. Isermann, "Modeling and Control of a Refrigerant Evaporator," ASME Journ of Dynamic Systems, Measurement and Control, vol. 107, pp. 235-240, 1985..

DEFINITIONS, ACRONYMS, ABBREVIATIONS

MAC	Mobile Air Conditioning
m_{ref}	refrigerant charge (mass) [lbm]
N_c	number of clutch cycles [-]
N_{comp}	compresor speed [rpm]
OEM	Original Equipment Manufacturer
Q_{cai}	condenser air volume flow rate [cfm]
Q_{eai}	evaporator air flow rate [cfm]
RH_{eai}	evaporator air inlet relative humidity
s	standard deviation
T_c	mean clutch cycling period [sec]
T_{cai}	condenser air inlet temperature [°F]
T_{eai}	evaporator air inlet temperature [°F]
T_{ero}	evaporator refrigerant outlet temperature [°F]
TTT	Time to Temperature Turning [sec]

970107

Fuzzy Controller for Thermal Comfort in a Car Cabin

Béatrice Gach and Michael Lang
Valeo Systèmes Thermiques

Jean-Christophe Riat
PSA-Peugeot Citroën

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

This paper presents two fuzzy logic based systems, developed by Valeo Thermal Systems and PSA Peugeot-Citroën, for controlling the thermal environment in a car cabin. This study aimed to simplify the control systems set up, while improving the cabin passengers comfort by taking into account the subjectivity of thermal sensation.

The first system regulates the internal cabin temperature from a temperature fixed by the user on the climate control panel. The second system proposes a new "intelligent" control panel in order to ensure a better thermal balance for the car passengers.

The two systems were installed and tuned on a Peugeot 605 vehicle, on which a standard automatic controller is already available. So, it was possible to compare the fuzzy and the classical series controllers on the same vehicle. The results show good regulation performances and demonstrate that the use of fuzzy logic reduces the development time.

1. Introduction

1.1. Context of the study

Introduced by L.A. Zadeh in 1965 [1], Fuzzy Logic has since been used successfully in many industrial applications. Its practical interest for system control is now demonstrated, especially when no model of the system to control is available.

The automotive industry, where embedded control systems are becoming commonplace, has been taking an interest in fuzzy logic for many years [3], [5], [6], [7]. A lot of automobile patents describe systems based on fuzzy logic and several standard cars are sold with fuzzy logic systems [5].

The goal of this study is to evaluate the use of fuzzy logic for controlling an automatic HVAC system. In fact, some characteristics of such a system appear to favor the use of this technology : non-linear behavior, lack of global model, subjective aspect of human thermal feeling [9].

1.2. Automatic HVAC systems

1.2.1. Principle of a car HVAC system

Every car is equipped with a heating and ventilating system in order to ensure comfortable climate conditions when the weather is cold. Today, these comfort conditions can be extended to warm seasons thanks to the addition of an AC loop. Figure 1 below shows the complete standard system.

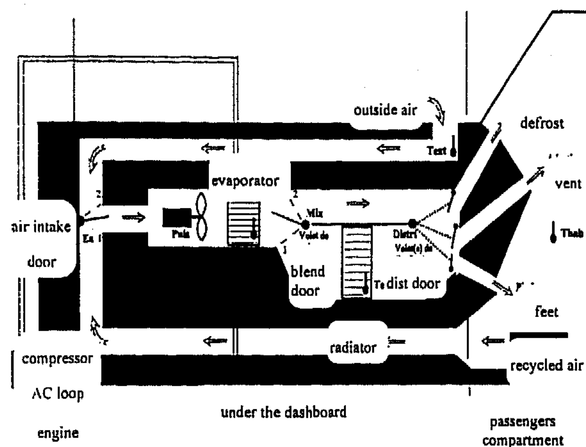


Figure 1 : Heating and air conditioning system

The inlet air (from inside or outside the car) is blown into the system by means of a blower and goes through the AC loop evaporator. A part of this air then crosses the radiator in which circulates the coolant from the motor cooling system. The other part is not heated up. The sharing of these two air masses, that are mixed downstream of the radiator, determines the final outlet air temperature.

1.2.2. Thermal comfort in a car cabin

The thermal comfort can be defined as the preservation of the body thermal balance, whatever the changes of the climatic environment. The thermal comfort conditions for buildings are well known today [2], [8]. But the results cannot be directly transposed to the car cabins that are subjected to specific constraints : forced convection and air speed, due to the closeness of air vents, great radiation influence due to the size and closeness of the windows, permanent contact with the seats, immobility, etc. Therefore, it is often difficult to tune the system to quickly obtain satisfying comfort conditions.

1.2.3. Automatic climate control in a car cabin

To more easily reach an optimal thermal comfort, an increasing number of vehicles is equipped with an electronic temperature and air flow control system. This system consists of a set of sensors and actuators , an electronic board where the controller is implemented and a control panel allowing the users to set their own desired climatic conditions.

1.3. Description of the experimental car

1.3.1. Standard series equipment

The HVAC hardware on which the fuzzy controllers were implemented is standard :

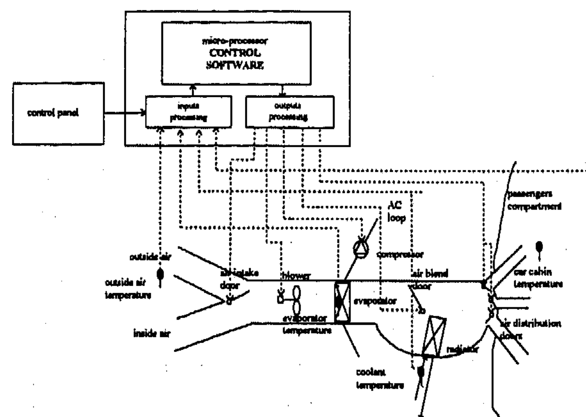


Figure 2 : Standard temperature control system

It consists of four temperature sensors, ie outside, coolant and evaporator temperature sensors to get information about the thermal state of the process and its environment, and a car cabin air temperature sensor, which is the variable to be controlled.

It also includes five actuators : **1** the blend door, to determine the temperature of the air blown in the cabin, **2** the blower voltage, that defines the airflow rate, **3** the distribution servomotor that fixes the way the air is blown in the cabin (towards the feet, the face or the windshield), **4** the air intake door, to determine the origin of the air entering the system and **4** the compressor clutch to turn the AC loop on or off.

1.3.2. New realizations

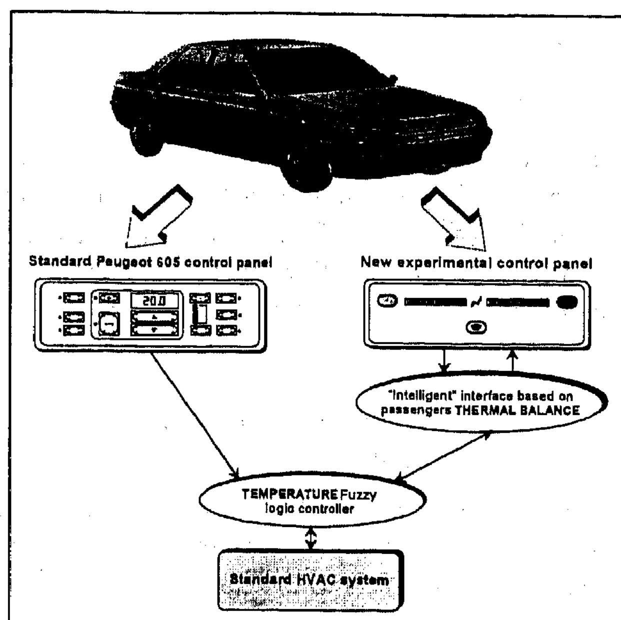


Figure 3 : systems realized

In figure 3, the system on the left consists of replacing the standard HVAC controller by a fuzzy logic one. The control panel and the HVAC system are the standard ones for the Peugeot 605.

This fuzzy controller is used by the system on the right, based on a thermal balance. New sensors are added (the same as for the previous system, plus hygrometry and car speed) and a new user interface is implemented in order to improve taking into account the subjective aspect of thermal feeling : the user doesn't indicate an absolute temperature but asks the system for cooler or warmer interior conditions.

1.3.3. Development system

The fuzzy control software is written in C. In order to test this software and to tune the controller, a specific hardware was utilized.

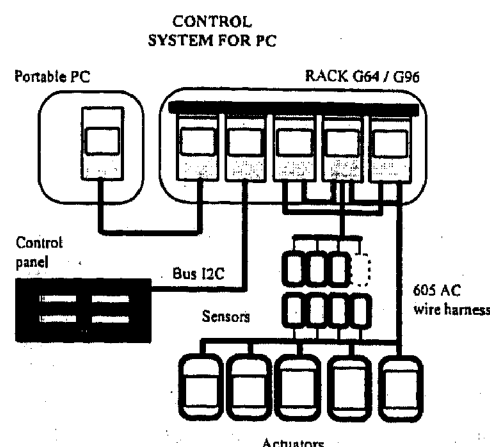


Figure 4 : on-board development system

This hardware (Figure 4) consists of a PC on board the vehicle and of a rack of Europe boards based on the G64/G96 bus, containing the interfaces. This system replaces the series control board and is connected to the vehicle wire harness.

For the thermal balance control application, the new control panel is connected to the rack with a I2C interface, that allows it to be physically separate from the control system.

The advantage of this solution is to be modular, which readily allows the addition of new sensors or actuators.

2. Fuzzy temperature regulation

2.1. Approach

Four steps were followed : 1 collection and structuration of Valeo expertise in the field of automotive climate control, **2** linguistic formulation of the desired behavior of the controller, **3** fuzzification of the system parameters, and **4** writing of the rules basis.

2.2. Fuzzy controller description

2.2.1. General structure

The fuzzy controller manages the blend, distribution and air intake doors, and the blower voltage. The compressor, which can

only be ON or OFF, is commanded by a classical binary logic.

2.2.2. *Fuzzification and defuzzification*

The input variables are fuzzified with piecewise linear membership functions. The output variables are crisp.

For each rules basis, the defuzzification method chosen is the mean of the activated outputs, weighted by their degree of activation.

Using this fuzzy process, we were able to reduce significantly the size of the final code compared to the standard control, which is a great advantage for our application.

2.2.3. *Example : blend door control*

The inputs are *epsit=tset-tint* (difference between the user's chosen set temperature and the actual car cabin air temperature) and the change of the car cabin air temperature *dtint=tint(t-1)-tint(t)*. The output is the change of the blend door position : *dv*.

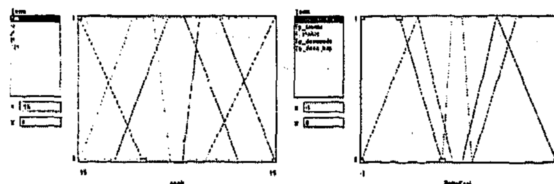


Figure 5 : Blend door control inputs fuzzification

dv	TTN	TN	N	Z	P	TP	TTP
value	-3.5	-2	-1	0	1	2	3.5

dtint \ *epsit*	TN	N	Z	P	TP
great_tp_increase	TTN	TTN	TN	Z	P
tp_increase	TTN	TN	Z	Z	TP
stable_tp	TTN	N	Z	P	TTP
tp_decrease	TN	Z	TP	TP	TTP
great_tp_decrease	N	Z	TTP	TTP	TTP

2.3. Results on the vehicle

2.3.1. *Controller tuning*

The controller was tuned from road and wind tunnel tests. This tuning lasted 4 months and took about ten days of wind tunnel tests, which is a significant reduction of time development compared with the classical controllers.

2.3.2. *Comparison with the classical series controller*

The tests were conducted in a wind tunnel in order to be able to submit the two systems (fuzzy and classical) to the same conditions.

The comparison presented here concerns the cold climates. Indeed, the heating mode is the most difficult to manage because the heat production is over-sized compared with the needs and is subject to several disturbances.

The series and fuzzy controllers were compared in two cases :

- engine ignition from a vehicle preconditioned state : the vehicle structure and the coolant liquid are at ambient temperature
- set temperature steps

The two following figures (6 and 7) show the evolution of the temperature near the driver's head and the set temperature, obtained first with the fuzzy and then with the series controller.

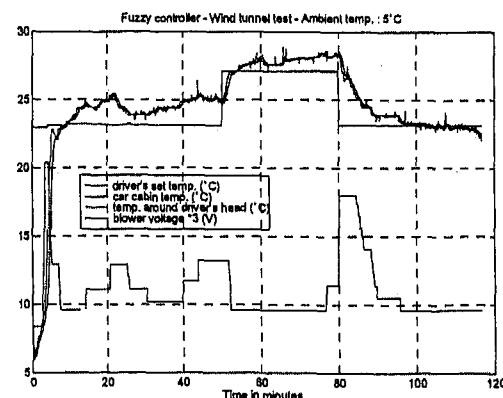


Figure 6 : fuzzy controller-wind tunnel at 5°C

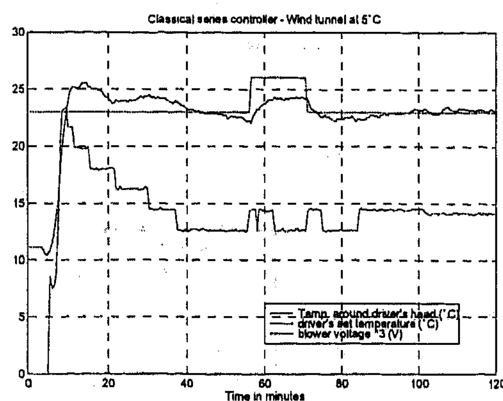


Figure 7 : classical series controller-wind tunnel at 5°C

Comparative tables :

- engine ignition from preconditioned state :

	Time for a 10°C increase near driver's head	Over shoot	Time before steady state	Max blower voltage	set temp. - cabin temp.	set temp.- driver's head temp.
Series	5 mn	1°C / 5 mn	12 mn	7.5 V	0 to 4 °C (not stab.)	1 to 2°C
Fuzzy	6 mn	3°C / 10 mn	20 mn	6.5 V	0.5 to 1°C	0°C

The fuzzy controller performances are equivalent to the classical ones, since the rise times are related to the blower voltage : the higher the blower voltage, the shorter the rise time.

- set temperature increase :

	Rise time	Overshoot	Time before steady state
Series	4 mn	no	4 mn
Fuzzy	1 mn	no	1 mn

- set temperature decrease :

	Fall time	Undershoot	Time before steady state
Series	6 mn	no	6 mn
Fuzzy	5 mn	no	5 mn

In the case of a set temperature change, the fuzzy controller has better performances.

These tables show better performances with the fuzzy controller.

2.4. Conclusion

The linguistic formulation gives the advantage to the fuzzy controller that allows :

- better understanding of the physical phenomenas involved, leading to a better controller maintainability,
- easier tuning, leading to a significant development time reduction,

for performances that are at least equivalent as those of the series controller.

The current studies concern the implementation of the fuzzy controller on the 605 series control board and its use on a Citroën ZX.

3. Thermal balance regulation

3.1. Principles of thermal balance

One characterization of thermal comfort is that the heat produced by the body is absorbed by the surroundings at the same rate, so that we do not accumulate heat or feel cold. The International Standard Organization defines PMV (Predictive Mean Vote) in ISO-7730 as an index of comfort. This approach based on the computation of a thermal balance is only validated for housing.

The thermal balance (noted L in Figure 8) is a mathematical function of six input variables (for more details see [8]) :

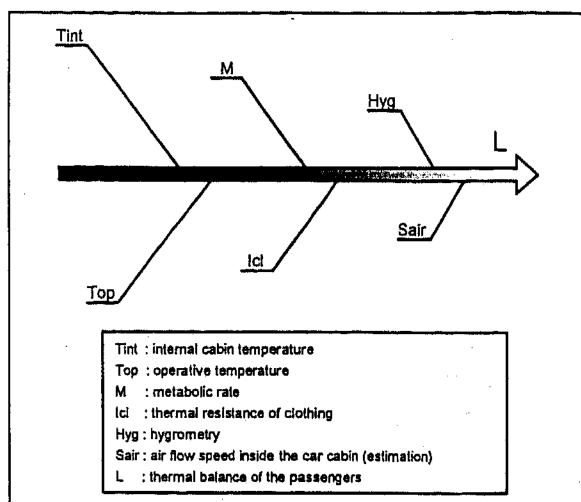


Figure 8 : parameters of thermal balance

The goal of this study is to compute this function value in the car and to control the air-conditioning so as to achieve maximum comfort level (L must be around zero). So it is possible to test this approach in an automotive context.

3.2. Architecture of global system

The principle is to use the temperature regulator presented in the second paragraph of this paper. This regulator is combined with a supervisor program which regularly calculates the value of the thermal balance.

This value is used to compute the best temperature for the thermal comfort in the car cabin. This temperature is the one which corresponds to a thermal balance around zero. This value is the new signal transmitted to the temperature regulator (Figure 10).

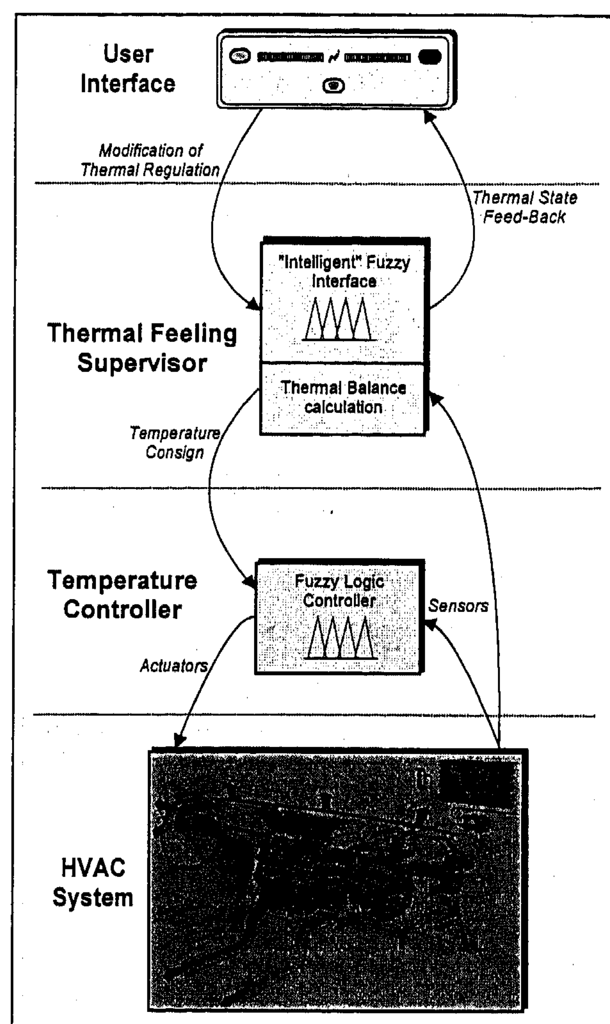


Figure 9 : architecture of global system

3.3. Working of the new control panel

The new control panel allows the user to ask for more heat or more cold. He indicates the intensity of his demand by lighting more or less L.E.D's. on the control panel (see Figure 10).

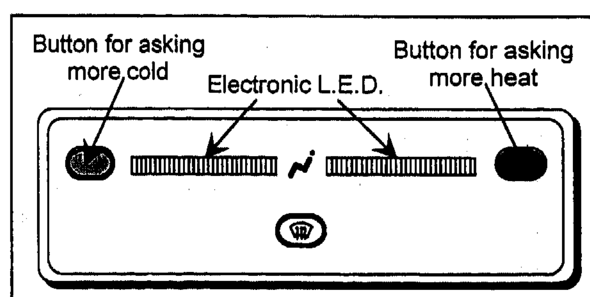


Figure 10 : new control panel

The idea is to use the action on the control panel to modify the value of the Icl parameter

(thermal resistance of clothing) in the calculation of the thermal balance. This parameter is initialized at a standard value when the system starts and is modified after each user action.

To calculate this modification (noted dIcl), we have used a fuzzy logic system with three inputs. The first one is the current value of the Icl parameter. The second one corresponds to the value of the thermal balance. The third parameter is the level of comfort wanted by the user which is quantified by the number of L.E.D's. he lit on the control panel.

For example, this figure shows the result of the calculation of dIcl depending on the desired comfort and the thermal balance :

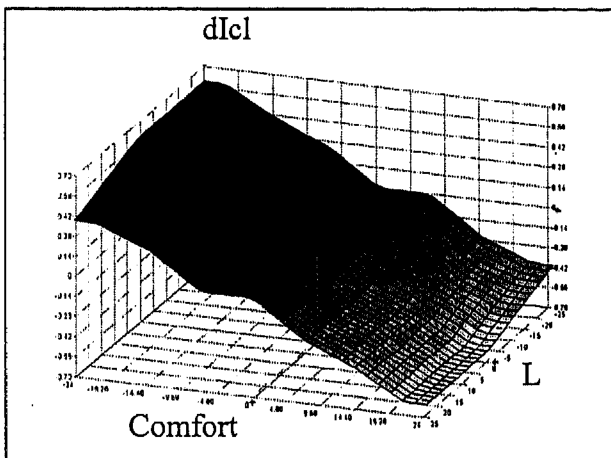


Figure 11 : control surface of dIcl

3.4. Experimental results

The complete system has been tested on ıoad by several different users. It is easy to understand though it is quite different from standard climate control panel.

To illustrate the principle of working of the new system, the graphics of the figure 12 shows the evolution of the parameters Icl and L during a test on road.

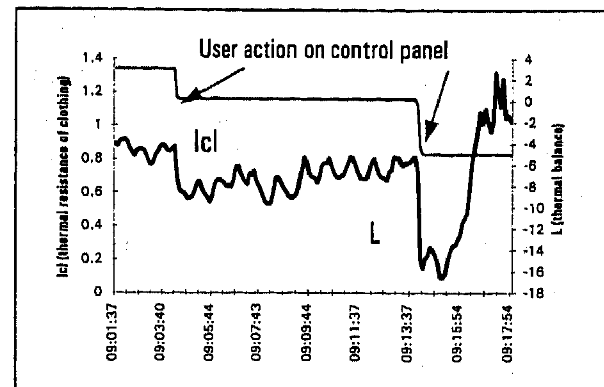


Figure 12 : evolution of Icl and L parameters

As shown in Figure 12, the user asks two times for more hot. At each time, the fuzzy interface of the control panel modifies the value of the Icl parameter. That's why the thermal balance calculated by the thermal feeling supervisor goes away from zero. A new temperature consign is calculated and transmitted to the controller. So the thermal balance returns progressively around zero.

4. Conclusions

This paper presented an actual case of use of fuzzy logic in an automotive context. It shows that for controlling a HVAC system, fuzzy logic allows to rapidly develop a first system with minimal effort. This system is then tested on roads and in a wind tunnel in order to tune it correctly.

A significant advantage of fuzzy logic is the case of tuning without writing any line code (modification or addition of rules). More over, the elaboration of the control strategy is simplified by its formulation with linguistic rules which simplify dialog with experts of the system to control.

5. Future prospects

These studies are continuing with several purposes :

- portage of the fuzzy controller on a Citroën ZX in order to evaluate its degree of portability

- implementation on a series control board;
- an ergonomic study in order to evaluate the performances of the thermal regulation not with technical criteria but with the users' point of view;
- development of a new fuzzy regulator directly based on thermal balance;

6. References

[1] L.A. Zadeh
"Fuzzy sets"
Information and control, Vol. 8, 1965, pp. 338-353

[2] P.O. Fanger
"Thermal confort analysis and application in environmental engineering"
McGraw-Hill, 1970

[3] L.I. Davis, T.F. Sieja, R.W. Matteson, G.A. Dage, R. Ames
"Fuzzy logic for vehicle climate control"
FUZZ-IEEE 94, pp.530-534

[4] F. Mingrino, G. Toscano Rivalta
"An automatic climate control based on the concept of equivalent temperature"
SAE International Conference 1995, SP-950022

[5] *"Development of a fuzzy controller for an automotive HVAC system"*
Mitsubishi Technical Information, 1992
(translated from Japonese)

[6] Nizar Al-Holou, Kashyap H. Shah
"Fuzzy logic based system to control climate in Automobile"
SAE International Conference, Mercy 1994

[7] S. Merlin, C. Melin, J.C. Riat
"An application of fuzzy logic to automatic transmission"
CESA'96 IMACS Multiconference, Lille France 1996, pp. 913-917

[8] ISO 7730
"Moderate thermal environments - Determination of the PMV and PPD indices and specification of the conditions for thermal comfort"

[9] L.C. Pairin
"Automatisation de la gestion du confort hygrothermique dans l'habitat et le tertiaire"
French CNAM report, 1993

970108

A Real-Time Computer System for the Control of Refrigerant Flow

Andy Bartlett, David Standaert, and Eric Ratts
University of Michigan-Dearborn

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

This paper presents a real-time computer system for the control of refrigerant flow in an automotive air conditioning system. This is an experimental system used to investigate the potential advantages of electronic flow control over conventional flow control (using an orifice tube or thermal expansion valve). Two features of this system are presented. First, the system organization is described. Second, the control and interface software are presented. The emphasis is on the software.

The system is organized as a closed loop control system. The inputs to the controller are measurements of the refrigerant system. In particular, thermocouples are used to measure the refrigerant temperature before and after the evaporator. The analog thermocouple signals are converted to digital form by an off-the-shelf, portable, data acquisition system (DAQ). Via a parallel port link, these digital measurements are transfered to a laptop computer. The laptop software processes the measurements and determines a desired refrigerant valve opening. Via the parallel port, the desired valve opening is passed back to the DAQ. The DAQ converts the desired setting from a digital value to a low power analog signal. Using this signal as an input, an analog power amplifier drives the valve to the desired opening. The new valve opening increases or decreases flow of refrigerant into the evaporator. This change in flow effects the temperatures measured by the thermocouples. This closes the control loop.

The software includes a real-time digital controller and a user interface. The software was custom designed using Visual Basic and runs under Windows. The DOS/Windows environment is generally problematic for real-time control, but has proven to be suitable for refrigerant flow control. Successful operation under Windows is made possible by resources provided by the DAQ, by careful limitation of Windows activities, and by the relatively slow dynamics of the air conditioner. In addition to the real-time controller, the software simultaneously provides a graphical user interface (GUI). The GUI allows the user to select and change controllers on the fly in real-time. The GUI also provides the user with a "virtual" strip chart. The strip chart aides the user in tuning the controller during operation. This software has been used on the road in a test vehicle and has proven to be very effective in the development of refrigerant flow control methods.

1.0 Introduction

This paper presents a real-time computer system for the control of refrigerant flow in an automotive air conditioning system. This is an experimental system used to investigate the potential advantages of electronic flow control over conventional flow control. Two features of this system are presented. First, the system organization is described. Second, the control and interface software are presented. Before presenting these features, previous work in the area will be reviewed.

2.0 Prior Work

This subsection will describe the traditional methods of controlling refrigerant flow and some electronic control methods that have been tried. The traditional ones will be described first. Advantages, limitations, and current status will be noted as appropriate.

One of the two most common traditional methods of refrigerant control in automobiles is the orifice tube (OT). As shown in Figure 1, the orifice is a fixed size restriction. By focusing just on the evaporator subsystem, it would appear that the orifice tube had no way to regulate the refrigerant flow. By looking at the system as a whole, it can seen that an orifice tube equipped system does make some adjustment of refrigerant flow to meet operating conditions [1]. However, it is generally agreed that the orifice tube system does not adjust flow as well as systems whose expansion orifice size is actively adjusted. The advantage of an orifice tube systems is in lower cost and increased reliability.

The second of the two most common traditional methods of refrigerant control in automobiles is the thermal expansion valve (TXV). A TXV system is shown in Figure 2. The opening of the TXV is actively adjusted based on a balance of forces. One of the forces is due to a sensing bulb placed at the evaporator output. As the outlet gets hotter, the gas in the sensing bulb gets hotter. Since the bulb and the capillary tube are essentially a closed volume system, the pressure in the bulb and tube increases. This pressure creates a force that increases the valve opening thereby increasing the flow. The increased flow reduces the superheat which reduces the bulb pressure. Eventually a balance of pressure and flow is reached. Some TXV systems use more complicated mechanisms to achieve a desirable balance [1]. The advantage of TXV systems is that they actively regulate superheat to a moderately positive value. Their disadvantage is they are less reliable than orifice tubes due to their moving parts and their relatively fragile capillary tubes.

Electronic control retrofits have been designed for some TXV systems [5]. Figure 3 shows an example. In this system, the temperature of the evaporator inlet and outlet are measured by a computer. The difference of these two temperatures gives a good approximation of superheat if the pressure drop across the evaporator is small. Based on these measurement, the computer adds more or less heat to the sensing bulb using a small electric heating pad. The goal of the retrofit is to have the system achieve a smaller superheat than the moderate one TXV systems are usually designed for. A standard TXV uses the relatively high superheat to avoid sending liquid to the compressor. The electronics can use a faster, more sophisticated design than is possible to mechanically fabricate. This is why the retrofit can safely decrease the superheat in both the transient and steady state periods.

The retrofit system is primarily for large stationary

refrigeration systems. One example is a supermarket display case. If an automobile manufacturer is going to use electronic control of refrigerant control, then it does not make sense for them to use a retrofit design. It is natural to expect that a direct design would have cost and performance advantages. In addition, the need to significantly modify the control algorithm is likely because an automotive system has more frequent and dramatic transients.

Two patents for direct electronic control of refrigerant flow were recently awarded [3] and [2]. Figures 4 and 5 show that both systems use an electronically controlled expansion valve. From a system architecture point of view, the main difference appears to be the method of measuring superheat. In Figure 4, the same approach as the retrofit system is used. Evaporator inlet and outlet temperature are measured and used to approximate superheat. In Figure 5, temperature and pressure at the evaporator outlet are measured. Using the properties of the refrigerant, these two measurements can be used to compute superheat. Based on the superheat, the computers will command new settings to the electronic expansion valves (EXV). The algorithm used to determine the valve adjustment will strongly affect the performance of the system.

It is interesting to note an SAE paper [4] related to the subject of the two patents. This paper was presented one year before the two patents were filed. This paper [4] shows what appears to be the same architecture as Figure 5. The paper [4] also shows experimental data obtained from this system.

Electronic control of refrigerant flow in automotive A/C systems is not a new idea as attested by the 1987 paper and the 1989 patents. Furthermore, electronic flow control is a feature available for large stationary A/C systems. Despite these facts, conventional flow control methods, OT or TXV, totally dominate the automotive market. The goal of the work described in this paper was to develop a prototype vehicle to investigate any potential advantages of electronic flow control.

3.0 System Organization

This section briefly describes the organization of a prototype system. This description covers the major hardware components, the instrumentation, and the electronic/computer hardware as shown in Figure 6. The computer software is not covered until the following section.

The prototype vehicle was a 1994 Lincoln Mark VIII. The prototype used production components for the condenser, evaporator, compressor, OT, accumulator, and electronic automatic temperature controller (EATC). Note that the EATC regulates cabin air temperature by controlling the compressor clutch and the air handling system, but it does not meter the flow of refrigerant.

The test vehicle was modified so that the air conditioning system could operate in conventional or electronic flow control mode. The hand operated Y-valve shown in Figure 6 allows either the OT or the EXV to meter the flow of refrigerant. Aside from some small changes in plumbing, selection of the OT system essentially gives the production Ford system.

The other setting of the Y-valve causes the refrigerant to be controlled by an EXV. Adequate electronic sensors were available so that Figure 4 or 5 could be implemented. Temperature was sensed by sheathed T-type thermocouples that were instrumented into the refrigerant flow stream. A specialized thermocouple card, IOTech DBK19, provided the interface to the small signal levels of these sensors. The pressure sensors came with built in electronics and could be interfaced directly to any general purpose Analog to Digital (A/D) converter. The A/D converter used in this project is part of a portable data acquisition board, IOTech DAQBook 100. This data acquisition board (DAQ) interfaces to any DOS/Windows compatible computer via a standard parallel port. Laptops were especially convenient for this purpose. Via the port, the sensor measurements are eventually made available to the computer. Based on these measurements, the computer uses some chosen algorithm to calculate a desired valve opening. This is desired opening is passed to the DAQ again via the parallel port. The DAQ system has a Digital to Analog (D/A) converter which it uses to put the desired opening in analog form. This analog signal is low power and can't drive the EXV. For this reason, the low power signal is passed through an analog power amplifier. The output of the power amplifier drives the valve. This completes the equipment required for the electronic flow control prototype.

In addition to electronic flow control, electronic monitoring of the entire system was also desired. For this reason numerous thermocouples and pressure sensors were directly connected to the refrigerant system. Additional thermocouples and humidity sensors were also used to measure the ambient and cabin conditions. These measurements could have been incorporated in the DAQ plus Laptop used for flow control. However, the extra measurements would have severely restricted the control capabilities given the many speed limitations of the DAQ plus laptop system. For this reason, a second laptop plus DAQ was used. One laptop plus DAQ was dedicated to control; the other was dedicated to data recording.

4.0 Real-time Software

The two laptop plus DAQ systems run two different sets of software. The data monitoring system uses a Windows based program that was supplied with the DAQ hardware. This software had adequate data collection capabilities, but it did not have real time control capabilities. Custom software was developed for the laptop plus DAQ system that performed the real time control. The overall feedback system is shown in Figure 7 where the real time software represents the shadowed blocks. The system can also run in a "test mode" as shown in Figure 8. The real time control software will be described in this section.

4.1 Oversampling and Conditioning

The real time software was written to use the scheme shown in Figure 4. This means that two thermocouple measurements are processed by the control system. Initially, two types of processing are required. First, the nonlinearities inherent in the thermocouples and the scaling due to the amplifiers must be accounted for. This shall be referred to as conditioning. Second, the noise in the measurements must be dealt with. Subroutines for conditioning the thermocouple signals were supplied by the DAQ manufacturer, and these were employed. For dealing with the noise, a custom strategy was developed.

The noise arises in many locations including the sensor, the wire harness, the interface electronics, and the A/D converter. The measurement noise could be attenuated in the digital controller or by using oversampling. The prototype system used a combination of both approaches.

A linear digital controller is identical in structure to a digital filter. If the control loop were running at a very high sampling rate, then the digital controller could include many fancy filtering effects. However, on the prototype vehicle, the controller will normally be running at less than a 10 Hertz sampling frequency. This relatively slow rate puts some constraints on high frequency filtering. Signal aliasing may also be a problem at this slow rate. Unfortunately, running at a faster rate is likely to overburden the computation and communication abilities of the laptop plus DAQ system. At a faster rate, the software may not meet the real time requirements.

To achieve the desired high frequency filtering, oversampling was used. Like most of today's A/D boards, the DAQ used in the prototype contains just a single A/D converter. Via high speed multiplex switching, many different signals can be linked one at a time to this converter. An allowable alternative is to link the same signal or sequence of signals to the converter several times in a row. This arrangement allows oversampling.

For the DAQ, a "new" signal can be linked to the A/D converter every 10 microseconds (100 kHz). If the control loop is executing every 100 milliseconds (10 Hz), then theoretically, the A/D could measure 10,000 signals in each cycle of the control loop. Due to communication and memory limits, the actual number of measurements must be smaller. At compile time, the custom software is set up to oversample each signal a certain number of times. Ten times oversampling was found to be effective at greatly reducing the high frequency noise. Measuring

each of the two thermocouples 10 times requires a total of 20 measurements by the A/D during each control loop cycle. Each set of ten samples is simply averaged to produce a filtered measurement that is used by the control loop every 100 milliseconds.

The oversampling does place a burden on the real time software. If just one sample per cycle were taken, then the control loop could be sped up a little bit. However, the speed penalty of moderate oversampling is relatively small and the noise reduction benefits are significant. Oversampling was found to be an efficient use of the prototype resources.

4.2 Superheat Calculation

The superheat calculation carried out by the software was extremely simple. It was assumed that two phase refrigerant was entering the evaporator and that there was no pressure drop across the evaporator. Under these assumptions, any superheat at the outlet would be equal to the temperature rise across the evaporator. This calculation of superheat simply involved subtracting the inlet temperature measurement TCinlet from the outlet temperature measurement TCoutlet.

Experimentation showed that there actually was a pressure drop across the evaporator of approximately 14 kPa. This pressure drop results in an error in the superheat calculation. The exact amount of error would depend on both the pressure drop and on the absolute pressure at the evaporator inlet. The evaporator inlet pressure had a fair amount of variation, but the pressure drop was always quite close to 14kPa. For the normal range of inlet pressures, the effect of a 14kPa pressure drop on R134a superheat was investigated. It was found that over the entire range the 14kPa pressure drop corresponded to a 2 K increase in the superheat. Therefore the calculation of superheat, TCoutlet - TCinlet, could be improved by adding 2 degrees Kelvin to it. From a control point of view, it would be equivalent to lowering the desired superheat by 2 K. This latter approach was used.

4.3 Desired Superheat

The desired superheat is the target value that the feedback control loop will try to achieve. When the computed superheat is higher than the desired value, the controller must eventually permit more refrigerant to flow. Conversely, the flow is decreased when the computed superheat is below the desired value. The desired superheat is simply a fixed value. Using the graphical user interface (GUI), the desired value can be readily changed during operation. When setting the desired value, the user should keep in mind that the computed superheat is approximately 2 K too low. So if the user wants a true superheat of plus one degree Kelvin, then the desired superheat should be set to minus one degree Kelvin. The lower left corner of Figures 9 and 10 shows the location in the GUI where the superheat goal is entered.

4.4 Digital Controller

The digital controller processes the superheat error to come up with a desired voltage for the EXV. The digital controller is linear for the most part so it is equivalent in structure to a digital filter. The controller does include a simple nonlinearity for anti-windup purposes. The anti-windup could restrict the controller output to the 0 to 5 volt range of the D/A. A smaller range corresponding to the response limits of the EXV were used in the prototype.

The linear portion of the controller can be described by a Z-domain transfer function (ZTF). Direct selection of a ZTF by the user faces some difficulties. If a good linear time invariant model of the plant is available, then there are many nice canned programs for automatically selecting a digital controller's ZTF. However, a good model, especially a linear one, of the refrigerant system was not available. Therefore, it was expected that the user would use heuristic rules to select and tune control features. For a control engineer, selecting features often involves selecting pole and zero locations. Many control engineers have an excellent feel for the effects of poles and zeros in the S-plane of continuous time system. Control engineers typically have a far poorer understanding of pole and zero locations in the Z-plane for discrete-time systems. One reason is that pole and zero locations in the Z-plane depend on the chosen sampling rate of the digital controller. For these reasons, the ZTF of the controller was indirectly specified.

The GUI allows the user to select a continuous time S-domain transfer function (STF). Figures 9 and 10 show the screen by which the user selects poles and zeros in the S-plane. The user may select up to three real poles, three real zeros, two pair of complex conjugate poles, and two pairs of complex conjugate zeros. The default steady state, i.e. DC gain, of the system is plus one. The user may override this choice by selecting an overall, a Proportional , an Integral, or a Derivative gain. Depending on which of the PID gains are selected, the STF can have an additional pole at the origin and/or up to two additional zeros. At any time, the user can change in real-time the number and location of poles, zeros, and gains by pressing the update button.

The prototype system is based on a digital computer so the STF selected by the user must be converted to a ZTF. The first issue in this conversion is the selection of a sampling rate. Near the top center of Figures 9 and 10 is where the user can select the desired sampling frequency. The selected frequency must not be so fast as to over burden the system. For refrigerant control with a full set of graphical displays, the system was able to readily handle rates less than 10 Hz.

After the selection of sampling rate, the method of mapping from the Z-plane to the S-plane must be selected. The bilinear transformation alias Tustin's approximation was employed in the software and was found to be effective. Given a STF H(s) and sampling rate h, the corresponding ZTF G(z) is given by simple substitution of s.

$$G(z) = H\left(\frac{2}{h}\ \frac{z-1}{z+1}\right)$$

By selecting the appropriate check box in Figures 11 and 12, the GUI allows the user to view either the ZTF or the STF.

Once the ZTF has been formed, implementing the digital controller is straight forward. Conversion of the ZTF to a difference equation is trivial. Because of the GUI allows the user a limited quantity of poles and zeros, the controller can be ninth order at worst. This means that implementation of the difference equation involves 19 multiplications and 18 additions at worst. The implementation must also store nine or fewer recent inputs and nine or fewer recent outputs. These computations were a light burden for a 486DX 25MHz laptop computer.

Before closing this subsection, some comments on the software will be made. Visual Basic is not particularly computationally oriented, but it did not hinder the implementation of the small digital controller. The formation of the STF from the selected poles, zeros, and gains had the potential to be some what cumbersome. In addition, the bilinear transformation appears to be a moderately involved symbolic processing problem. There were three keys to making the software manageable. First, a data structure was created that could handle any polynomial up to some specified maximum order. Second and third, subroutines for adding and multiplying any pair of polynomials were written. Note that the STF and ZTF are just the just the ratios of polynomials. Adding a new pole or zero to the denominator or numerator of the STF, respectively, simply involves polynomial multiplication. The bilinear transformation only involves multiplying and adding polynomials of the form q, 2z+2, and hz-h as seen in the example below.

$$H(s) = \frac{2\ s + 3}{4\ s^2 + 5\ s + 7}$$

$$G(z) = \frac{2\left(\frac{2z-2}{hz+h}\right) + 3}{4\left(\frac{2z-2}{hz+h}\right)^2 + 5\left(\frac{2z-2}{hz+h}\right) + 7} =$$

$$\frac{2\ (2z-2)\ (hz+h) + 3\ (hz+h)\ (hz+h)}{4\ (2z-2)\ (2z-2) + 5\ (2z-2)\ (hz+h) + 7\ (hz+h)\ (hz+h)}$$

The zeroth order polynomials q are the coefficients of STF. The bilinear transformation has a regular pattern and can be fully implemented with a set of nested For loops.

4.5 Graphical User Interface

The graphical user interface has two important screens. The change control GUI is illustrated in Figures 9 and 10. The use and operation of this screen was described in the previous subsection. The main GUI is illustrated in Figures 11 and 12. The key features of the main GUI will be described in this section.

The main GUI is divided into four main areas. The bottom portion of the GUI contains several check boxes, buttons, etc. that control the operation of the system and the GUI. This controls will be referenced as the other three sections are described.

The top portion of the GUI shows in tabular form the ten most recent values of the key signals. When the "Test Mode" check box is selected as in Figure 11, only the input and output of the digital controller. When the "Super Heat" mode is checked as in Figure 12, only the thermocouple and superheat values are also displayed. To reduce visual flicker and to minimize computational burden, the newest data is added to the table in a circular buffer fashion. The star at the top of the table marks the newest data, and data immediately to the right is the oldest data. By checking the Disable Output Grid box, no more new data will be inserted in the table. The real time system can operate at a higher sample rate when the various displays are individually or collectively disabled.

The upper-middle portion of the GUI shows the transfer function of the controller. By checking Continuous Time or Discrete Time, the STF or ZTF, respectively, will be displayed. For example, Figure 12 shows the following STF.

$$H(s) = \frac{-1\ s - 1}{33.3\ s^2 + 1\ s + 0}$$

The upper left corner of this region also gives the sampling rate.

The lower-middle portion of the GUI provides a virtual strip chart. By checking the appropriate boxes any or all of the signals associated with a given mode can be plotted. Figure 11 shows a plot of the Test Mode inputs and outputs. The input is a square wave from an external function generator and the output corresponds to the under damped second order system selected for the controller. Figure 12 shows all five of the signals from Super Heat Mode. The graph is not batch updated; it is a true stripchart and slowly scrolls from left to right. The signal plots, the axes, and the enable check boxes are all color coordinated to allow easy identification. To aid the user, the signals were heavily filtered before plotting so as to remove noise and jitter. This explains why the square wave in Figure 11 shows very slow transitions.

5.0 Performance

The real time system described in this paper was evaluated in Test Mode and in actual operation on the prototype vehicle. The detailed results of the refrigeration control are beyond the scope of the paper. The comments here will be restricted to the systems ability to meet the constraints of real time operation. In super heat mode, the system was fully operational at 10 Hz. This rate along with full use of the displays and of controller updates was satisfactory for the refrigeration control experiments.

In test mode, the system was capable of operating better than ten times faster. The omission of the thermocouple signal conditioning reduced the processing load. In addition, if the GUI displays of data were disabled, then much higher rates could be sustained. Figure 13 shows an oscilloscope trace where the real time system is producing an underdamped second order response at 150 Hz. When a Pentium based desktop machine was substituted for the Laptop 486, sampling rates of 500Hz were obtained.

Because the system was running under the DOS/Windows environment, it was not possible to obtain guaranteed real time performance. In the vast majority of cases, the controller output was properly updated each sample time. However, the system did fall behind in some situations. Falling behind was noticed in higher speed test mode operation and could be captured on an oscilloscope. In Figure 13, the digital system output is a staircase plot and each step has a consistent width very close to 1/(150 Hz). If Windows operating system, attempted a time consuming task such as updating an interactive screen, then the staircase plot could show one very long step followed by several short steps. In this case, samples were taken by the DAQ hardware at the appropriate times but the computer was to busy some Windows task to process the samples. Eventually Windows would return control to the real time system, and this system would try to "crunch through" the backlog as quickly as possible.

6.0 Conclusions

This paper has presented a real time computer system for the control of refrigerant flow in a prototype automobile. The system carries out digital control and provides a GUI that allows the control to be tuned on line. The GUI also provides graphical and tabular displays that assist in the tuning process. Testing on the prototype vehicle has shown that the system has the real time control capabilities required for the refrigeration control project. On road testing has shown that the system can simultaneously provide the graphical user interface and maintain a 10 Hz control loop.

7.0 Acknowledgements

This project was made possible through the support of several organizations. The authors wish to thank the Center of Engineering Education and Practice (CEEP) for three years of generous support. The authors also wish to thank the Ford University Research Program for two years of extremely helpful support. Ford Motor Company's Components Division (formerly Climate Control Division) also supported the project by providing the first and third author with faculty internship for two summers each. Ford's Alpha Division gave crucial support by providing the Mark VIII test vehicle. The University of Michigan-Dearborn Campus Grant Program gave seed support. All of this support was highly appreciated by the authors and by the many student research assistants who gained valuable experience because of it.

This project had many interdependent parts each of which was crucial to the whole. The authors would like to acknowledge the many people who did not contribute directly to the real time software but did play valuable roles in the project. Richard Terry was the key player in the mechanical aspects of the instrumentation. Tim Philippart designed and constructed the power amplifier. Dennis Kubica, Josh Isser, and Joe Jaegger assisted in the electrical aspects of instrumentation. Paul Rogers played a lead role in much of the data collection and analysis. The UMD Electrical and Mechanical Engineering Shop technicians: Jesse Cross, Martin Stenzel, Don Haidys, and Mike Solstad, also aided in the completion of the project. In addition, engineers at Ford Motor Company contributed their experience and expertise in selecting and evaluating the project. Key contributors were Chris Rockwell, Gary Dage, Bob Matteson, George Wiklund, and David Zietlow. This project lasted three long and hectic years, and there are undoubtedly other contributors that have been overlooked. The authors regret that the failings of their memories prevent them from individually acknowledging and thanking each contributor.

8.0 References

1. Cuffe, K. "Air Conditioning and Heating Systems in Trucks," SAE Paper 780001, 1978.

2. Barthel, R.C., Malone, P.J., Orth, C.D., and Jarosch, G.W., "Controlling Refrigeration," United States Patent No. 4,848,100, July 18, 1989.

3. Torrence, R.J. "Controlling Superheat in a Refrigeration System," United States Patent No. 4,835,976, June 6, 1989.

4. Mitsui, M., "Improvement of Refrigerant Flow Control Method in Automotive Air Conditioners," SAE Paper 870029, 1987.

5. Marsala, J., "An Electronic Control for Thermostatic Expansion Valves," Proc. of Building Retrofit Innovative Concepts Fair - An Experimental Technology Exchange, Washington, DC, Oct. 24, 1989.

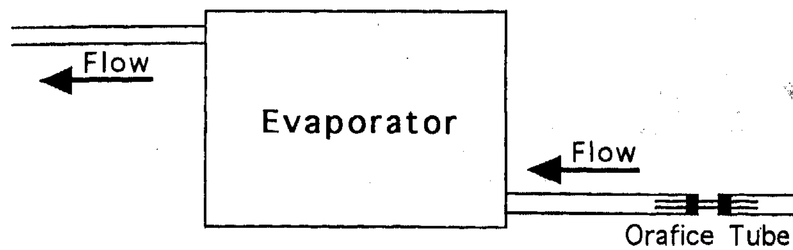


Figure 1: Refrigerant expansion using a orifice tube.

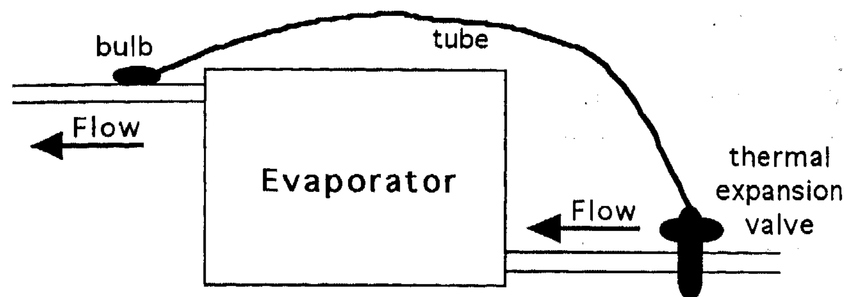


Figure 2: A basic thermal expansion valve system.

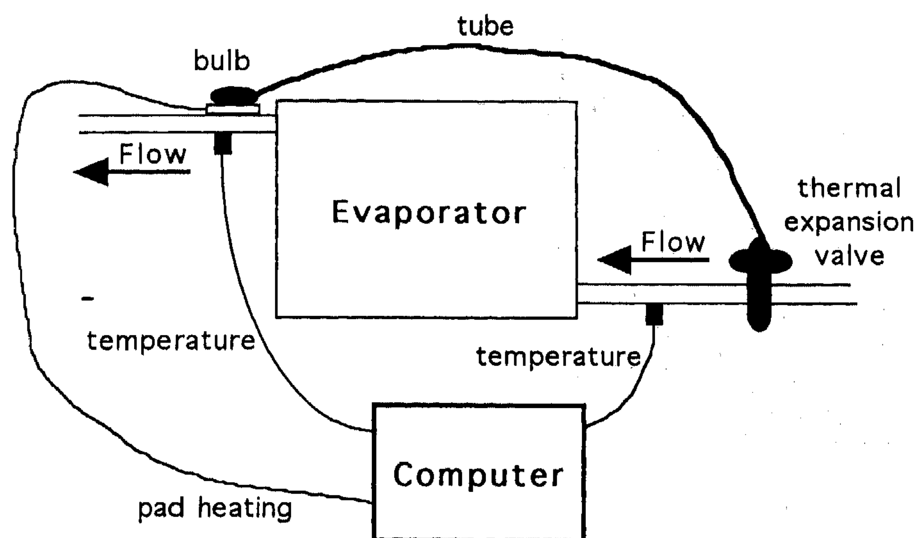


Figure 3: A thermal expansion valve system with electronic control retrofit.

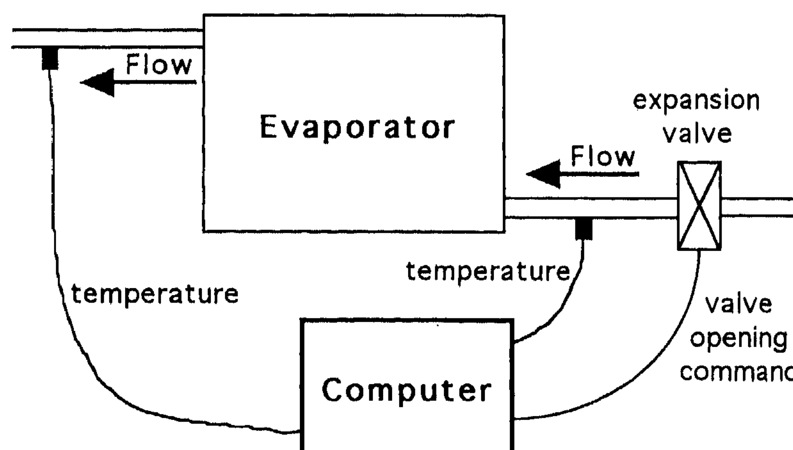


Figure 4: Electronic expansion valve with two temperature measurements.

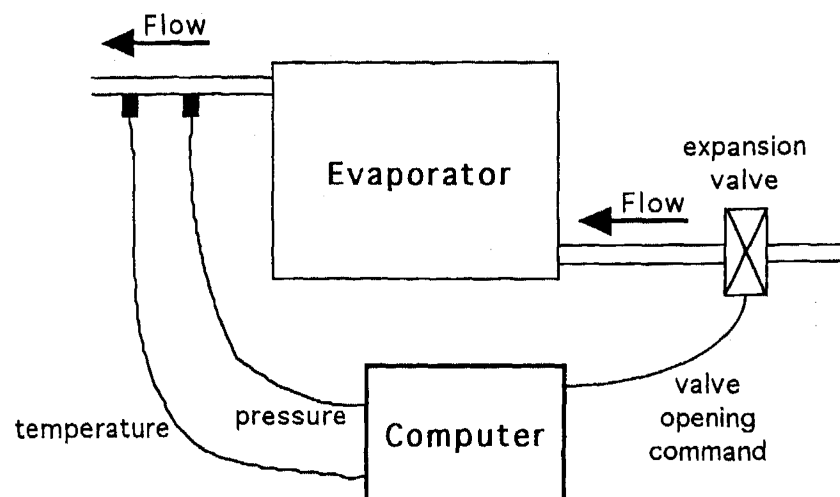


Figure 5: Electronic expansion valve with temperature and pressure measurements.

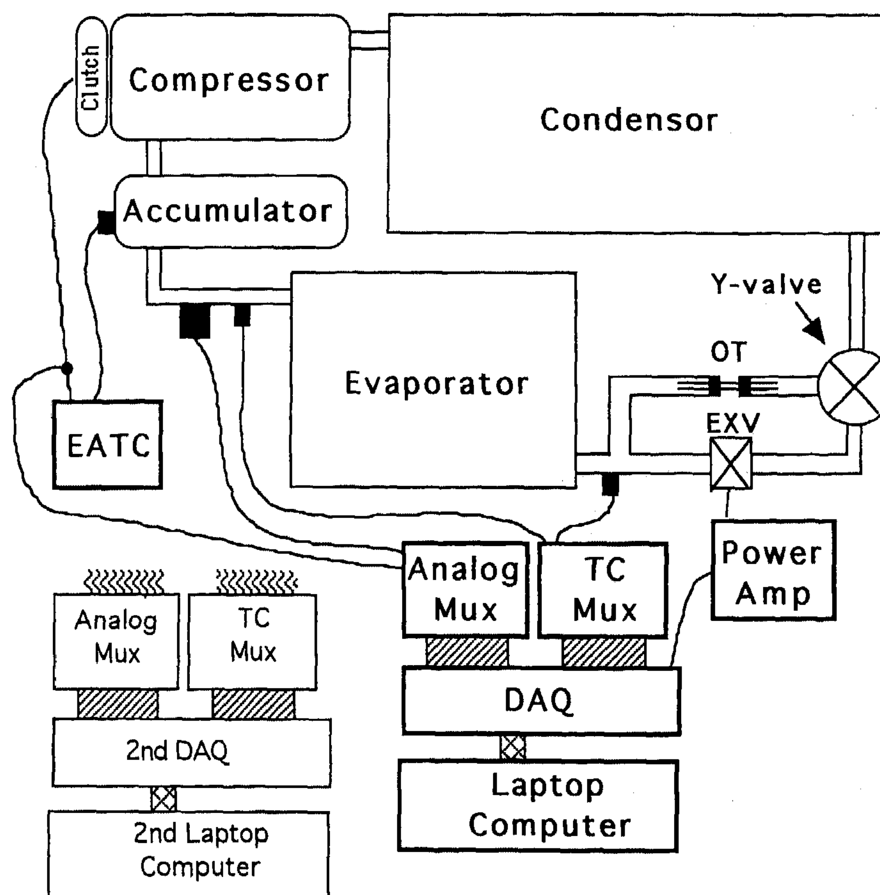


Figure 6: Electronic expansion valve with temperature and pressure measurements.

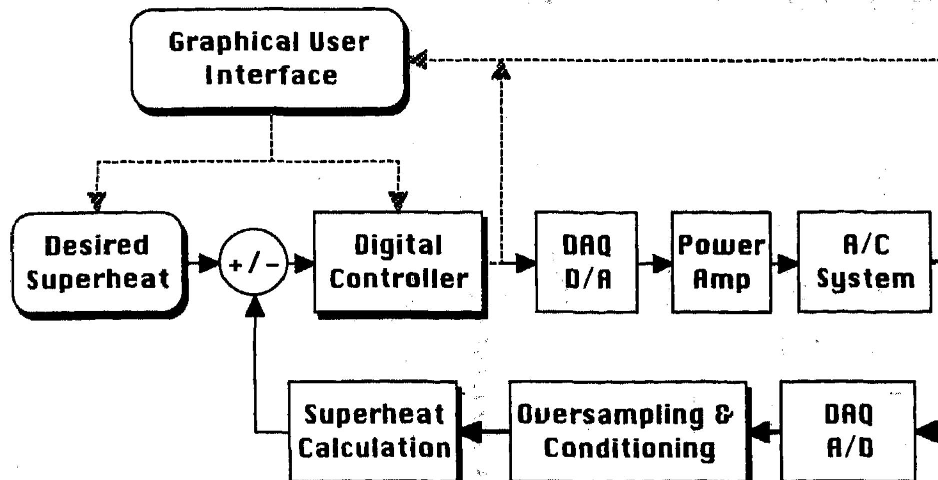


Figure 7: Diagram shows closed loop system for the control of refrigerant flow. Shadowed boxes represent real-time software running on PC.

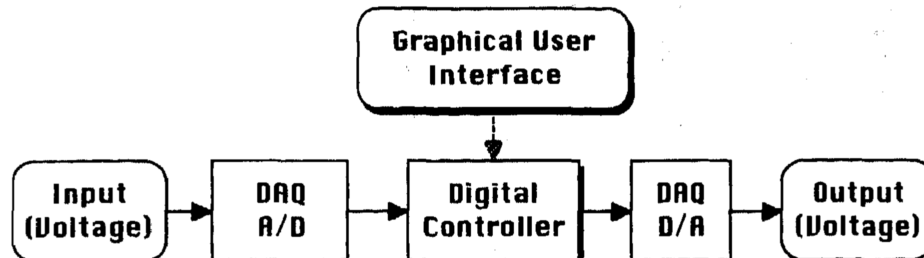


Figure 8: Diagram shows test mode for real time system. This mode is useful for testing and debugging the system. This mode could also be used in a general control loop provided the inputs and outputs are electrical signals.

VELab Change Refrig Flow Controller

Update | Hide | Frequency 150 | **Overall Gain** ☐ Ko 1

PID: ☐ Kp 1 ☐ Ki 1 ☐ Kd 1

Real Poles: ☐ P1 1 ☐ P2 1 ☐ P3 1

Real Zeros: ☐ Z1 1 ☐ Z2 1 ☐ Z3 1

Complex Poles: ☒ ZetaP1 0.1 ☒ WnP1 100
☐ ZetaP2 1 ☐ WnP2 1

Complex Zeros: ☐ ZetaZ1 1 ☐ WnZ1 1
☐ ZetaZ2 1 ☐ WnZ2 1

Desired Superheat (C) 2 Output Bias (V) 0

Figure 9

VELab Change Refrig Flow Controller

Update Hide

Frequency 10

Overall Gain ☒ Ko -1

PID: ☒ Kp 1 ☒ Ki 1 ☐ Kd 1

Real Poles: ☒ P1 -0.03 ☐ P2 1 ☐ P3 1

Real Zeros: ☐ Z1 1 ☐ Z2 1 ☐ Z3 1

Complex Poles: ☐ ZetaP1 1 ☐ WnP1 1
☐ ZetaP2 1 ☐ WnP2 1

Complex Zeros: ☐ ZetaZ1 1 ☐ WnZ1 1
☐ ZetaZ2 1 ☐ WnZ2 1

Desired Superheat (C) 2 Output Bias (V) 0

Figure 10

Vehicle Electronics Lab Refrigerant Flow Controller

New				*						
TCinlet										
TCoutlet										
SuperHea										
InCont	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.75	1.76	1.76
OutCont	1.72	1.74	1.77	1.78	1.82	1.80	1.77	1.74	1.72	1.71

10 Hz	S^9	S^8	S^7	S^6	S^5	S^4	S^3	S^2	S^1	S^0
Num										1.00E+00
Den								1.00E+00	2.00E-01	1.00E+00

5	5	16	32	32
4	4	12	24	24
3	3	8	16	16
2	2	4	8	8
1	1	0	0	0
0	0	-4	-8	-8

Time between vertical tics 5.0 Seconds

Enable Graph of:
- Evap Inlet Temp
- Evap Outlet Temp
- Evap Super Heat
- [x] Test Input (V)
- [x] Output (V)

Show Transfer Function
- (•) Continuous Time
- ( ) Discrete Time

Run Controllers
- ( ) Super Heat
- (•) Test Mode

Disable
- Output Grid
- Graph
- TF Grid

Exit

Start Acquisition

Change Controller

Electrical and Computer Engineering,
Mechanical Engineering,
VELab, FURP, & CEEP
University of Michigan-Dearborn

Figure 11

Vehicle Electronics Lab Refrigerant Flow Controller

New							*		
TCinlet	31.50	31.00	30.70	30.80	30.40	30.30	30.50	30.70	32.00	30.80
TCoutlet	28.70	28.90	28.80	29.30	29.10	29.00	28.80	28.60	28.80	28.70
SuperHea	-2.80	-2.10	-1.90	-1.50	-1.30	-1.30	-1.70	-2.10	-3.20	-2.10
InCont	4.80	4.10	3.90	3.50	3.30	3.30	3.70	4.10	5.20	4.10
OutCont	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

10 Hz	S^9	S^8	S^7	S^6	S^5	S^4	S^3	S^2	S^1	S^0
Num									-1.00E+00	-1.00E+00
Den								3.33E+01	1.00E+00	0.00E+00

5	16	16	32	32
4	12	12	24	24
3	8	8	16	16
2	4	4	8	8
1	0	0	0	0
0	-4	-4	-8	-8

Time between vertical tics 5.0 Seconds

Enable Graph of:
- [x] Evap Inlet Temp
- [x] Evap Outlet Temp
- [x] Evap Super Heat
- [x] SuperHeat Err (C)
- [x] Output (V)

Show Transfer Function
- (•) Continuous Time
- ( ) Discrete Time

Run Controllers
- (•) Super Heat
- ( ) Test Mode

Disable
- [ ] Output Grid
- [ ] Graph
- [ ] TF Grid

Exit

Start Acquisition

Change Controller

Electrical and Computer Engineering,
Mechanical Engineering,
VELab, FURP, & CEEP
University of Michigan-Dearborn

Figure 12

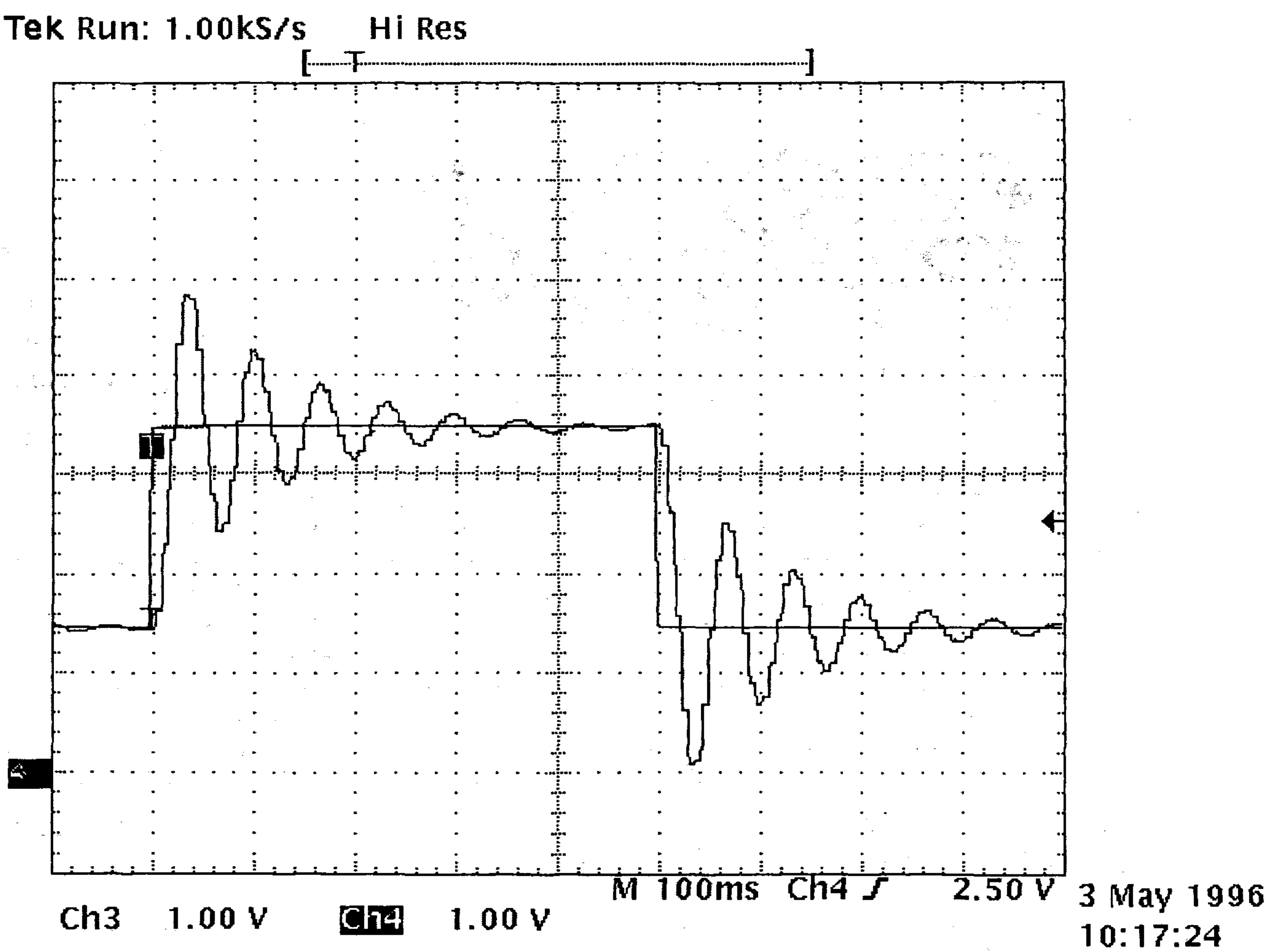

Tek Run: 1.00kS/s Hi Res
Ch3 1.00 V
Ch4 1.00 V
M 100ms Ch4 ∫ 2.50 V
3 May 1996
10:17:24

Figure 13

970109

Hot-Wire Velocity Measurements of Defroster and Windshield Flow

Bashar S. AbdulNour
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Hot-wire anemometry is used to experimentally determine the velocity field in the vicinity of the defroster nozzle jet flow and windshield interior surface. It provides quantitative velocity measurements necessary to determine the heat transfer coefficient on the windshield. Measurements are useful when used as a design tool to predict defroster and windshield flow. A modified traverse device was constructed to carry the hot-wire probes. The results obtained from the modified technique are more accurate and provides a valuable insight into the boundary-layer structure over the windshield.

INTRODUCTION

Measurements of jet flows are of interest in many automotive applications. For example, hot-wire measurements were used to investigate the velocity field due to automotive air registers (AbdulNour et al., 1996). In particular, these measurements can be used for design and development purposes. Also, if numerical codes can successfully compute these 3-D flows, then the verified computations can be used as an analytical design tool. Velocity surveys on the interior windshield surface of a 1995 Lincoln Mark VIII vehicle due to a defroster jet are reported herein. The present study identifies the techniques and results from this investigation. Specifically, it identifies the preparation for the flow visualization and testing results. It also identifies the hot-wire techniques that were used and presents the velocity field results.

The standard technique to perform hot-wire measurements on the windshield entails the use of a plexiglas template to replace the windshield. A straight hot-wire anemometer is inserted into the holes of the template to a specified depth into the interior surface (usually 7mm) to carry out the measurements. The probe is rotated until a maximum signal is detected on the signal processor. The magnitude and direction of the velocity is then determined, with the direction being the normal to the straight hot-wire. A typical result is shown in Fig. 1.

The present experimental technique provides a significant improvement over previous practices, since the data are far more informative than those wherein one fixed z-location (normal to the windshield interior) were to be used. Also, measurements on the glass surface (and not the composite material with upstream holes) are clearly more representative of the prototype flow field. Flow disturbance effects by such holes is of concern, and the theoretical smoothness constraint of covering the holes represents a formidable challenge to the conduct of an unbiased experiment. Mean and rms velocity data were also acquired in systematic traverses above the exit plane of the defroster nozzle. These surveys showed significant non-uniformities. The details come at the expense of development time and material as well as machining costs, but are deemed worthy for the enhanced contribution of the results.

FLOW CONDITIONS

Throughout this investigation, the total volumetric flow rate of air through the defroster nozzle was 101.4 cfm. This flow rate corresponds to a vehicle in defrost/max heat mode. The average uniform velocity U at the exit of the nozzle was calculated to be 5.2 m/s based on the measured effective exit area of 0.0092 m^2. Examining the Reynolds number of the flow based on the hydraulic diameter D_H, we obtain

$$Re_H = \frac{UD_H}{\nu} \approx 1.21 \times 10^4$$

at the nozzle exit. In the above, D_H is defined as four times the vent exit cross-sectional area divided by the wetted perimeter (D_H = 0.037 m), and v is the kinematic viscosity at a temperature of 300 K and atmospheric pressure. This value for the Reynolds number suggests that the flow is indeed turbulent.

FLOW VISUALIZATION

A template, using a nominal 100x100 mm square grid array, was used to define the measurement locations. Flow directions were obtained using wool tufts attached to the inside of the windshield. Velocity surveys of the streamwise mean velocity magnitude $\bar{V}_s = [\bar{u}^2 + \bar{v}^2]^{1/2}$ and rms intensity $\tilde{V}_s = (\overline{V'^2})^{1/2}$ were acquired at the accessible sites; u and v are the x-component and y-component velocity, respectively. The former are characterized by typical wall jet velocity distributions: $V_s = 0$ at z = 0 (plane of windshield), $\bar{V}_s$ = maximum at $z \approx \delta$, and $V_s \rightarrow 0$ as $z/\delta \geq 5$. The $V_s(x,y,z)$ time series data were taken at 7-10 z-locations over the windshield surface. The above nomenclature along with the general shape of the wall jet velocity profile are illustrated in Fig. 2.

The template, previously used as a replacement windshield with holes for probe inserts, was used to mark (via a grease pencil) the measurement locations on the outside of the windshield. Wool yarn was then strung through these marked points and secured to the outside of the windshield using transparent tape. Wool tufts for flow visualization were then attached to the inside of the windshield at the crossing points of the yarn grid. These were made by pulling a few threads of the yarn a short distance from the full body of the yarn tuft. (The latter was of order 15-20 mm in length.) It was ensured that the tufts were supported without bias, *i.e.*, that the bending strength of the few threads was sufficiently small.

Photographs were taken of the tuft orientations in the defroster jet flows by exposing the film for a suitably long period to capture the extent of the tuft oscillations. No attempt was made to average the resulting images. The mean flow direction, $\bar{\theta}(x, y)$, where $\theta = \tan^{-1}(v/u)$ as shown in Fig. 3, was inferred as the midpoint of the extremes of the tuft positions. The very slow response time of the massive tuft (i.e., $\rho_{tuft} \gg \rho_{air}$) made it problematic to use information such as relative light exposure to infer the percentage of time spent in a given portion of the angular range. In most cases, this inference is quite acceptable in the following sense. The hot-wire will essentially rectify the velocity reading as

$$V_{indicated} = V_{actual} \cos(\theta_{actual} - \theta_{probe}) .$$

Hence, even if the excursions of the tuft directions are ±20 degrees, the uncertainty is bounded by $V_{indicated} \leq (\pm 0.94) V_{actual}$ with a most probable excursion nominally equal to 40% of this θ value. Hence, for this estimate, the cosine is 0.99 and the estimated actual velocity is

$$V_{actual} = \frac{V_{indicated}}{0.99} = 1.01\, V_{indicated} .$$

The data from the photographs was transcribed to a graphical representation to show the mean values of the orientations $\bar{\theta}$; refer to Fig. 3. Specifically, x = 0, y = 0 are at the center and lowest point of the measurement grid and with y > 0 representing positions on the driver's side. Hence, a right-hand coordinate system has z > 0 pointed into the passenger compartment.

VELOCITY FIELD MEASUREMENTS

Velocity Survey Capabilities and Probe Alignment - An inability to specify a priori the most effective heights for the hot-wires due to the variable boundary-layer thickness led to a reconsideration of the standard approach. Additionally, the presence of cutout holes upwind of the measurement sites was a concern. If sealed by plugs it would be difficult to ensure a satisfactorily flush condition (≤ 0.25 mm). If sealed by tape, there would be a wide field of distributed roughness elements with dimensions large enough to disturb the flow. Note that any such roughness should be less than 5 *wall units* (Potter and Foss, 1982) to have a negligible effect. It is difficult to estimate v/u_τ for this flow field. However, if one uses $200 \doteq \delta/(v/u_\tau)$ as a minimal ratio between δ and a wall unit, then for $\delta \doteq 0.7$ cm, $v/u_\tau = 0.035$ mm, and a roughness element should be smaller than 0.18 mm, which is a challenging target. The present technique, in which there are no upwind disturbances for a given measurement, is a clear improvement.

An initial traverse device was constructed and used for initial measurements at 10 x,y-locations and 7 z-locations off the surface. The probe axis was aligned with the previously determined flow direction at each point and was affixed (on the interior of the windshield) through a magnet on the outside of the windshield. This device, although serviceable, incorporated one z-location measurement for one traverse position. It was elected to fabricate another unit, depicted in Fig. 4, that could accommodate 2 or 3 probes on a common support. This support could be removed and transferred to the

traverse rig for pre- and post-calibrations without risking a break in the electrical connections. (Note that such a disconnect/connect sequence can significantly alter the calibration values for a hot-wire.)

The second unit was then used to re-execute the z-surveys at the original 10 locations as well as at the remaining measurement sites that were accessible given the windshield-dash height restrictions. The time series data from these surveys were then used to create $\bar{V}_s(x, y, z)$ and $\tilde{V}_s(x, y, z)$ values. This traverse device presents a relatively bluff obstruction to the flow. A series of measurements were executed that showed there to be, at most, a negligible contamination of the measurements as a result of the traverse device.

Reference Velocity - Given that the flow system might not operate at a fixed condition, care was taken to continuously monitor the pressure at an arbitrary position in the supply plenum. A lead was inserted upstream into the defroster nozzle for this purpose. This lead was inserted into the plenum and the indicated pressure was used to define the state of the system. A reference velocity was then defined by the differential pressure between the plenum and the closed passenger compartment; that is

$$u_{ref} = \sqrt{\frac{2}{\rho}[p_{pl} - p]} \ .$$

The time averaged u_{ref} for a given 20-second sample $(\ )_i$ of the hot-wire was recorded and used as follows: $\bar{V}_s(x, y, z)]_i/u_{ref}]_i$ and $\tilde{V}_s(x, y, z)]_i/u_{ref}]_i$ were tabulated for each location. An ensemble mean value was then computed as

$$\langle u_{ref} \rangle = \frac{1}{N}\sum_{i=1}^{N} u_{ref}]_i$$

and the individual measurements are brought into relative agreement as

$$\bar{V}_s(x, y, z) = \frac{\bar{V}_s(x, y, z)]_i}{u_{ref}]_i}\langle u_{ref} \rangle \ .$$

This process was repeated for the standard deviations $\tilde{V}_s$. Using this procedure, the dimensional velocities at all locations are made compatible with each other in spite of the flow system variations.

RESULTS AND DISCUSSION

Velocity surveys over the windshield at surfaces spaced at z = 1.78, 6.5, and 17.8 mm from its interior surface are shown in Fig. 5a-c; the rms values of the velocity measurements at z= 6.5 mm are shown in Fig. 6. Also, detailed velocity measurements within the shear layer at several representative x- and y-locations (see Fig. 3) on the driver side of the windshield are shown in Fig. 7a-f. All the results are consistent with the wall jet velocity profile (in magnitude and direction) due to the impingement of the defroster jet flow on the windshield.

In addition, a hot-wire probe was traversed along the length of the defroster exit opening and at a distance of 24 mm above the grill surface; a schematic is shown in Fig. 8. We also determined the values of the velocity at 32 mm. Since the pressure gradient and effective shear terms (for the majority of the flow) were expected to be quite small, it was expected that these two planes would exhibit quite similar velocities. Because of this expectation, only a few repeat measurements were executed at 32 mm and 24 mm from the grill.

The results of the longitudinal survey along the centerline at 24 mm above the defroster exit on the driver side are presented in Fig. 9, while the details of the traverse velocity survey at positions 4 and 8 are shown in Fig. 10a-b. Subsequent processing, however, showed that substantial variations existed in these sample data. One source for such discrepancy is the presence of strong spanwise flows. Strong non-uniformities are evident in the velocity surveys. Note that the peak velocity is representative of the maximum stagnation pressure provided by the fan. The lesser velocities in the survey indicate inefficiencies of the delivery system in that the peak velocities show excess kinetic energy with respect to that required to provide a given flow rate. This effect can be expressed as the *kinetic energy flux correction factor* (Potter and Foss, 1982).

CONCLUDING REMARKS

The standard hot-wire measurement technique is susceptible to error due to the presence of holes in the windshield template. The employed modified technique provides more accurate results and clarifies the boundary-layer structure over the windshield. In addition, it is flexible and does not require a plexiglas template. Measurements can be used as a design tool to predict the velocity distribution due to defroster nozzle jet flow and determine the heat transfer coefficient on the windshield.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. John F. Foss and the research group at the Turbulent Shear Flow Laboratory, Department of Mechanical Engineering, Michigan State University for their assistance in collecting the hot-wire measurements.

REFERENCES

AbdulNour, B. S., Martin, C. J., and He, Q., 1996, "Experimental and Computational Investigation of the Flow Field Due to Automotive Air Registers", *Advances in Automotive Climate Control Design,* SAE Publication SP-1152, SAE Paper No. 960686.

Potter, M. C., and Foss, J. F., 1982, *Fluid Mechanics,* Great Lakes Press, Okemos, MI.

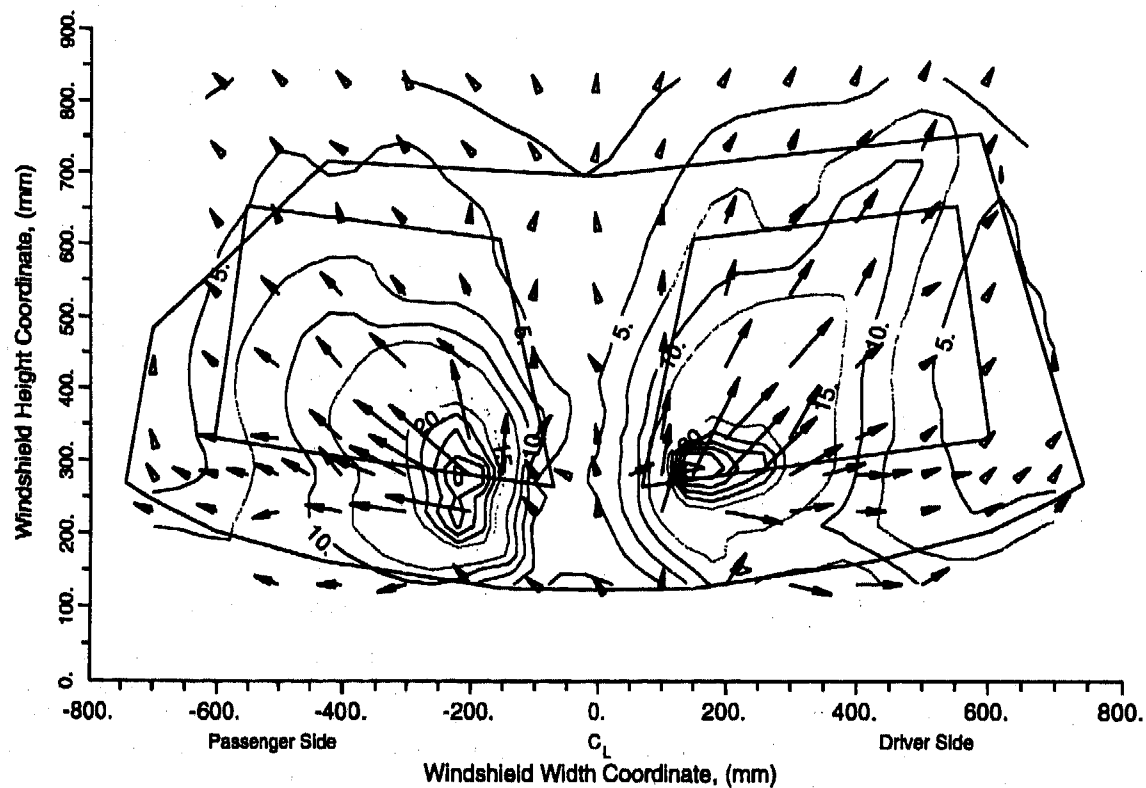


Figure 1. Experimental hot-wire velocities obtained from standard measurement technique with a plexiglas template windshield; vectors indicate flow direction.

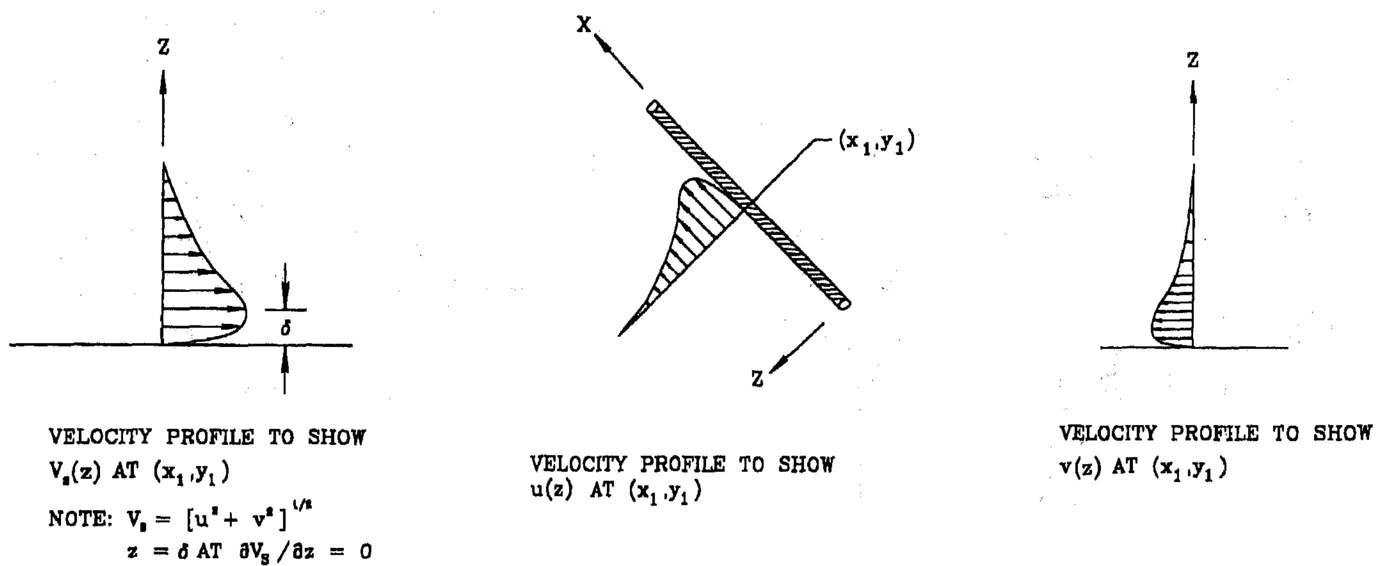


Figure 2. Typical representation of the wall jet velocity profile with nomenclature.

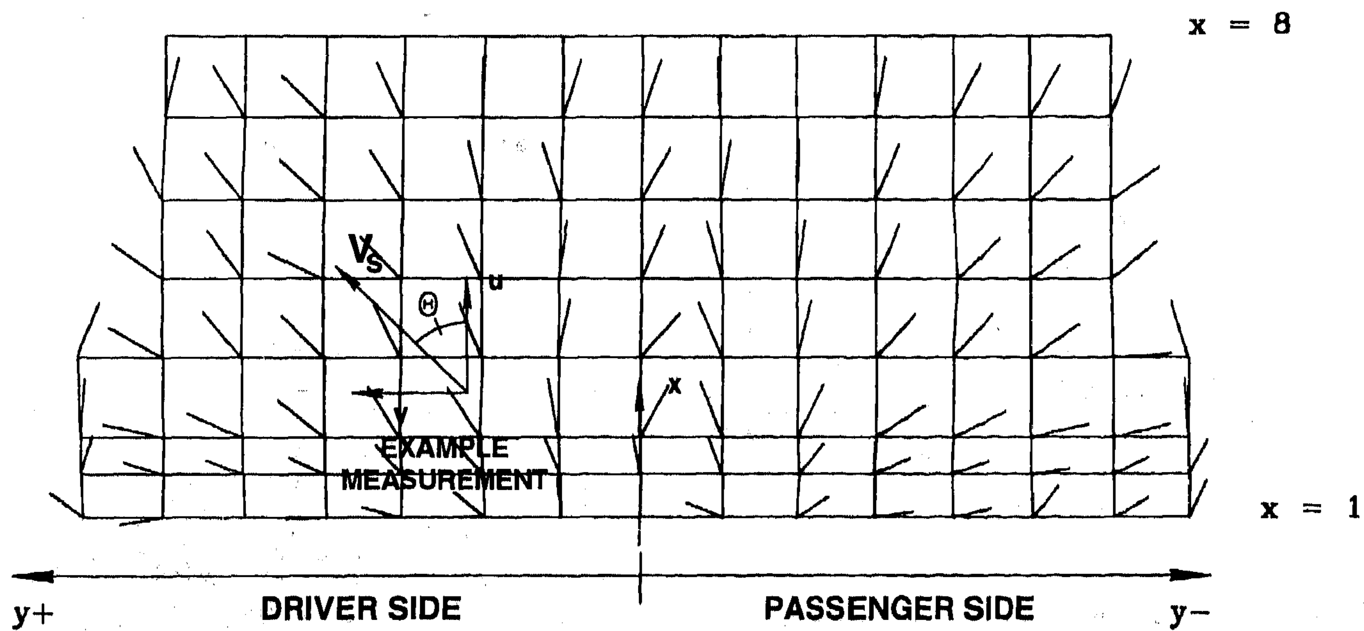


Figure 3. Coordinates for windshield velocity measurements and mean tuft orientations.

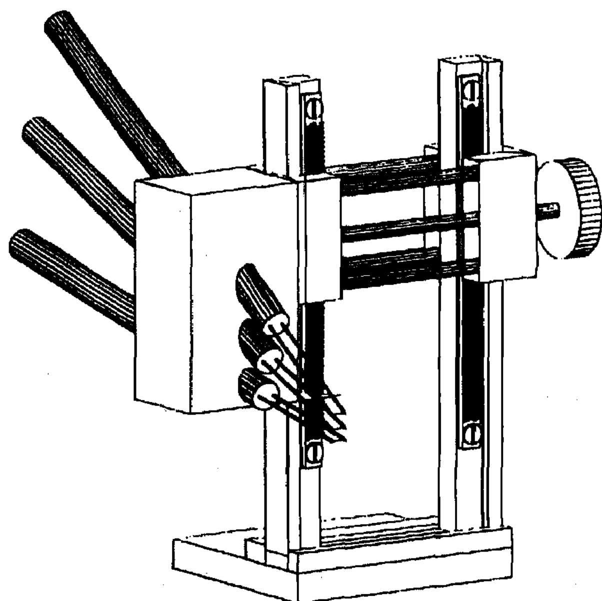
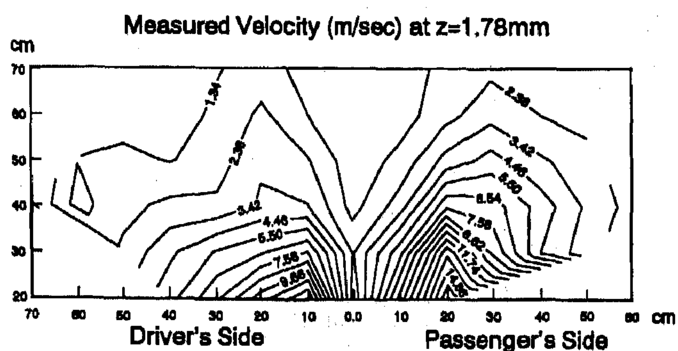
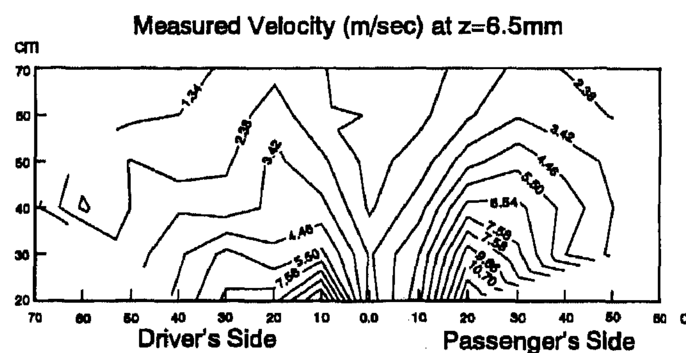


Figure 4. Perspective view of the 3-probe device.

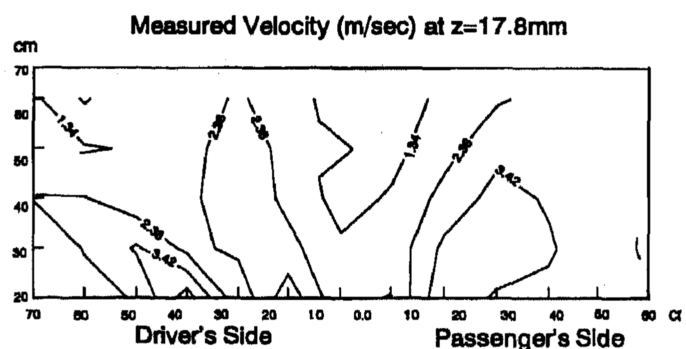


(a) z = 1.78 mm.

Figure 5. Survey of windshield wall jet velocity measurements at all x,y-locations and different z-locations.



(b) z = 6.5 mm.



(c) z = 17.8 mm.

Figure 5. (continued)

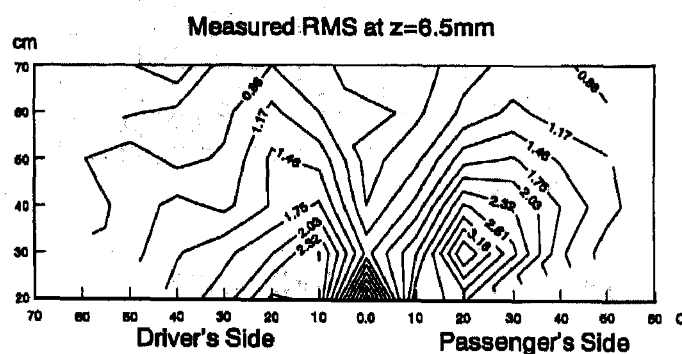
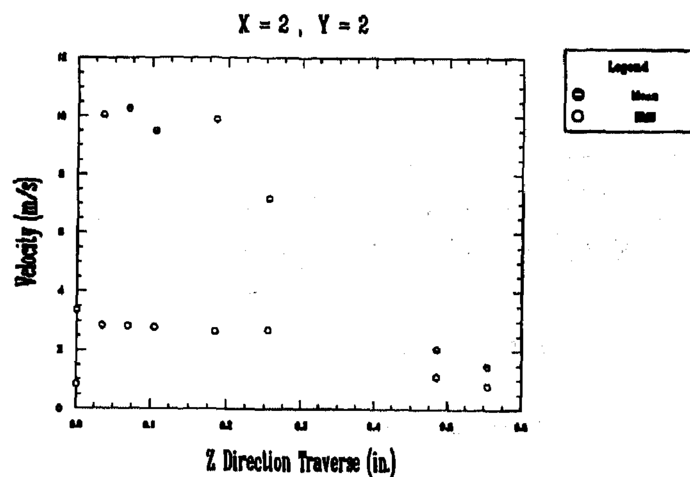
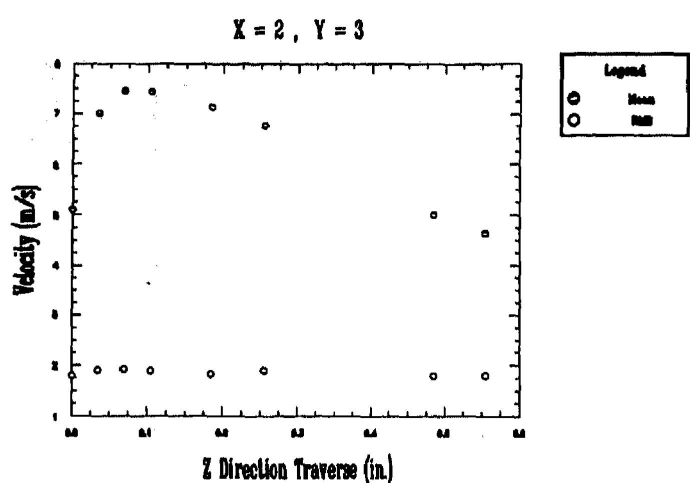


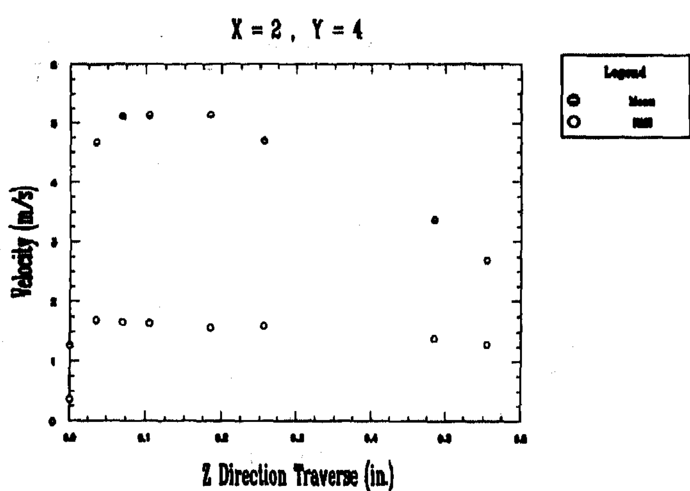
Figure 6. Survey of rms values of wall jet velocity measurements at z = 6.5 mm.



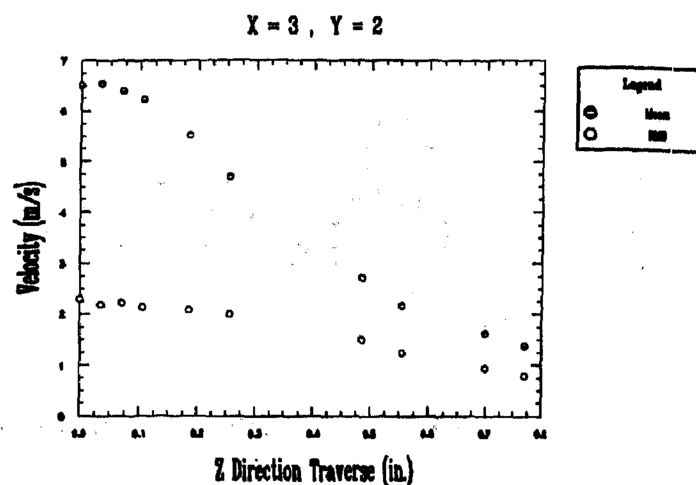
(a) Row x = 2 (280 mm); column y = 2 (200 mm).



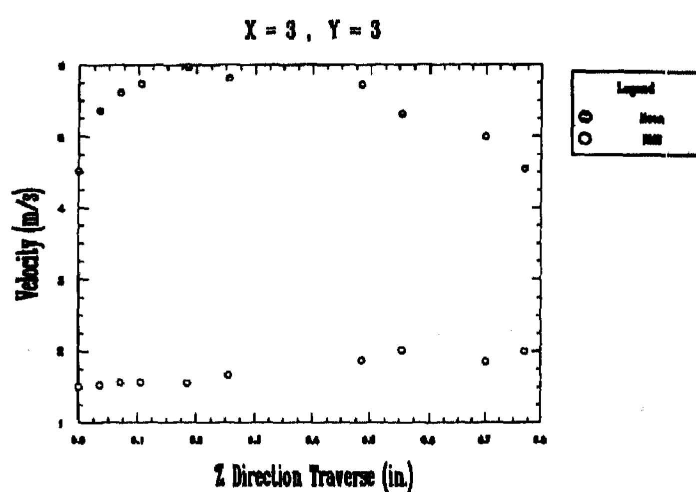
(b) Row x = 2 (280 mm); column y = 3 (300 mm).



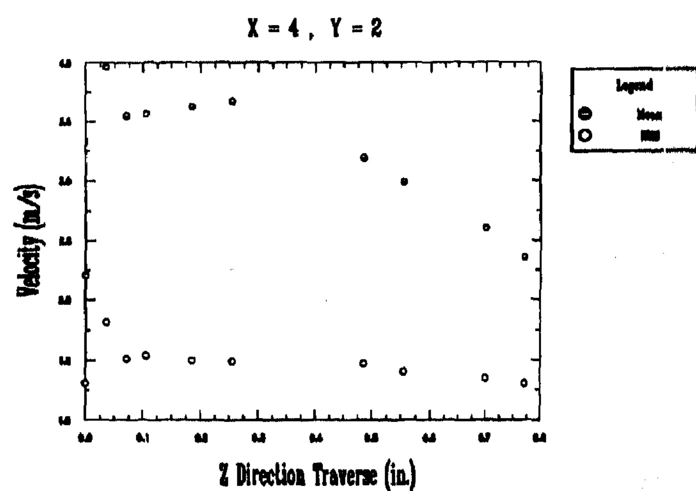
(c) Row x = 2 (280 mm); column y = 4 (400 mm).



(d) Row x = 3 (330 mm); column y = 2 (200 mm).



(e) Row x = 3 (330 mm); column y = 3 (300 mm).



(f) Row x = 4 (430 mm); column y = 2 (200 mm).

Figure 7. Detailed mean and rms velocity measurements at several representative x- and y-locations on the driver side of the windshield; $0.0 < z < 20.3$ mm.

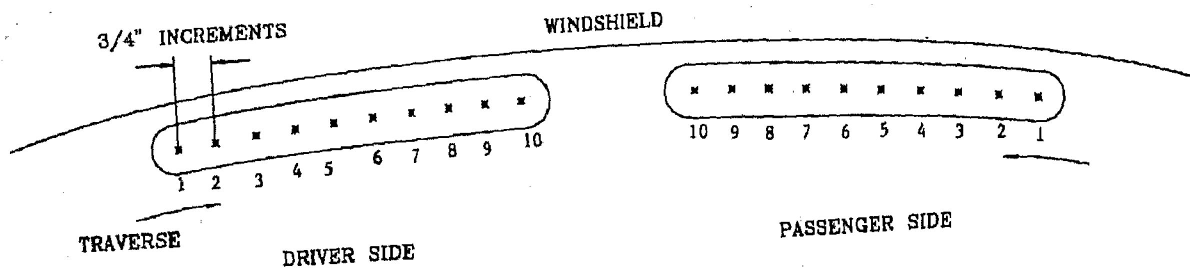


Figure 8. Schematic representation of velocity survey locations at defroster exit.

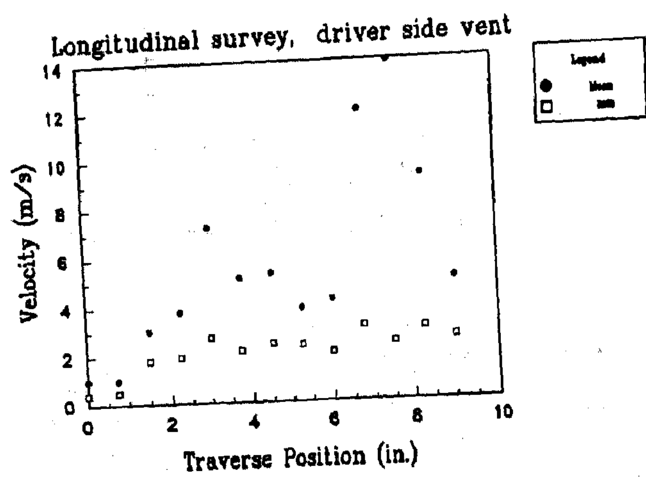
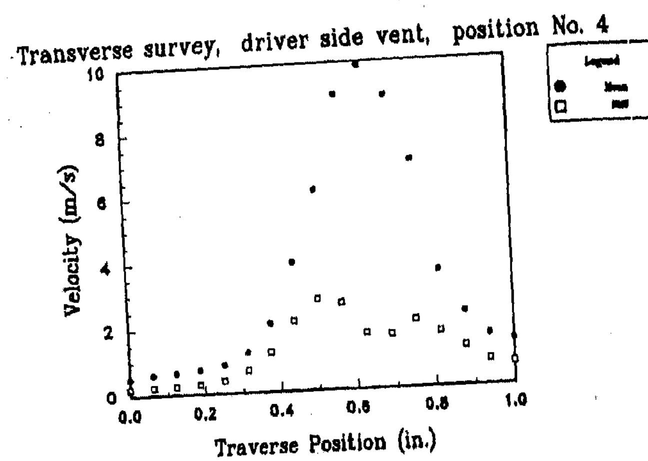
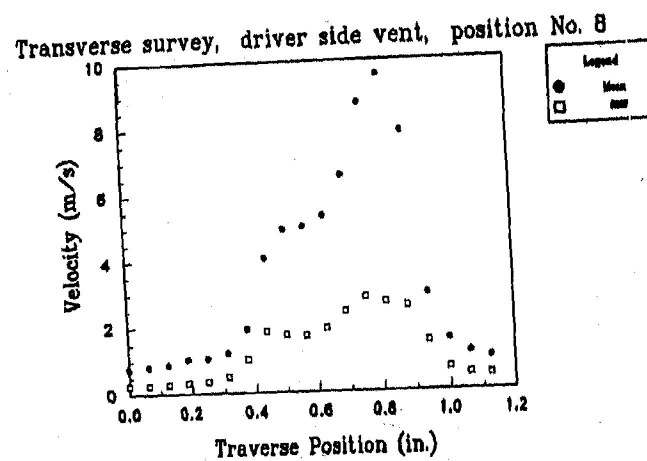


Figure 9. Velocity survey along the centerline at 24 mm above the driver side defroster exit.



(a) Position 4.



(b) Position 8.

Figure 10. Traverse velocity survey at 24 mm above the driver side defroster exit.

970110

Development of Sub-Cool System

Yasushi Yamanaka, Hiroki Matsuo, Kaoru Tuzuki, and Toshio Tsuboko
Denso Corp.

Yoji Nishimura
Toyota Motor Co. Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The important problems to be tackled from the point of view of preventing global warming are to save the power consumption of car air conditioning systems and reducing the refrigerant used in those systems.

Our approach to the task is to enhance system efficiency by utilizing sub-cooling at the outlet of the condenser and also to reduce the amount of refrigerant through a refrigerant behavior pattern analysis.
This approach has created a simple system construction using an integrated condenser/receiver device.

INTRODUCTION

The car air conditioning system has become indispensable in the provision of comfort and safety. The system, however, affects fuel economy and presents lingering concerns about global warming due to the HFC134a refrigerant, which has recently been adopted as a refrigerant to thoroughly solve the ozone depletion problem.[1][2]
Accordingly our most important task from an environmental point of view is to make a system which consumes less power and uses less refrigerant.

The purpose of this paper is to explain the behavior of the refrigeration cycle from the standpoint of the amount of refrigerant, performance and stability, and to achieve these requirements, in addition to suggest a car air conditioning system which fulfills the important factor of being easy to install.

The car air conditioning system is the simplest vapor-compression type refrigeration cycle, composed of a compressor, condenser, expansion valve, evaporator and refrigerant tank.

In conventional car air conditioning systems which are expensive and not easy to install, improved efficiency was achieved by improving the performance of system components such as condenser and compressor, rather than modifying the structure of such systems.
From the point of view of preventing global warming, the reduction of the amount of refrigerant is also important in addition to obligation regarding refrigerant recovery and the refrigerant recovery machine.
However, only attempts to reduce the size of each system component's inner volume have been made, and inroads into the review of the entire system have rarely been made.

This paper describes the system which achieves high efficiency together with a reduction of the amount of refrigerant through a review of the entire system.

BEHAVIOR OF CURRENT REFRIGERATION CYCLE

In many car air conditioning systems (Japan and EU, someparts of US), a receiver incorporating a dryer, a filter and a sight glass to check the amount of refrigerant are installed in the downstream of the condenser (Fig. 1). This receiver absorbs fluctuations in the required refrigerant amount inside the refrigeration cycle caused through changes in operation conditions, while maintaining constant performance and quality against leakage of refrigerant from hoses. Thus, the system is filled with refrigerant so that the liquid surface is formed in the receiver and only transparent liquid with no gas is observed through the sight glass.
Fig. 2 shows the behavior of the refrigeration cycle vs refrigerant amount. At point (A), only liquid is observed through the sight glass after gas completely disappears. Fig. 3 shows the location and amount of the refrigerant for this point. Recently, the condenser is filled with a smaller inner volume of refrigerant due to the introduction of the small capacity multi-flow condenser, while in receiver, there still exists a lot of refrigerant even at this point (A).

This extra amount of refrigerant (150g) should be attentioned. This is because disturbance of the liquid surface caused by refrigerant flowing in from the top of the receiver draws in gas from the outlet at the bottom of the receiver, thus requiring more refrigerant to maintain a certain liquid surface. In other words, this extra amount of refrigerant is required at least to make sure that no gas enters so that only the liquid flows out. From the standpoint of environmental protection, this extra amount of refrigerant must be attentioned.

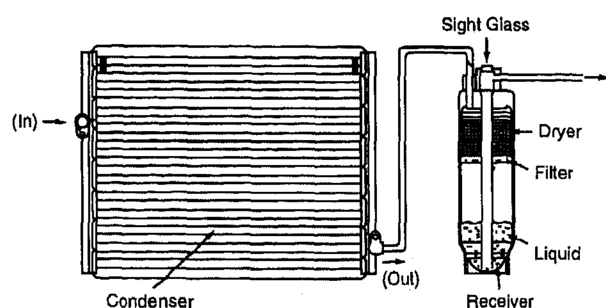


Fig. 1 Current Refrigeration Cycle

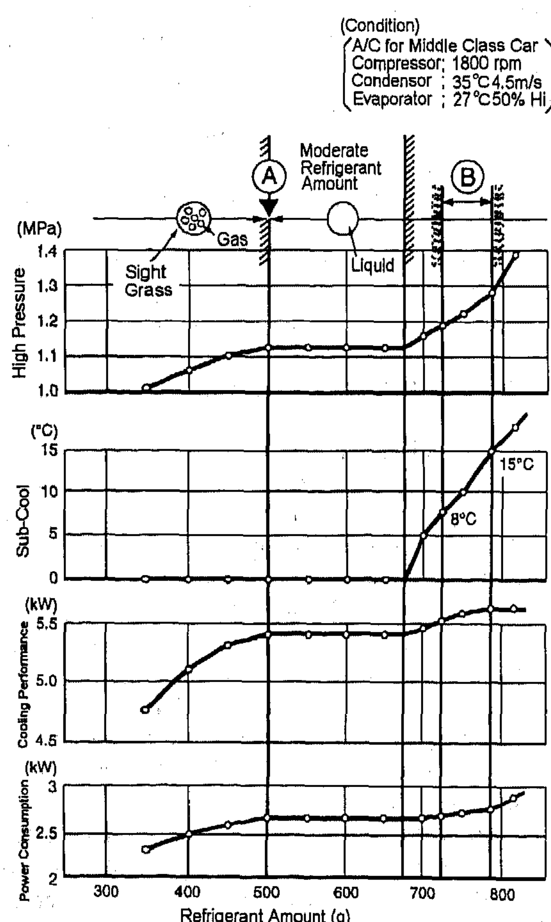


Fig. 2 The Behavior of the Refrigeration cycle vs Refrigerant Amount

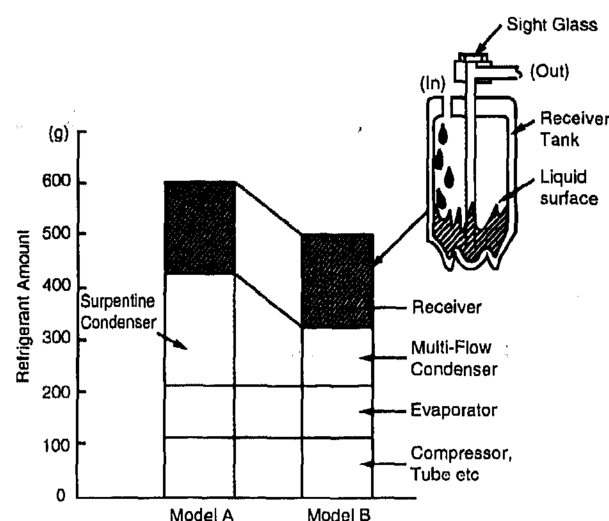


Fig. 3 The Location and Composition of the Refrigerant Amount for point (A)

In normal operation, liquid refrigerant supplied to the expansion valve is a saturated liquid having nearly zero sub-cooling, since liquid surface is always formed so that both liquid and gas exist inside the receiver. When this high-temperature saturated liquid is decompressed by the expansion valve, part of the liquid remains as gas, cooling down the liquid refrigerant. Thus, gas (approx. 40%) which does not contribute to cooling is generated and flows into the evaporator (Fig. 4).

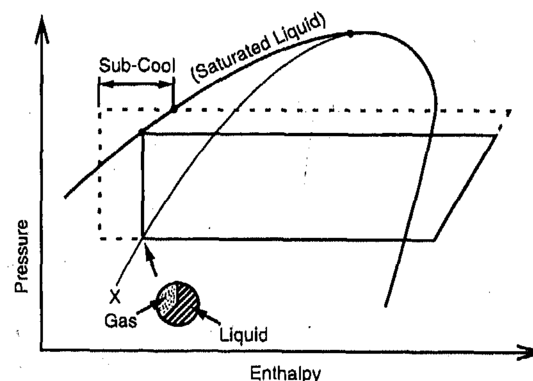


Fig. 4 The flows into the evaporator

If this high-temperature liquid refrigerant is cooled in advance, a smaller amount of gas will be generated, thus improving cooling performance. For instance, the amount of refrigerant can be increased to cause the liquid inside the receiver to overflow, thus making the downstream side of the condenser a liquid region. This region is then cooled by ambient air to cause sub-cooling. Fig. 2 shows this performance characteristic. Adding a further 50 to 100 g of refrigerant in addition to the refrigerant which has overflowed from the receiver can maintain sub-cooling

between 8 and 15°C, thus improving cooling performance by approximately 5% due to reduction of gas generated during decompression (Fig. 2 (B)).

However, it is still difficult to guarantee performance against fluctuation in refrigerant amount caused by changes in operation conditions and leakage of refrigerant. Furthermore, an increase in refrigerant amount is not preferable from the view point of environmental protection.

CONCEPT OF SUB-COOL SYSTEM

A receiver, which guarantees stable performance and quality, is installed in the passage of the refrigerant flow in the condenser, not at the exit of the condenser. Refrigerant is then supplied until a liquid region is developed inside the receiver, and the cooling section located downstream of the receiver is used as a sub-cool section where the liquid is cooled. Apart from that, this cooling section guarantees stable performance and also maintains quality since it liquefies gas which happen to flow from the receiver. The amount of refrigerant (150g) required at least to prevent the escape of gas from the receiver of the conventional system is no longer necessary. This results in the saving of refrigerant, even if an increase in refrigerant amount (50~100g) is expected in the sub-cool section. As a result, the sub-cool system not only helps in both improvement of performance and saving of refrigerant, but also reduces costs and weight, furthermore improving ease of installation by utilizing one of the headers of the multi-flow condenser to integrate a receiver with the header as shown in Fig. 5.

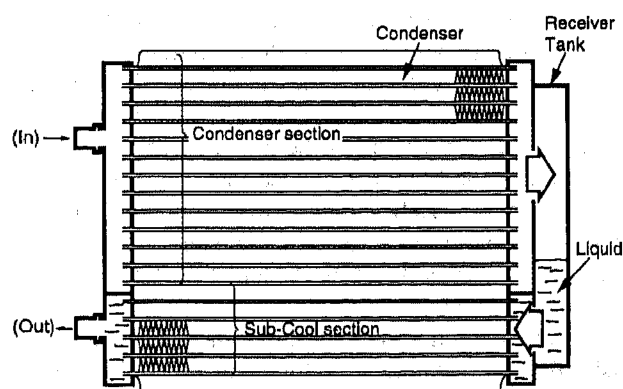


Fig. 5 Concept of Sub-Cool System

EXAMINATION OF OPTIMUM SUB-COOL SECTION RATIO

We now examine the optimum sub-cool section ratio for the multi-flow condenser.

Computer simulation was performed, and the results are shown in Fig. 6. From the point view of cooling performance and efficiency, the figure shows a wide range of the optimum sub-cool section ratio. Taking various operation conditions into account, it can be from 15% to 20%.

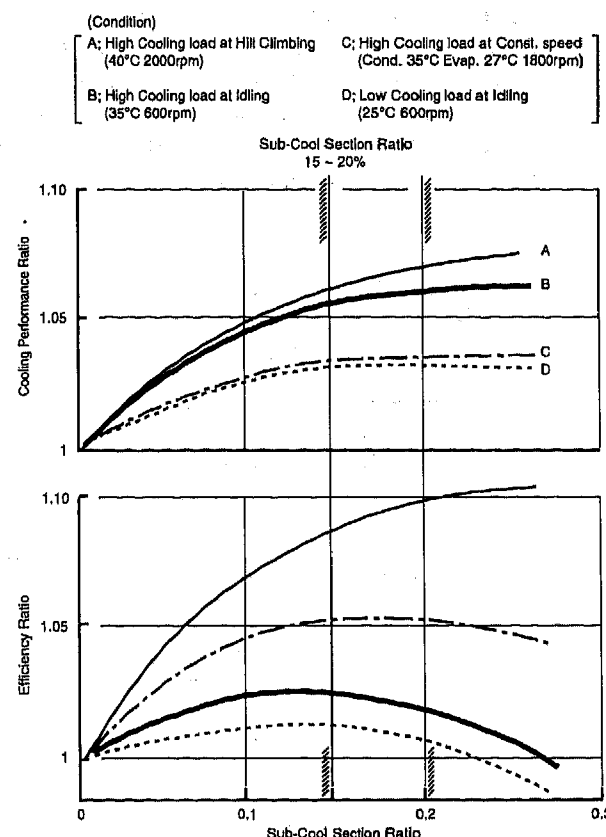


Fig. 6 Ultimate Sub-Cool Section Ratio

EXAMINATION OF LIQUID STORAGE FUNCTION

To integrate a liquid tank with the condenser's header and facilitate installation of the tank, the tank must be long and thin. If refrigerant flows into the receiver from the top of the tank and flows out from the bottom as in the conventional receiver, no liquid surface will be developed, thus liquid and gas will be mixed if the tank is thin. As a result, sub-cooling cannot be constant irrespective of the refrigerant amount, as shown in Fig. 7.

We therefore examined a method in which a liquid storage function is provided in upper space of refrigerant flow. Fig. 8 shows its operating principle.

(1) Some of the gas flowing into the entrance enters the gas region inside the tank due to buoyancy, but the gas condenses on the cooled inner surface, thus resulting in a balanced liquid surface. (Fig. 8-b)

(2) Accordingly, when more gas flows into the entrance, more gas enters the gas region, causing the liquid level to drop, thus supplying refrigerant into the refrigeration cycle. (Fig. 8-a) Supply of refrigerant

continues until state (1) is reached.

(3) When refrigerant no longer contains gas and sub-cool is developed, liquid refrigerant inside the tank also cools down. This causes gas in upper space to condense. As a result, inner pressure decreases, liquid surface rises, and finally extra refrigerant in the refrigeration cycle is stored. (Fig. 8-c) This continues until state (1) is reached.

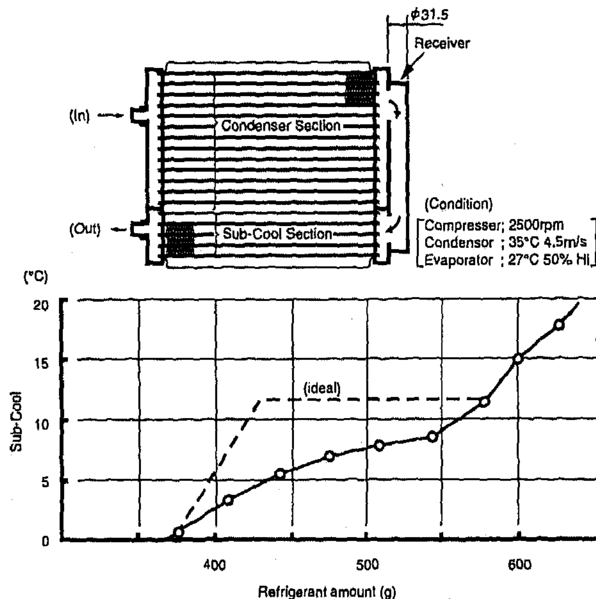


Fig. 7 The Behavior of Sub-Cool vs Refrigerant Amount

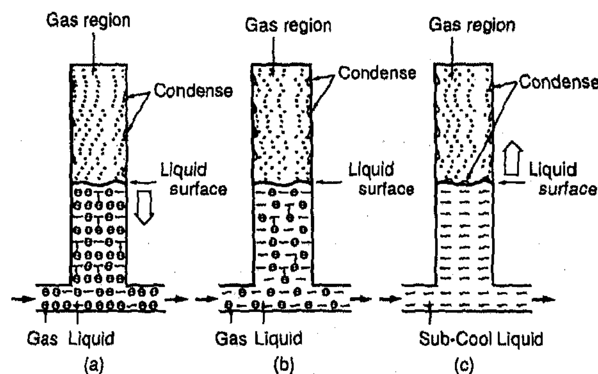


Fig. 8 The behavior of Liquid Storage Function

Fig. 9 shows the characteristic of the refrigerant storage which employs the above principle. From the figure, development of the liquid surface can be observed. It can also be observed that the more the refrigerant amount increases, the more refrigerant is stored in the tank.
Furthermore, stable sub-cooling is maintained at the exit of the sub-cool section as long as the liquid surface exists.

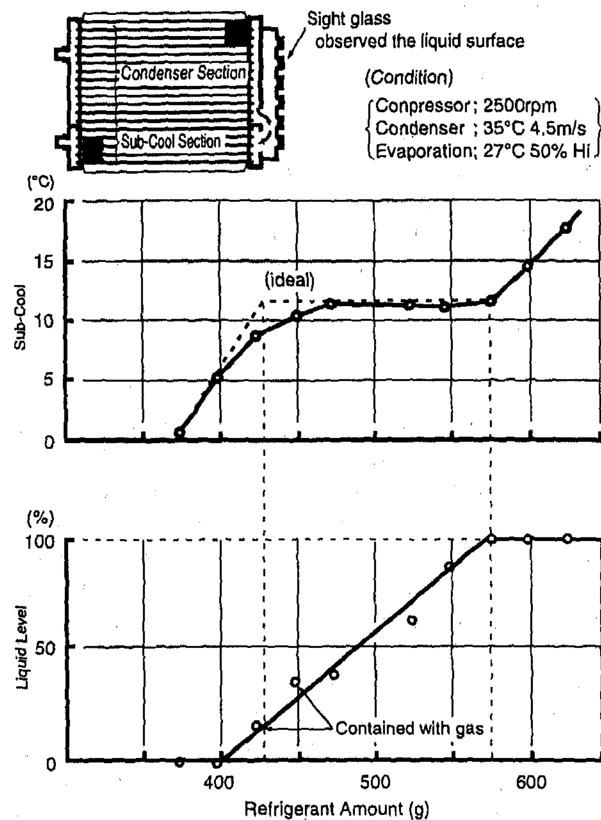


Fig. 9 The behavior of Sub-Cool vs Refrigerant Amount

STRUCTURE OF THE DEVELOPED SYSTEM

Fig. 10 shows a detailed structure of the developed system and its external view. A multi-flow condenser is used and a receiver tank with an inner diameter of 31.5 mm is integrated in one of the headers. To use approximately 20% of the lower section of the heat exchanger as the sub-cool section, the header is divided into condenser and sub-cool sections. Refrigerant flows from the condenser section into the receiver tank and then out into the sub-cool section. Liquid is stored on the top of the tank.

A dryer enclosed in a felt bag is installed inside the tank, and a filter is installed at both the entrance and exit at the bottom of the tank.
As with the conventional receiver, functions other than the liquid storage function are also provided. Like the conventional system, the sight glass located in the liquid line is used to check for shortage of refrigerant.

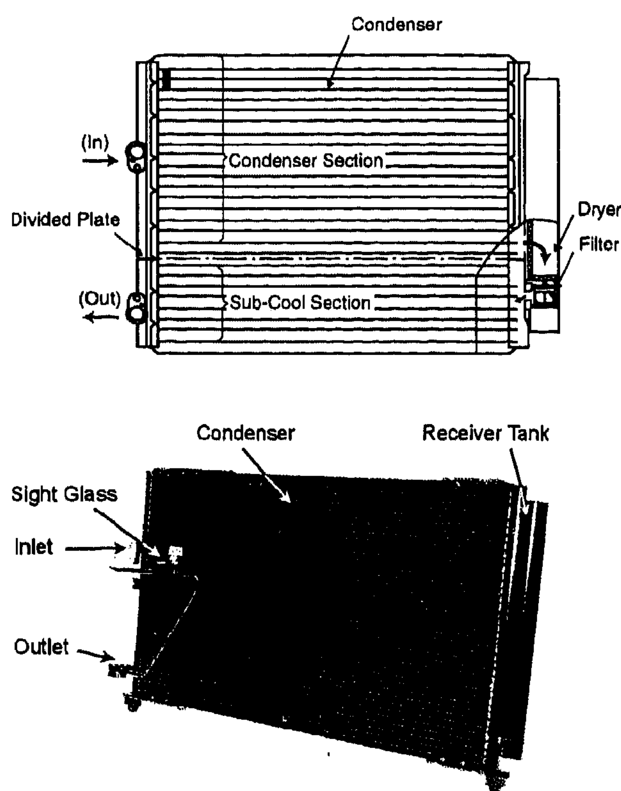


Fig.10 Structure of Sub-Cool System

EFFECTS

Evaluation results for effects of the developed sub-cool system are given below.

Fig. 11 shows comparison of refrigerant amount characteristic between the systems using the conventional receiver and the developed sub-cool system, in which the same size condenser used. The conventional system is stable in the range where no sub-cooling exists, whilst the developed system is stable at approximately 13°C of sub-cool.

This sub-cool has improved the cooling performance by 5% as shown in Fig. 12. This means that outlet air temperature can be reduced by 1.5°C and room temperature reduced by 1°C in the case of cool-down performance of an actual vehicle. Reducing the discharge displacement of compressor to provide the same cooling performance results in a saving of power by 10%.

In addition, when compared to a system using the conventional receiver which requires extra refrigerant to prevent the escape of gas, the developed system saves an average of 100 g of refrigerant as shown in Fig. 11.

CONCLUSION

We have developed a sub-cool system which can improve the efficiency of car air conditioning system and reduce the amount of refrigerant. The system employs a liquid storage function in the passage of the refrigerant flow in the condenser, and cools down the liquid in the cooling section to cause sub-cool.

Since the liquid storage function is provided as a long thin receiver tank integrated into the multi-flow condenser' s header by welding, the function can be installed easily. This was made possible since it was found that a long thin tank can be provided in upper space of the refrigerant flow to store liquid.

This system has a power saving of 10% whilst maintaining the same cooling performance and a reduction in refrigerant amount of 100 g.

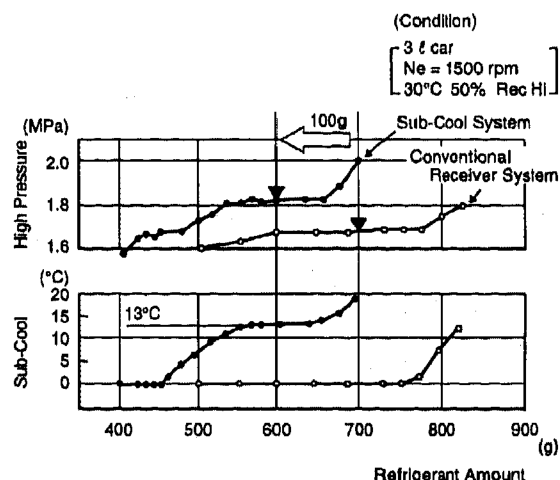


Fig. 11 Refrigerant Amount Characteristic

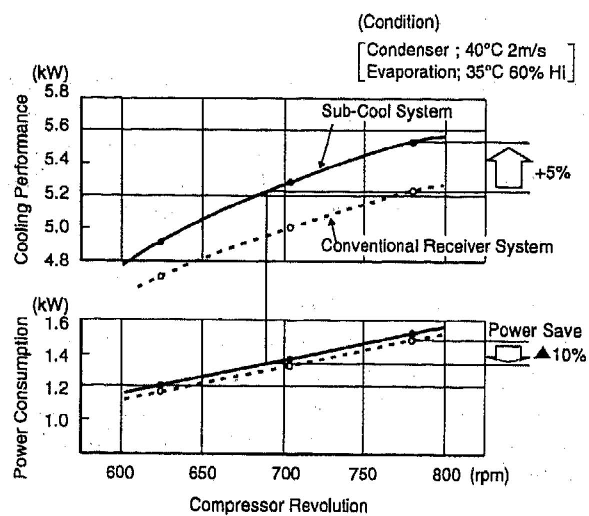


Fig. 12 Effect of Sub-Cool System

REFERENCES

(1) Toshio Hirata: "Automotive Air Conditioning system Using HFC-134a Comparison of Refrigeration Cycle Characteristics of CFC-12 and HFC-134a"; SAE 930229

(2) J. R. Morley: "Trends in Environmental Issues and Implications for Automotive Air Conditioning"; C496/006/95, UTMS, London"

970112

Determination of Adsorption Parameters of Active Carbon Filters Measuring Only the Beginning of the Breakthrough Curve

Thomas Wolff
Helmut Sandler GmbH & Co. KG

Heinz Bittermann
Wix-Helsa Filtration Products, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The adsorbed mass is an important parameter for the characterization of adsorptive filters. For its calculation it is necessary to determine the breakthrough versus time for a given upstream concentration. In order to reduce testing time, emissions and the amount of contaminants required, an evaluation method was developed that allow prediction of the breakthrough curve by extrapolating an aborted measurement. This new method reduces the amount of contaminants by more than 80%. Two additional parameters, which describe the characteristics of an adsorptive filter, result from the evaluation procedure.

INTRODUCTION

The total adsorbed mass and the breakthrough after a few seconds are parameters used to characterize an adsorptive filter. They have to be determined, for example, during the production procedure to ensure the quality of the produced filters. To determine the adsorbed mass, the integral over the breakthrough curve $P(t)$ has to be calculated. Using the ideal gas equation, the adsorbed mass is derived by calculating the difference between the whole amount of agent which the filter has been exposed to and the amount of gas having penetrated the filter (i.e. the integral over the downstream concentration vs. time). All these calculation steps are summarized in equation (1).

$$M_{ads} = M_{tot} - \frac{p\dot{V}}{RT} M_{mol} \int_0^{t_E} P(t)dt \qquad (1)$$

where M_{ads} is the adsorbed mass, M_{tot} the whole mass and M_{mol} the mol mass of the contaminant in g, p the ambient preassure in Pa, $\dot{V}$ the flow rate in m^3/min,

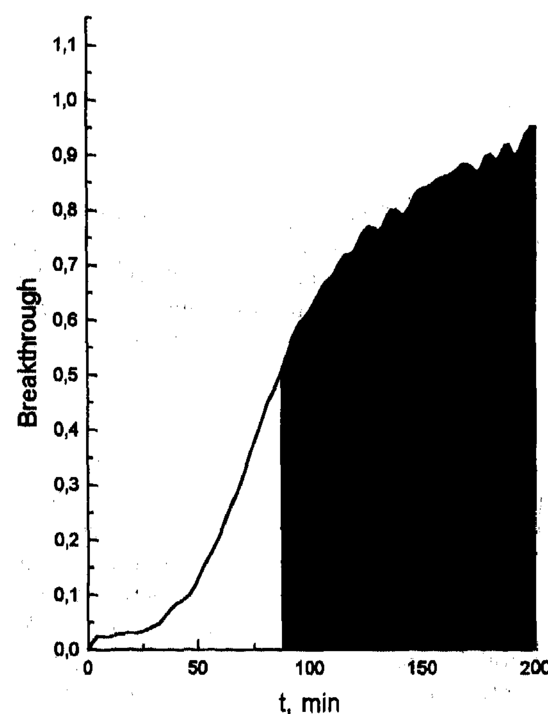


Figure 1: Typical breakthrough curve of an activated carbon filter using toluene as the contaminant. The shaded area under the curve represents the amount of gas, which, starting from the point at about 90 minutes, is not adsorbed by the filter.

a reasonable evaluation method with sufficient accuracy which reduces the required test time and the amount of contaminant by at least 50%. The emissions can be reduced by approximately 80%.

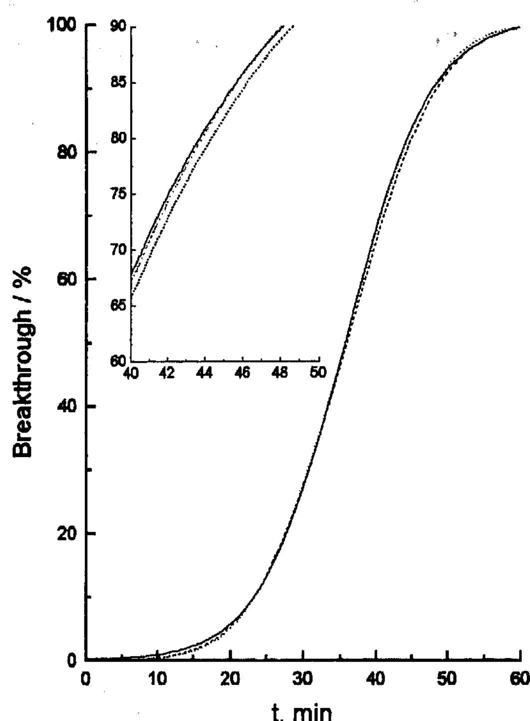


Figure 2: Two fits of equation (2) to a set of measured data (solid curve). The dotted curve is a fit to the whole dataset and the dashed curve is a fit to the first half portion of the same data. The inset shows an enlarged portion to demonstrate the difference between the two fitted curves.

$R = 8,3144 J/mol\,K$ the gas constant and T the ambient temperature in Kelvin. Equation (1) shows, that if $P(t)$ is known, the adsorbed mass can be determined. In figure 1 a typical breakthrough curve of an activated carbon filter, using toluene as the contaminant, is shown. The measurement time for such a curve takes up to several hours, depending on the filter type and contaminant. During the later phase of the test the major amount of the agent penetrates through the filter. The shaded area under the curve in figure 1 represents the amount of gas, which, starting from the point at about 90 minutes, is not adsorbed by the filter. This area is much larger than that one under the first section of the curve. It is therefore of major importance to find an evaluation procedure which enables us to determine the adsorbed mass from the data of an incomplete breakthrough curve.

It is the aim of this paper to demonstrate that a fit of the data of the initial section of a measured breakthrough curve to an appropriate function provides a reasonable evaluation method with sufficient accuracy which reduces the required test time and the amount of contaminant by at least 50%. The emissions can be reduced by approximately 80%.

EVALUATION PROCEDURE

The easiest way to derive $P(t)$ from the data of an initial fragment of a breakthrough curve is fitting the data to a function that describes exact the curvature of $P(t)$. In the example shown in figure 1 and in all other cases where $P(t)$ is given by a symmetrical S-curve, which begins at a given offset P_0 and reaches 100 %, the appropriate function is the error function $\mathrm{erf}(t) = 2\pi^{-1/2}\int_0^x \exp(-t^2)dt$. The complete form of the function, which can be taken for the fit is given by equation (2):

$$P(t) = \frac{P_s}{2}\left\{1 + \mathrm{erf}\left(\frac{t-t_b}{t_w}\right)\right\} + P_0 \qquad (2)$$

where P_s is a scaling factor and P_0 represents an offset. t_b and t_w are two additional parameters for the characterization of an adsorptive filter. The subscript b stands for breakthrough and the subscript w for width. The larger t_b, the longer is the "lifetime" of the filter. t_w describes the kinetics of the adsorption process. It can be compared to the width of a Gaussian distribution which leads into a convolution with a Heaviside function $\theta(t-t_d)$ to the error function in (2).

$$\mathrm{erf}(t-t_d)/t_w) = 2\pi^{-1/2}\exp(-(t/t_w)^2) * \theta(t-t_d)$$

The unit of t_b and t_w is *minutes*. The integral over the function in equation (2) is easy to solve and we get the following expression for the total adsorbed mass

$$M_{ads} = M_{tot} - \frac{p\dot{V}}{RT} M_{mol}\, P_s\,(\mathrm{F}(t_b, t_w, t_E) - \mathrm{F}(t_b, t_w, 0)) + P_0\, t_E \qquad (3)$$

with

$$\mathrm{F}(x,y,t) = \frac{y}{2\sqrt{\pi}}\exp\left(-\frac{(t-x)^2}{y^2}\right) + \frac{1}{2}(t-x) + \frac{1}{2}(t-x)\,\mathrm{erf}\left(\frac{t-x}{y}\right)$$

Using an appropriate method for a non linear fit, e.g. a Gauss-Newton or a Marquart-Levenberg method, with the four fit parameters P_s, P_0, t_b, and t_w, equation (2) can be fit to a set of measured data. In figure 2 two fits of (2) to a data set, derived from actual carbon filter media using n-butane as the contaminant, are shown. The dotted curve is a fit to the whole data set (solid

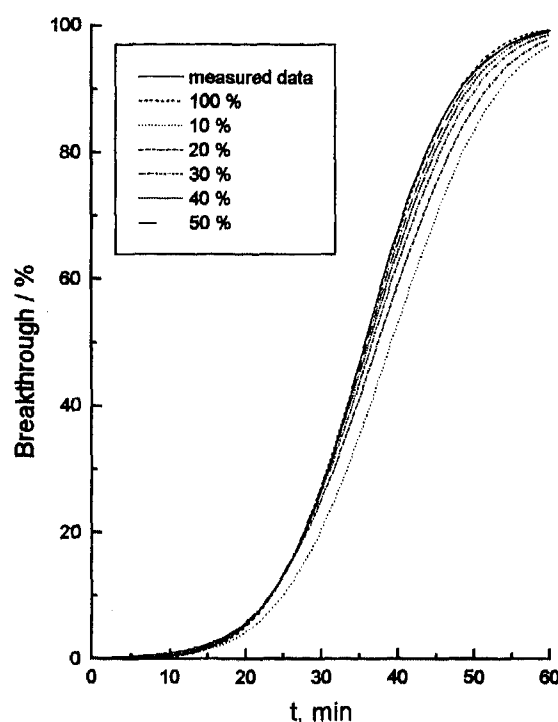


Figure 3: Fits of equation (2) to datapoints (solid line), which were cut off at six different levels (corresponding to table 1).

curve) and the dashed curve is a fit to the first half portion of the same data. The plot demonstrates, that both fits are in good agreement with the measured data (similar to the one shown in the example).

This indicates that measurements can be aborted whenever a breakthrough of about 50% is reached and the rest of the curve can be determined by fitting $P(t)$ to the real test data.

EVALUATION OF THE NEW METHOD

Figure 3 shows six different fits to the data of the last example. Each of the six curves corresponds to a value for the breakthrough, which represents the last point of the dataset used for the non-linear fit.

In table 1 the values of the breakthrough of the last datapoint are listed together with the corresponding error, which results from the difference of the real adsorbed mass to the value calculated from the fitted curve using equation (3). The values in table 1 indicate that the deviation from the actual value for M_{ads} is less than 1% if we interrupt the measurement at 50% breakthrough. This point represents the point of inflection of the curve. If the test is aborted close to this point the error will be less than 1%. Deviation increases as testing time is reduced.

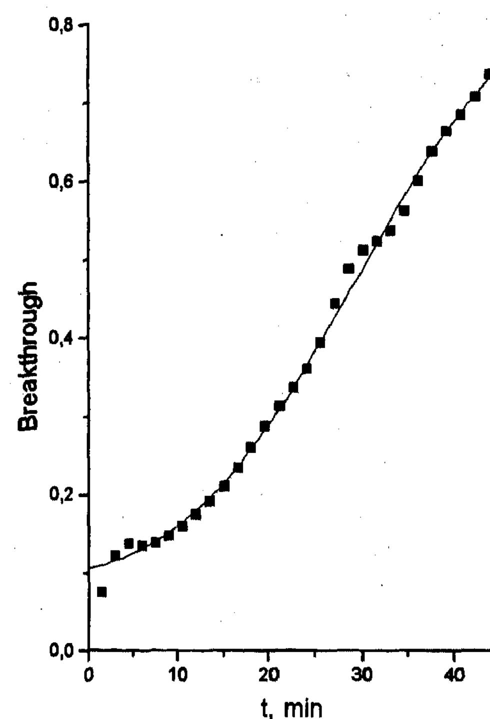


Figure 4: Fit of equation (2) to a breakthrough curve of a combination (thin bed) filter. The contaminant was toluene. The initial penetration (spontaneity) can be determined from the value of the fitted curve at $t = 0$.

Table 1

$P(t_e)$ / %	error / %
100	0.17
50	0.8
40	1.52
30	2.7
20	4.8
10	9.2

The size of the error depends on how good the measured data fits to equation (2). An evaluation of a large amount of measured curves of two different active carbon systems has shown, that the error, which we get when we cut the data near the point of inflection, is in the range of 0.7 – 3.0 %.

CONCLUSIONS

With the present evaluation method it is possible to cut down the testing time to 50% or less and the emissions can be reduced by about 80%. These two results are very important, whenever the adsorbed mass of a large amount of samples has to be determined, e.g. quality control during the production of adsorptive filters. The whole production and testing process can be accelerated and environmental pollution problems related to testing can be reduced to a minimum. The reduction of the costs for one measurement is another important argument for the application of this method.

Another very useful effect of this method is, that we can determine the breakthrough of the filter at the initial time of the measurement. This value, which we call the spontaneity of the filter, can normally be determined by extrapolation from the first datapoints to $t = 0$. If a fit to equation (2) was made, the spontaneity is directly represented by the value of $P(0)$. In figure 4 a fit to the breakthrough curve of a combination filter using toluene as a test agent is shown. The spontaneity according to the above mentioned proposal would read 10.5%. In this example the fluctuation of the first datapoints would have made an extrapolation inaccurate and difficult.

970113

Improvement of Scroll Compressor for Vehicle Air Conditioning Systems

Jiro Iizuka, Norio Kitano, Shigeru Ito, and Shinichi Otake
Sanden Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Mobile Air Conditioning Compressors with the scroll mechanism design are superior in noise, vibration and harshness (NVH) levels, and in efficiency.

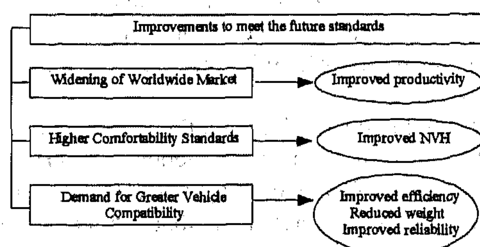
For further NVH improvements; for improved comfortability; to enhance the manufacturability to meet the increasing market volumes; and to give better vehicle compatibility in weight reduction and increased durability, the New Type TR Compressor is presented.

INTRODUCTION

BACKGROUND - Fifteen years has passed since mass production of the TR compressor started in 1981. During this period, four models of different displacements were developed, TR60, TR70, TR90, TR105. Additionally the controlled capacity compressor TRC model was developed and improvements were implemented to make the TR compressor line compatible with the new refrigerant HFC 134a.

DEVELOPMENT OF THE NEW TR MODEL

To meet the future performance standards for automobile air conditioners the New Type TR Compressor was developed.



PRODUCTIVITY

Improved productivity is gained by simplifying the parts structure.

EM BEARING - The current TR has a ball coupling as an anti rotation mechanism for the orbiting scroll member, and it serves as a thrust bearing. This ball coupling consists of a pair of races, a pair of retainer rings with beveled holes to retain the ball bearings and a set of ball bearings. This will be replaced by an Eccentric Motion Bearing (EMB) and will include only a pair of integrated races which constrain the ball travel, and a set of ball bearings.

Parts manufacturing and assembly operations are simplified and reduced with the EM Bearing and the quantity of parts is decreased.

BALANCING METHOD - The balance weight shown in Figures 1 & 2 must be offset with a clutch armature plate weight. The mass times distance of the counterweight, that is next to the scroll, is offset by the orbiting scroll centroid times the orbiting radius, and this couple alone provides adequate balance.

In the new TR compressor, the balance weight and armature weight are eliminated thus giving substantial productivity improvement and reduction in parts and weight.

Not only is weight reduced by the elimination of the balance weight and clutch armature counterweight, Figure 1 and cross-hatched section of Figure 2, but some mass is eliminated from the housing as well.

SEALING METHOD BETWEEN SUCTION AND DISCHARGE CHAMBER - In the current TR compressor the suction and discharge chambers are sealed off with an O-ring. In the New Type TR Compressor both chambers are sealed from each other by metal contact. This method improves assembly productivity. Figure 3.

SHAFT HOLDING METHOD - In the current TR the shaft position is held by the rear ball bearing on the crankshaft flange which butts against the shaft balance weight and the crankshaft flange. In the New Type TR Compressor, this ball bearing is replaced by a lighter weight, efficient needle bearing. Figures 4 & 5.

The shaft is held in position by the front bearing races being retained between the clutch armature hub and the snap ring. Figures 5 & 6.

These changes improve productivity of parts manufacturing and assembly.

HIGHER COMFORTABILITY STANDARDS

IMPROVEMENT OF NVH - In the current TR the radii of the beveled holes in the retainer rings are the same as the orbiting radius. The ball bearings, of smaller diameter than the beveled holes, then travel around the holes on the bevels as the scroll is traveling in an orbit.

The amount of bevel and the thickness of the retainer rings dictate the ball travel radius. The ball bearings can move off of the bevels at any change of condition because their travel is not constrained. Thus, a kind of ball bearing "backlash", occurs at many condition changes, particularly load changes. This degrades the NVH characteristics.

The New Type TR Compressor consists of an integral race and ball retainer which gives actual paths for the ball to travel in and thus the ball travel becomes constrained and true and backlash does not occur. Figure 7.

Also, the compressor moment of inertia becomes smaller due to the elimination of the clutch armature balance weight and the drive shaft balance weight. This then reduces the starting torque which improves the NVH characteristic at the numerous cycling events.

Figure 8 compares the sound pressure level in DBA between the standard TR and the New Type TR Compressor.

DEMAND FOR GREATER VEHICLE COMPATIBILITY

EFFICIENCY AND RELIABILITY - To meet the ever increasing A/C loads, the market demands an even greater capacity of the scroll compressor. The engine compartment packaging limitations, however, prevent the application of a larger compressor with a larger diameter to satisfy the needed increase in capacity.

Due to the cantilever aspect of the scroll wall at the discharge end, any increase in scroll wall height to increase displacement should be accompanied by a stronger root at the end wall section.

To gain toughness at the root center in the current design, the wall would have to be thickened and this would cause the starting points of the involute to be delayed, or in other words moved away from the center.

This would cause a shorter and more rapid compression process, and as a result, the gas leakage to the suction side at the center would be increased and this would degrade compressor efficiency (COP), and this would be unsatisfactory. Therefore, to increase efficiency and to give extra root strength for even taller scroll walls, the New TR will provide a new concept two step scroll wall. See Figure 9.

To improve efficiency the involute curve length has been increased by decreasing the angle from the center of the generating circle, A degrees and B degrees - current TR; to $A-\alpha$ degrees and $B-\beta$ degrees - New Type TR. The length increase of the involute curve is shown by comparing the involutes of Figure 9a with Figure 9b.

Figure 10 shows a Finite Element Analysis comparison between the current TR90 (Base Design) and the New TR design. The two step design shows a significant reduction in maximum stress and a 34% increase in the safety factor.

Figure 11 shows the crank angle vs. gas pressure. The two step center wall method gives a longer compression process at the final compression stage.

This method comes closest to reaching the ideal scroll compression cycle which has never been attained due to the difficulty in achieving sufficient strength at the scroll center wall.

By this improvement, a smoother torque, higher efficiency and lower pressure variations are achieved.

WEIGHT - Cars are becoming more efficient year by year as the demand increases for better fuel economy and lower emissions. Weight reduction at every area of an automobile carries the highest priority.

As discussed in the preceding paragraphs, the weight of the New Type TR Compressor is reduced by eliminating certain counterweight parts, redesigning of certain parts and by simplifying the structure.

The quantity of parts has been reduced by 18%. The weight of the compressor has been reduced by 10% from 5.9Kg to 5.3Kg.

RELIABILITY - The overall reliability has been improved by replacing the ball bearing with a needle bearing and by reducing the starting moment of inertia. This also reduces weight and size.

Included is a newly designed shaft lip seal with a rubber coated edge face that replaces the O-ring.

The seal reliability is the same as the O-ring type, but the structure and assembly is simplified giving an overall reliability improvement. See Figure 12.

COMPARISON OF ANTI-ROTATION MECHANISMS

The EM bearing has been compared to other anti-rotation mechanisms such as the Oldham Coupling, which is constructed of a crossed sliding apparatus, and the Pin and Ring Coupling which is constructed of a pair of pins and guide rings.

The Oldham Coupling and the Pin and Ring Coupling have low mechanical efficiency due to frictional forces. Also, the Oldham Coupling has a fundamental limitation in inertia force balancing that results from its reciprocal motion. Both designs, however, have good noise characteristics, thus they were evaluated.

The Ball Coupling of the current TR and the EM Bearing have the best mechanical efficiency due to the use of ball bearings located between the orbiting scroll and the base.

The EM Bearing gives better productivity than the Ball Coupling due to a decrease in the number of parts. Further, the EM Bearing gives a big advantage in noise reduction compared to the Ball Coupling. See Table 1, which shows the ranking of these characteristics among the four designs. From this comparison, the EM Bearing was designed as the most suitable coupling for the advanced scroll compressor.

CONCLUSION

The scroll compressor is a superior design because it has the highest volumetric and isentropic efficiency and has the lowest NVH characteristics of all types of vehicle A/C compressor mechanisms.

The New Type TR Scroll Compressor has been developed to further improve the NVH characteristics, to improve efficiency, and improve reliability.

By this design, the productivity and vehicle compatibility were greatly improved, thus meeting the widening market demands and acceptance.

REFERENCES

1. Kiyoshi Terauchi, et. al., "The Characteristics of the Spiral Compressor for Automotive Air Conditioning", 1983 SAE Congress Paper 830641.

2. Yasuo Kikuchi, et al., "Scroll Compressor Wall Strength Improvement", 1996 SAE Congress Paper 960964.

Current TRS090	NEW TR 90
·Counter weight. ·Shaft balance weight. ·Clutch armature weight.	·Counter weight only.

Figure 1.

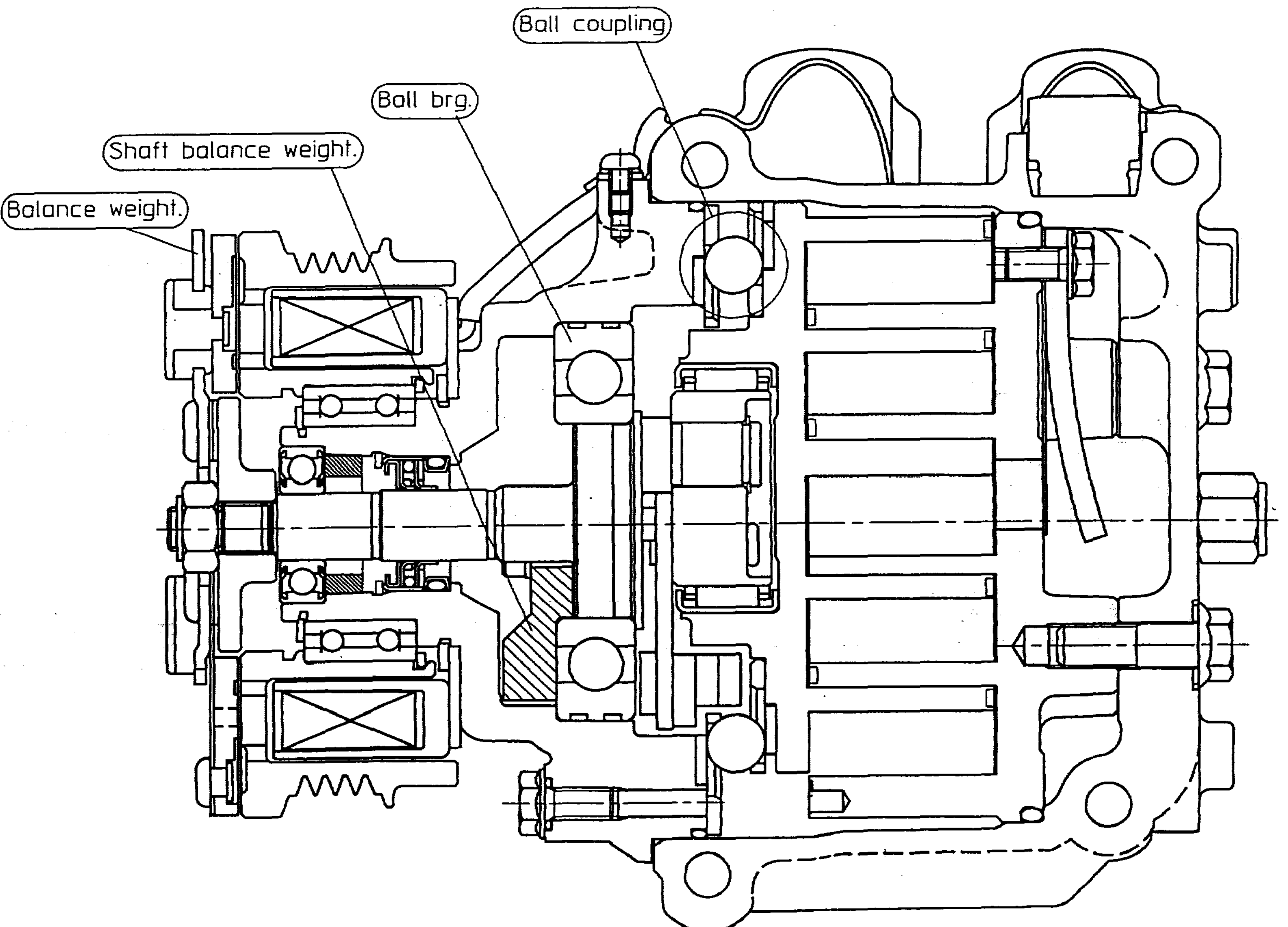

Current TRS090 compressor
Ball coupling
Ball brg.
Shaft balance weight.
Balance weight.

Figure 2.

CHANGE ITEMS	Current TRS090	NEW TR 90
Fixed scroll	·With protrusion. ·O-ring seal.	·Without protrusion. ·Without O-ring.

Figure 3.

CHANGE ITEMS	Current TRS090	NEW TR 90
Rear bearing	·Ball bearing.	·Shell type needle bearing. Compact. Light weight.

Figure 4.

CHANGE ITEMS	Current TRS090	NEW TR 90
Shaft holding method	*·Inner race of rear ball bearing between shaft balance weight and crankshaft flange.*	*·Inner race of front ball bearing between snap ring and clutch armature.*

Figure 5.

Change items on NEW TR compressor

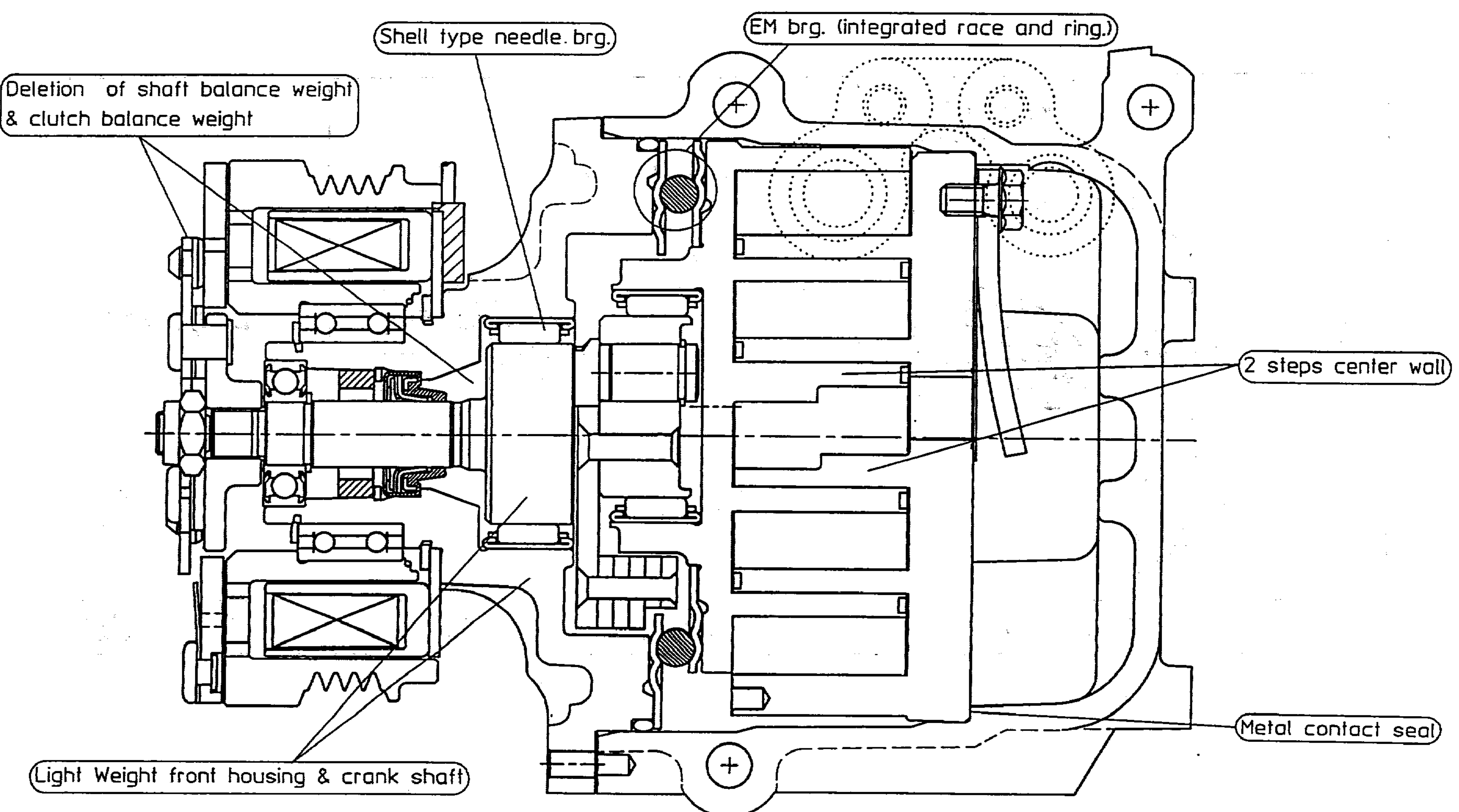
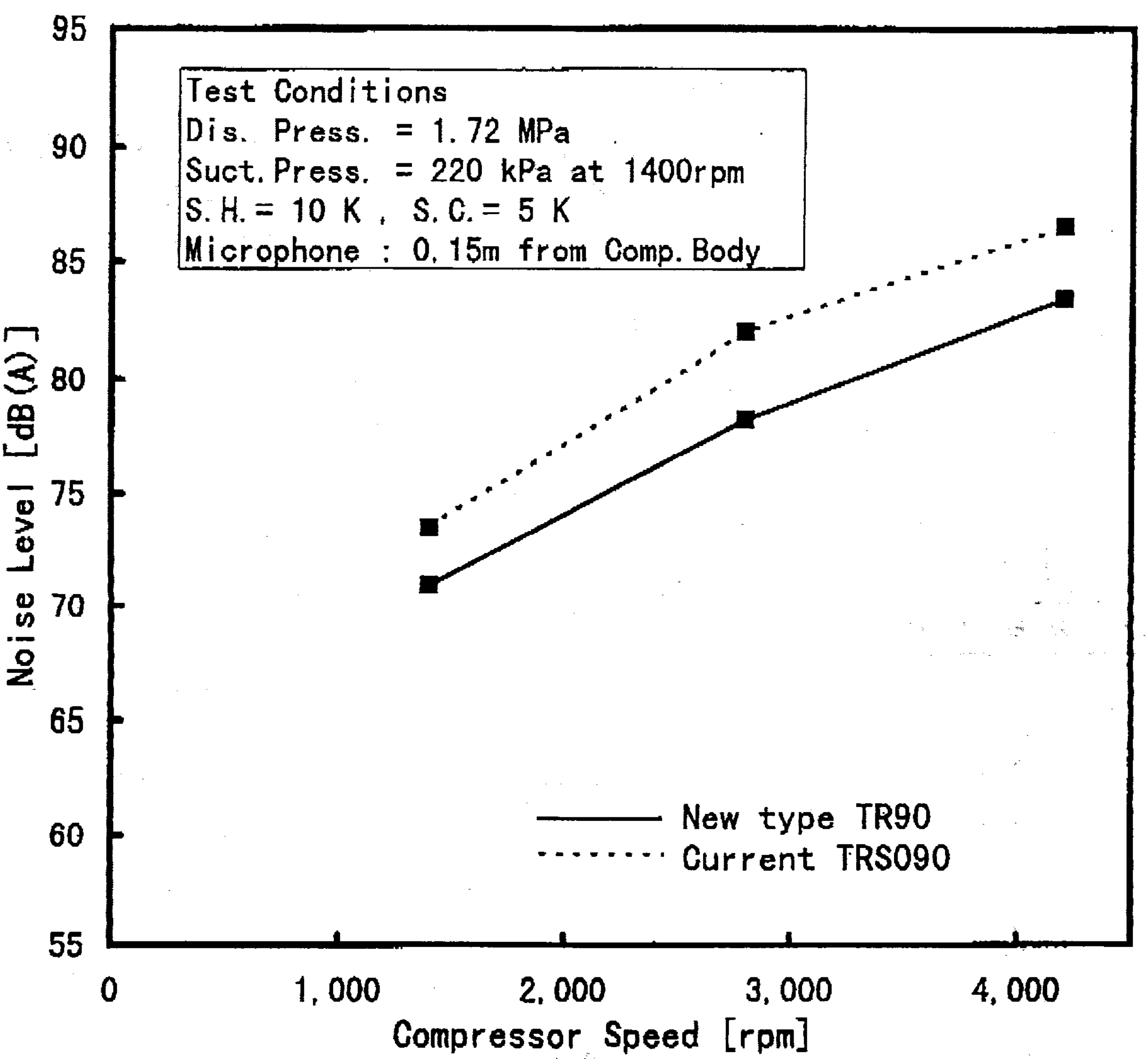


Figure 6.

CHANGE ITEMS	Current TRS090	NEW TR 90
Ball coupling	·Two parts of race and ring.	·Unified to one piece. Low noise. Light weight.

Figure 7.



Noise Level Comparison
between New type TR90 and Current TRS090

Figure 8.

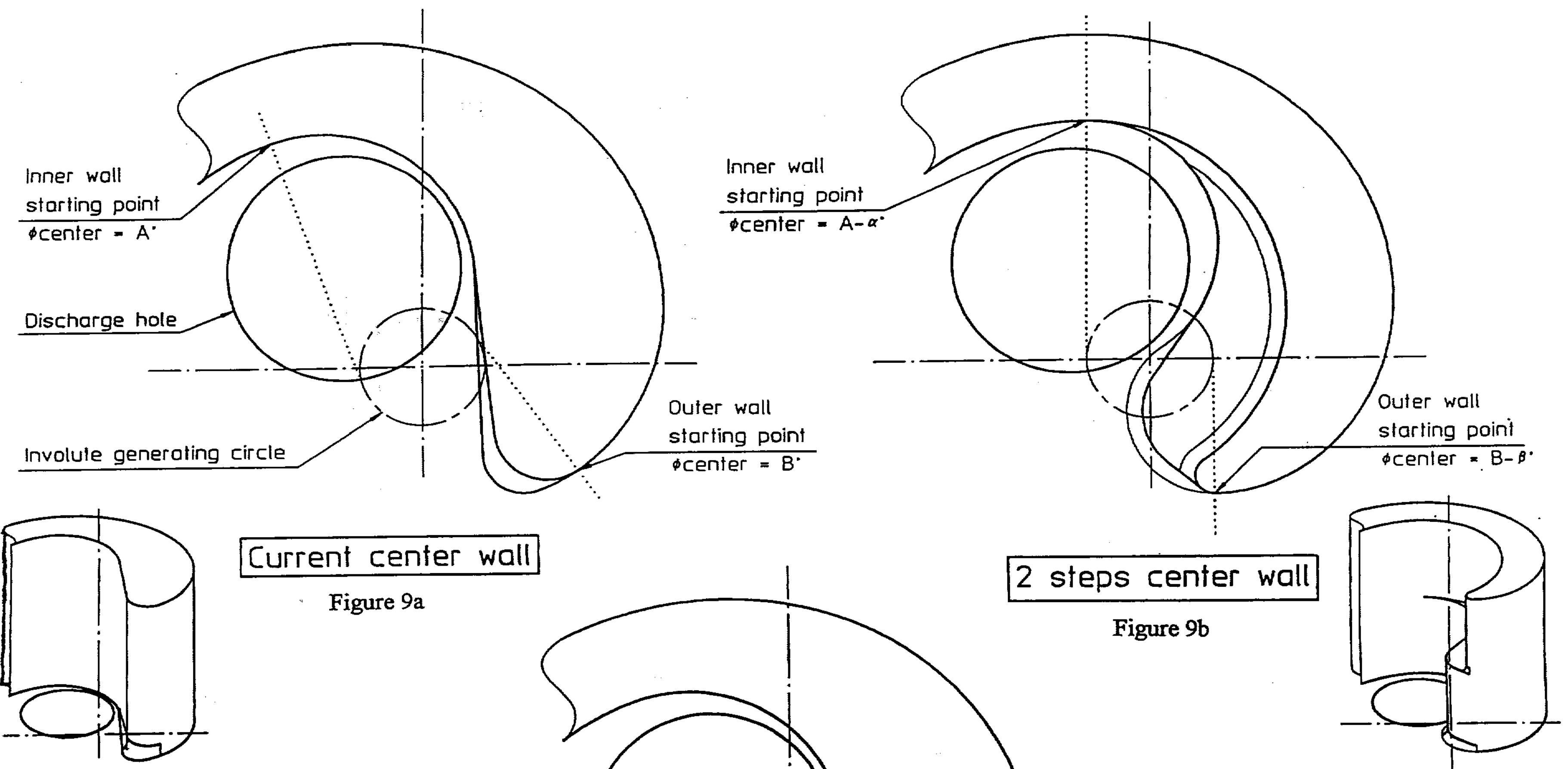


Figure 9a

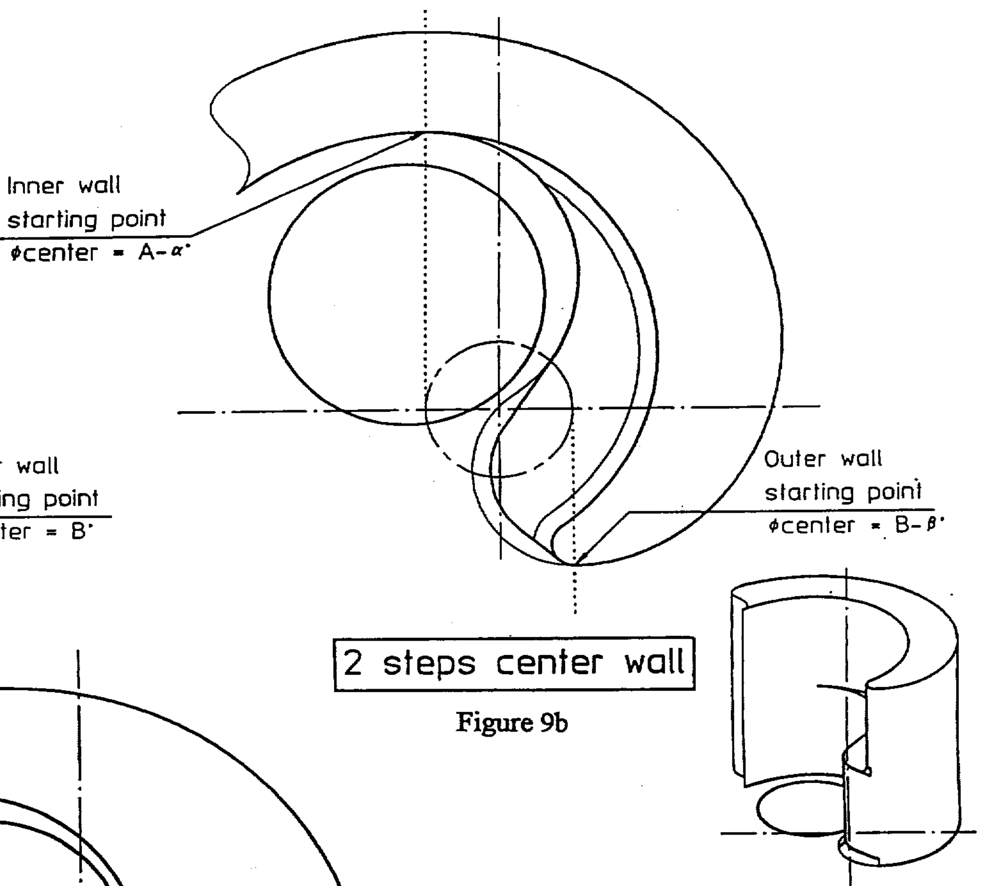
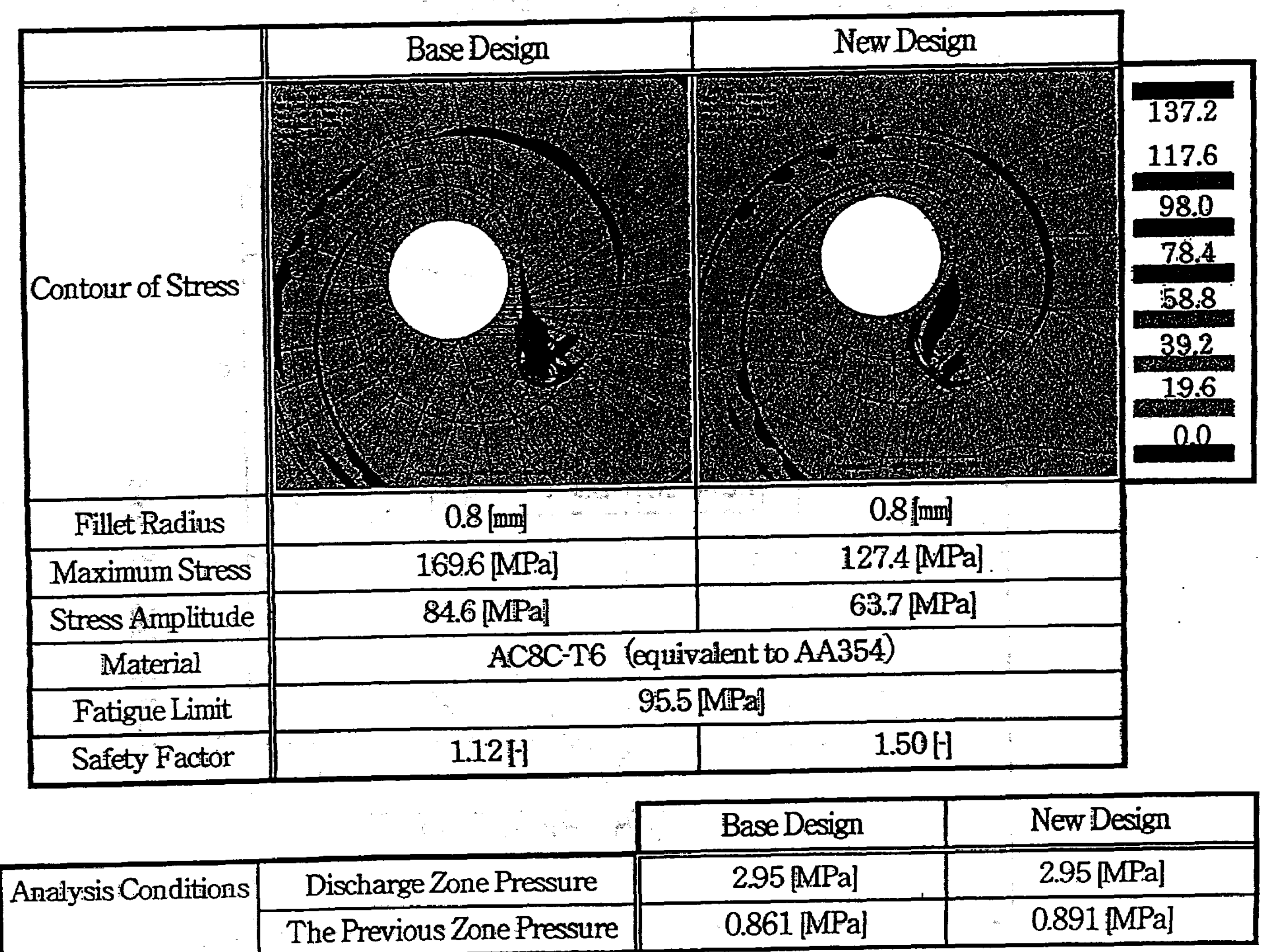


Figure 9b

Figure 9.

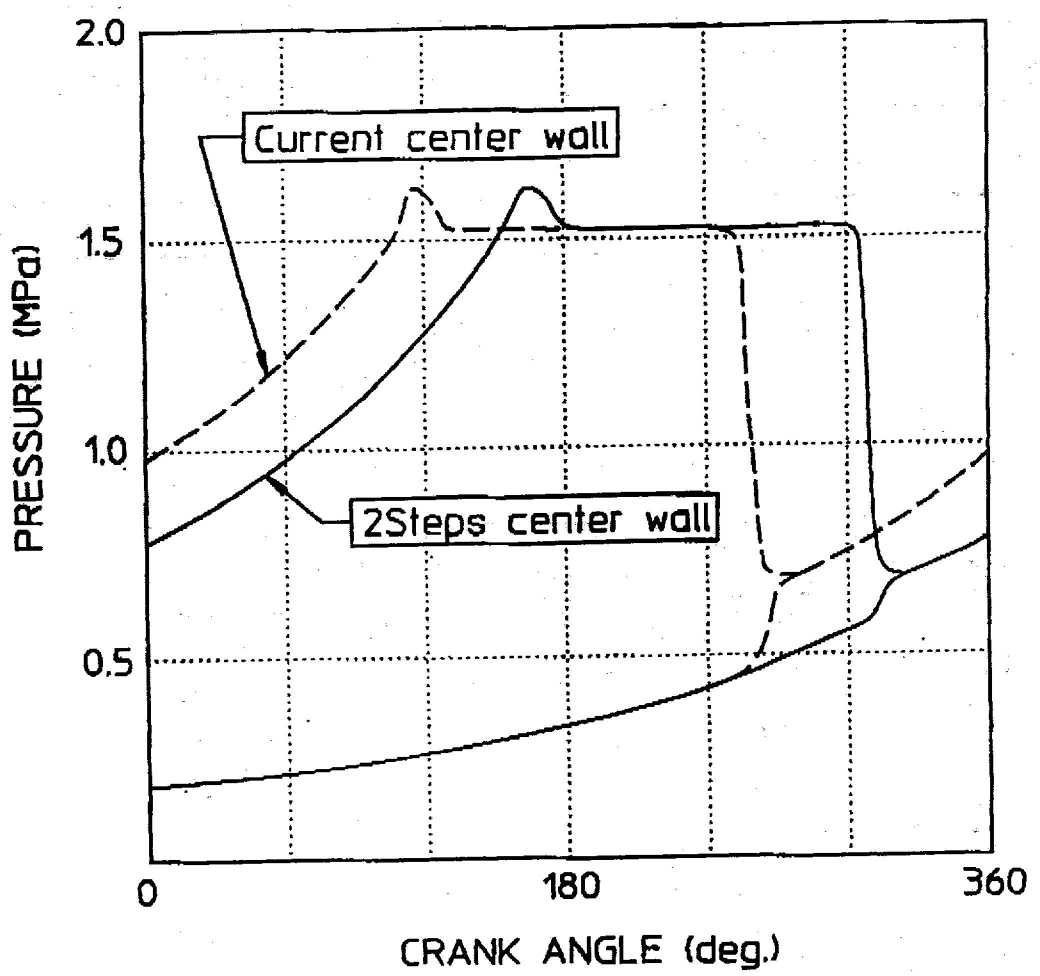
FEA Result on Scroll Center Profile



	Base Design	New Design
Contour of Stress		
Fillet Radius	0.8 [mm]	0.8 [mm]
Maximum Stress	169.6 [MPa]	127.4 [MPa]
Stress Amplitude	84.6 [MPa]	63.7 [MPa]
Material	AC8C-T6 (equivalent to AA354)	
Fatigue Limit	95.5 [MPa]	
Safety Factor	1.12 [-]	1.50 [-]

		Base Design	New Design
Analysis Conditions	Discharge Zone Pressure	2.95 [MPa]	2.95 [MPa]
	The Previous Zone Pressure	0.861 [MPa]	0.891 [MPa]

Figure 10.



Crank Angle vs Gas pressure Diagram

CHANGE ITEMS	Current TRS090	NEW TR 90
Lip seal	·With O-ring.	·Without O-ring. ·Rubber coated.

Figure 12.

Table 1. Comparison of Anti-Rotation Mechanisms

Mechanism/ Item	EM Bearing	Ball Coupling	Oldham Coupling	Pin & Ring Coupling
Noise	5	3	4	4
Mechanical Efficiency	5	5	2	2
Balancing	5	5	2	5
Productivity	4	3	3	4

NOTE: Rank 5 shows the best.

970114

In-Vehicle Adsorptive Filters Efficiency

Nathalie Lemaitre
Valeo Thermal Systems

Yves Godet
INERIS

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper presents the Valeo approach to qualifying adsorptive filters in real conditions in vehicle. The first part describes the choice of air quality indicators and the adapted methods used to measure them. The second part presents the application in vehicle and the results. The last part compares results to laboratory characterizations.

INTRODUCTION

Adsorptive filters are recently developed to improve air quality in the passenger compartment. Valeo is interested in integrating such filters in heating and air conditioning systems to provide a better air quality in vehicle for drivers and passengers. These filters contain activated carbon and non-woven material to treat gases and particles present in ambient air. They are generally called « combined filters ».

The common characterization test of adsorptive filters is based on DIN 71640 part 2, proposed for ISO standard. It can give the main features for adsorptive filters : efficiency and capacity of adsorption. This procedure is adapted to laboratory conditions and uses high concentrations of test gases. These concentrations are still much higher than those met in the atmosphere around the vehicle.

Also it is essential to know how these filters react in real conditions in vehicle. As well, it is important to compare laboratory characterization and real tests results.

In this objective, this paper will attempt to evaluate different filters in real conditions of use.

AIR QUALITY INDICATORS AND MEASUREMENT METHODS

AIR QUALITY INDICATORS - The indicators of air quality are chosen according to atmospheric pollutants found in the cabin and the ability to measure them in a vehicle.

Atmospheric pollutants - Main atmospheric pollutants are regulated because of their effect on health and are continuously monitored in ambient air by registered laboratories. These are sulfur dioxide (SO_2), nitrogen oxides (NO_2, NO), carbon monoxide (CO), ozone (O_3), volatile organic compounds (V.O.C) including hydrocarbons. The major sources are industry, heating and transportation. It is possible to measure these gases by direct or indirect methods. For in-vehicle tests, 6 indicator gases are selected:

- 3 gases used in DIN test : SO_2 - NO_2 - toluene,
- 2 gases to complete : benzene - xylene
- and 1 gas for traffic indicator : CO

MEASUREMENT METHODS - Specific methods are applied to measure the different gases.

Sulfur dioxide (SO_2) - In ambient air, SO_2 molecules are stable. Ultraviolet rays can transform them to excited state. The return to the stable state creates a fluorescent radiation emission. The fluorescence intensity is proportional to SO_2 concentration. Direct analyzers can be used in the vehicle to measure SO_2.

Nitrogen dioxide NO_2) - To be measured by chemiluminescence, NO_2 is first converted into NO by reduction in a molybdenum oven. Then chemiluminescence corresponds to the oxidation of NO molecules by O_3 molecules : $NO + O_3 \Rightarrow NO_2^* + O_2$. The return to a stable state creates a light radiation emission : $NO_2^* \Rightarrow NO_2 + h\nu$ the radiation is proportional to NO_x concentration. Without conversion, the radiation is only proportional to NO. The difference between NO_x and NO gives NO_2 concentration. Direct analyzers measuring both NOx and NO can be used in the vehicle to determine NO_2 concentration.

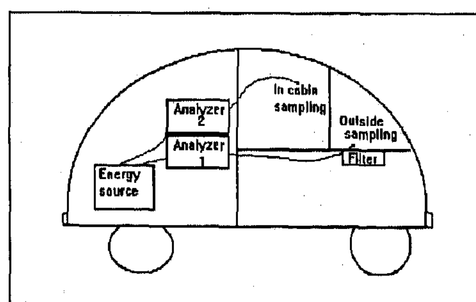
Carbon monoxide (CO) - CO molecules, absorb in infrared rays. This absorption, follows Beer-Lambert law and is proportional to CO concentration. Direct analyzers can be used in the vehicle to measure CO.

Volatile Organic Compounds (V.O.C) - To obtain specific hydrocarbon measurements, the method used is first a sampling on adsorptive tubes and then analysis in the laboratory by gas phase chromatography phase and mass spectrometry. A first sampling in the vehicle shows the main hydrocarbons found in the cabin come typically from urban pollution (3). Many compounds are identified : benzene, toluene, octane, tetrachloroethylene, ethylbenzene, p xylene, styrene, o-xylene, nonane... Majority compounds are : toluene, benzene, xylene. This indirect method can be used in vehicle.

VEHICLE APPLICATION AND RESULTS

VEHICLE APPLICATION - A vehicle is dedicated to the tests with a driver and a responsible of measurement.

Measurement on board - For each air quality indicator, two measurements are necessary to evaluate outdoor and indoor concentrations simultaneously. Each gas concentration measurement needs an analyzer or a sampler. Because of the size of the analyzers chosen, only one gas is measured during a test. However, the 3 hydrocarbons can be sampled at the same time with pumps. 2 identical analyzers are placed on the back seat and connected to the energy source. Sampling tubes are joined to analyzers (or pumps for V.O.C) : one from the outside air inlet, the other from the cabin at passengers head. For direct measurement, analyzers are linked to a portable computer which can display the results graphically.



Filters equipment - An adapted box is placed on the external air inlet of the air conditioning system on the vehicle. Filters can be easily replaced in this box during the trial. The filter surface is 5dm^2 with 35mm depth, and less than 200 grams of activated carbon to reduce the pressure drop. 3 filters were tested :

- 2 combined filters including non-woven fibers and activated carbon granules for particles and gases filtration, coming from 2 different manufacturers.
- 1 filter without activated carbon called a placebo.

The tests distribution is presented below in table 1 :

	Filter A	Filter B	Placebo
SO_2	❶	❶	❶
NO_2	❷	❷	❷
CO	❸	❸	❸
toluene	❹	❹	❹
benzene	❹	❹	❹
xylene	❹	❹	❹

❹ : concentrations with test day numbering

Trials organization - A loop of 17 kilometers was defined in Paris, mixing periphery and center town, needing 30 to 50 minutes depending on traffic intensity. The loop was repeated for each couple « indicator/filter », that is to say, 12 loops in all. The campaign is realized in 4 days for all measurements.

Relative humidity and temperature are monitored during the trials, and the air conditioning system preserves a constant temperature of 22 to 25°c in the cabin.

The airflow with the filter in place is fixed at 245m^3/h, in full cold mode.

RESULTS - Results are expressed directly in part per billion (ppb) or million (ppm), or converted in microgram per cubic meter (µg/m^3)

Loops concentrations profile - For direct measurement, it is possible to obtain the profile of concentrations during the loop.

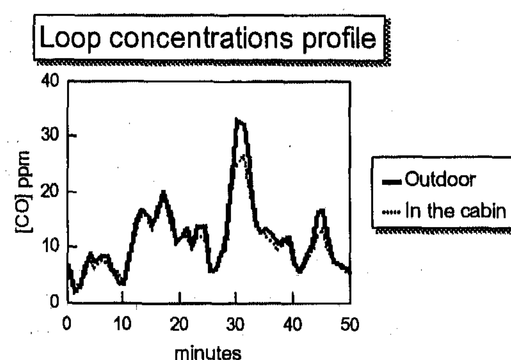


Figure 1 : example of a loop concentrations profile

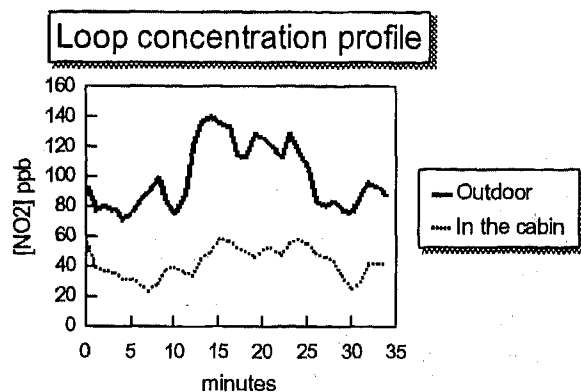


Figure 2 : example of a loop concentrations profile

The same type of profile is observed for SO_2. The concentration changes very rapidly around the vehicle according to the traffic environment. During a loop, the variation can reach a factor of 3 in 5 minutes. The concentration in the cabin follows the same variation, but at a lower level when the gas is adsorbed on a filter (figure 2). Driver and passengers are subjected to pollution peaks.

Trials concentrations profile - Results can be expressed in average of concentration for each loop.

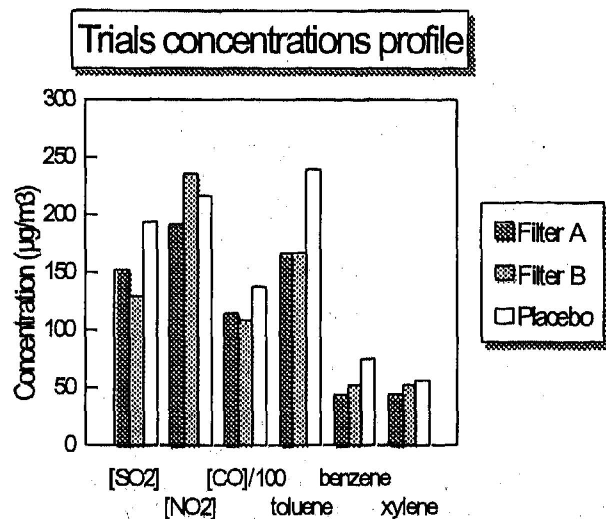


Figure 2 : Trials concentrations profile

The concentrations are quite homogeneous between the loops for a same indicator, measured the same day. The variations can only reach a factor 1,2 to 1,4.

The results can be compared to standards (1) for limit value of 1 hour exposure in ambient air. No standards are defined for 1 hour exposure for toluene, benzene and xylene.

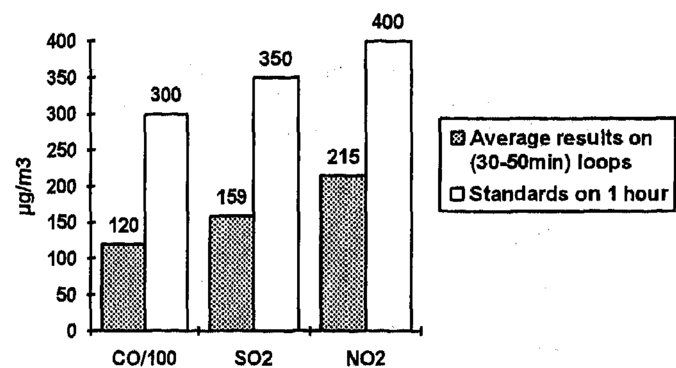


Figure 2 : Indicators concentrations

The concentrations obtained during this campaign are lower than standard values.

Air quality indicators results - The concentrations profile is presented for each indicator

• CO is not adsorbed on filter. To have an idea about the validity of the measurement, the lack of influence of this gas can be illustrated figure 4. The small difference measured between outside and in the cabin, could be attributed to the inertia time of both samples, or to the air modification in sampling tubes or air conditioning system.

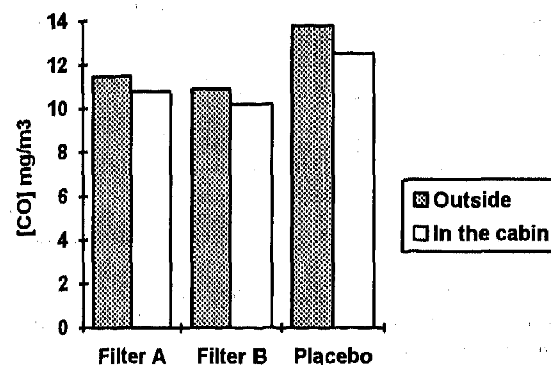


Figure 4 : CO concentrations measurement.

• SO_2 is adsorbed on filters. The high difference between outside and in the cabin for filters A and B, and the lack of difference for the placebo are shown in figure 5.

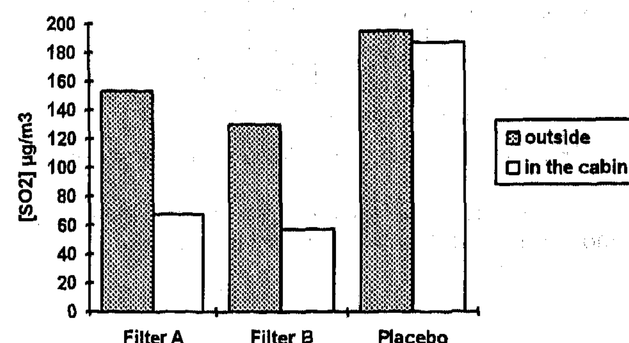


Figure 5 : SO2 concentrations measurement.

• NO_2 presents the same behaviour as SO_2, but with the placebo the highest concentration is for this case, in the cabin with the same explanation than for CO measurement.

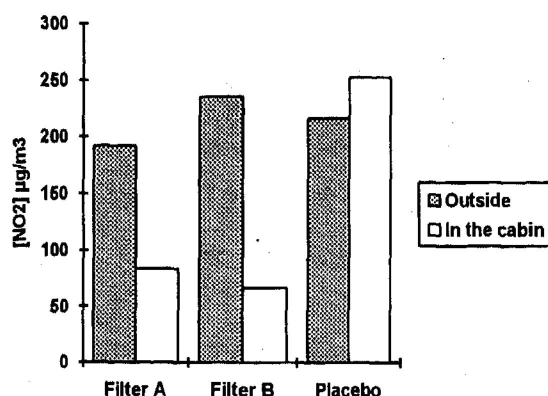


Figure 6 : NO2 concentrations measurement.

• Toluene is adsorbed on filters A and B, and not adsorbed with the placebo.

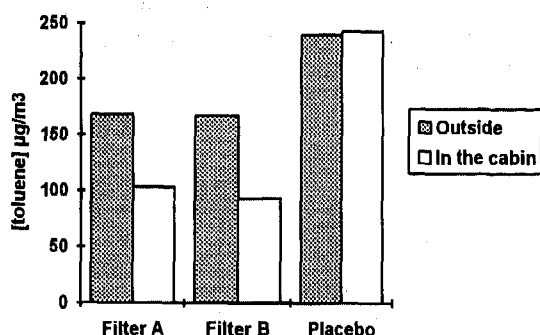


Figure 7 : Toluene concentrations measurement.

• Benzene concentrations are always lower in the cabin than outside, but more with filters A and B than with the placebo.

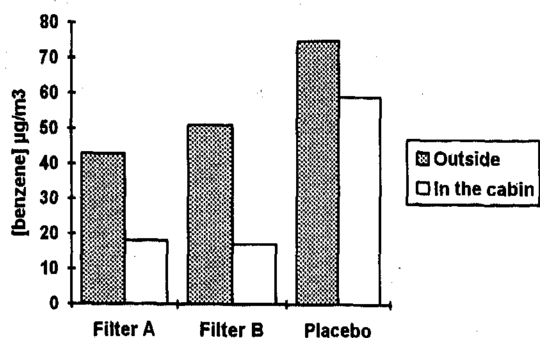


Figure 8 : Benzene concentrations measurement.

• Xylene concentrations are lower in the cabin, more with filters A and B than with the placebo.

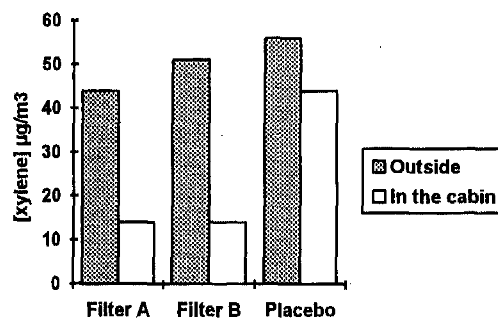


Figure 9 : Xylene concentrations measurement.

Filters efficiency - The efficiency is calculated according to the formula :

$$E\ (\%) = 1 - \frac{\text{Concentration in the cabin}}{\text{Concentration outside}}$$

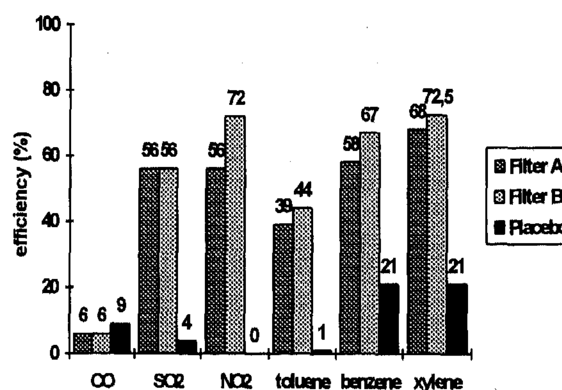


Figure 10 : Efficiency values

The efficiency of filters A and B is demonstrated in vehicle for SO_2, NO_2 and toluene. The efficiency is significant for indicators present in large concentrations. This is confirmed by the results of no efficiency obtained with the placebo. But for benzene and xylene, the level of concentration is too low to permit a precised efficiency calculation because the placebo shows 21% efficiency and this is not the true.

If a global efficiency is calculated on the 3 significant indicators, filter A is associated with a value of 48% and filter B with 57%. Filter B seems to be a little more efficient than filter A.

LABORATORY AND VEHICLE EFFICIENCY COMPARISON

TEST CONDITIONS COMPARISON - The following test parameters can be considered : temperature, relative humidity, airflow, gas concentration.

	VEHICLE	LABORATORY
Temperature	14 to 33°c	23°c
Humidity	30 to 50%	45%
Airflow	245m³/h	200m³/h
Concentration of		
toluene	80ppb	80ppm
SO_2	61ppb	30ppm
NO_2	113ppb	30ppm

The biggest difference comes from a factor 1000 between the gas concentrations. The gas feeding in laboratory is homogenous while it is more turbulent in vehicle. On top of that, the sample points are placed in a stable zone of the bench, that does not exist in vehicle.

RESULTS COMPARISON - Results of filter A and B characterization in the laboratory exist only for toluene.

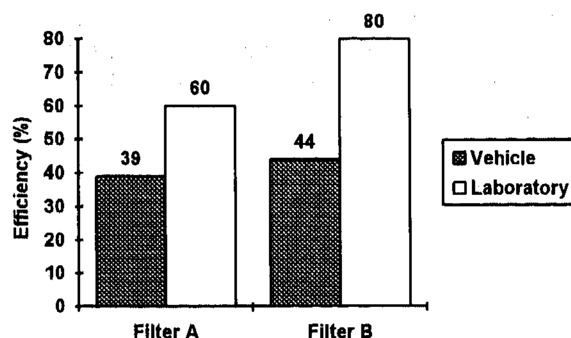


Figure 11 : Toluene effciency comparison

On this particular comparison, the laboratory characterization is more optimistic than the real evaluation. The efficiency is multiplied by 1,5 for filter A, and by 1,8 for filter B.

Obviously, conditions of tests are very different and can affect the results. Only a second campaign would confirm these first results.

CONCLUSION

This study shows that it is possible to evaluate filter efficiency in the vehicle. However, the necessary methods employed are heavy to implement and difficult to use systematically.

But it is clear that absorptive filters improve distinctly air quality in the vehicle. The tested filters reduce disturbing gases concentrations like SO_2, NO_2 and toluene by 40 to 70%. A little difference can be noticed between the two filters tested, A and B, but they are quite similar.

In this example and for toluene, the efficiency evaluation in vehicle is lower than the laboratory characterization. The conditions are very different but it would be interesting to confirm this comparison in a next campaign.

The other parameter which could be interesting to test, would be the real life time of adsorptive filters. Compare the capacity of adsorption measured in laboratory to real aging filters, could help to know the correlation between gram of test gases and filters life time. A next study could try to evaluate this feature.

REFERENCES

(1) Air quality CEE standards 1980-1985 / OMS EUR. 1987.

(2) Air quality monitoring systems in France. Example : AIRPARIF for Paris vicinity.

(3) ADEME, INRETS - Automotive pollution and health effects. February 1995.

970115

Analysis of the Magnetic Field and the Magnetic Force in a Magnetic Clutch of the Automotive Air Conditioner

Gunhee Jang and Kyungho Lee
Hanyang Univ.

Kwangil Kim, Yongkil Cha, Taeyoung Park, Kihyo Kim, and Minsoo Han
Halla Climate Control Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper investigated the characteristics of the magnetic field and the attractive magnetic force in a magnetic clutch of the automotive air conditioner. The magnetic force was calculated by using the finite element method, the principle of virtual work and the Maxwell stress tensor, and it was measured by experiment. Magnetic field is not affected by the pulley groove and the attractive magnetic force is also not significantly affected by the bearing because the total magnetic field in the airgap is not changed. The magnetic force increases rapidly as the airgap decreases. The bridges in the disk and the pulley decrease the magnetic force because they create the radial flux to saturation and reduce the effective axial magnetic flux in the airgap to the amount of the radial magnetic flux flowing in the bridges. Also the magnetic force is decreased proportionally as the increase of the cross-sectional area of bridge.

INTRODUCTION

The magnetic clutch is widely used as a device to transmit the power of an engine to a compressor to drive the automotive air conditioner. Figure 1 shows the magnetic clutch with four poles. The current in the coil produces the closed magnetic flux path along the pulley, the airgap and the clutch disk, which results in the attractive magnetic force in the air gap between the rotary pulley and the stationary disk in order to engage the clutch disk to the rotary pulley. The excessive magnetic force results in the high impact force when the disk is engaged, and too little force does not allow the engagement of the disk.

This paper investigated the magnetic field and the attractive magnetic force, and the performance of magnetic clutch due to the variation of design parameters of the magnetic clutch. A finite element method (FEM) was used to analyze the magnetic field in the magnetic clutch. Magnetic nonlinearity was included by using the BH curves of the disk and the pulley measured by the Hysteresisgraph system. The attractive magnetic force was calculated by the principle of virtual work or the Maxwell stress tensor, as needed. Also an experimental setup was built and the attractive magnetic force was measured in the universal tension machine to verify the result of FEM. Finally, the design parameters of the magnetic clutch were thoroughly investigated in order to see how they affect the magnetic field and the magnetic force.

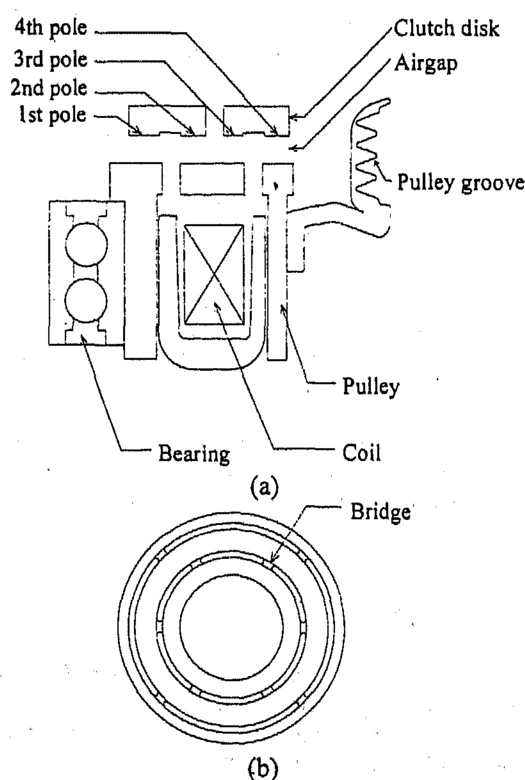


Figure 1: (a) Cross-section of the magnetic clutch with four poles
(b) Top view of the pulley

METHOD OF ANALYSIS

MAGNETIC FIELD AND MAGNETIC FORCE - The magnetic field generated by a magnetic clutch is described by the set of Maxwell's equations. Introducing the scalar potential into the Maxwell's equations, with some mathematics, gives a single partial differential equation for the scalar potential Φ.

$$\nabla^{T}\cdot\mu\nabla\Phi - \nabla^{T}\cdot\mu\mathbf{H}_s = 0 \qquad (1)$$

In this equation, μ is the permeability and $\mathbf{H}_s$, the field intensity due to the current in the coil may always be calculated directly by the Biot-Savart law.

$$\mathbf{H}_s = \frac{1}{4\pi}\int_{\Omega}\frac{J \times R}{|R|^3}d\Omega \qquad (2)$$

The governing equation, like the Poisson's equation for electrostatic fields, can be solved using the finite element method. ANSYS, a FEM solver for magnetic field, was used to calculate the scalar potential Φ and the magnetic field intensity, H.

A non-uniform distributed force per unit area at the interface between two materials is calculated by introducing the Maxwell stress tensor. Since the strain imposed on the material due to magnetostriction is small enough to neglect changes in the density, it can be assumed that the change of permeability is negligible. Thus, in tensor notation,

$$\sigma_{ij} = \frac{1}{\mu}\left(B_i B_j - \frac{1}{2}\delta_{ij}B_k B_k\right) \qquad (3)$$

where σ_{ij} is the Maxwell stress tensor. B_i is the magnetic flux density which is obtained by the multiplication of permeability to the magnetic field intensity. From the expression given by Woodson and Melcher(1985) for an interface between two materials a and b, the traction, f_i, is given by :

$$f_i = (\sigma_{ij}^{a} - \sigma_{ij}^{b})n_j \qquad (4)$$

The normal and the tangential traction can be decomposed.

$$f_n = (\sigma_{ij}^{a} - \sigma_{ij}^{b})n_i n_j \qquad (5)$$

$$f_t = \sqrt{|f_i|^2 - |f_n|^2} = |\mathbf{n} \times \mathbf{f} \times \mathbf{n}| \qquad (6)$$

Since $\mu_{air} << \mu_{iron}$, the magnetic traction can be simplified by the introduction of the cylindrical coordinate. Along the air gap, the normal and the tangential traction are as follows:

$$f_z \approx \sigma_{zz}^{a} = \frac{1}{2\mu_{air}}(B_z^2 - B_x^2 - B_y^2) \qquad (7)$$

$$f_x = f_y \approx \sigma_{xy}^{a} = \frac{1}{\mu_{air}}B_x B_y \qquad (8)$$

The magnetic force in the axial direction produced for one position can easily be derived from the integration of the normal force along the small air gap.

$$F_z = \int_S f_z \, dS \qquad (9)$$

The concept of virtual work may also be used to obtain the magnetic force. The stored energy in a magnetic field is as follows.

$$W = \int_{\Omega} B\cdot H \, d\Omega \qquad (10)$$

If the disk is moved by an infinitesimal distance δz, the work done is the product of the component of force in that direction F_z and δz. This work must come from the stored energy W of the system with the fact that the field distribution inside a closed surface in air remain unchanged if the external sources are not removed and replaced by currents and poles on the surface:

$$F_z = -\frac{\delta W}{\delta z} \qquad (11)$$

The magnetic force calculated by the virtual work has higher accuracy than that by the Maxwell stress tensor, because the energy computation is of an order of accuracy higher than the potential. However, the virtual work method does not produce the distribution of magnetic force.

FINITE ELEMENT MODEL - The magnetic clutch without bridges has the rotationally symmetric geometry with respect to z-axis, i.e. axisymmetry. Axisymmetric 2-dimensional FEM model was constructed as shown in Figure 2. 4 nodes quadrilateral and 3 nodes triangular elements are used, and it has fine mesh in the airgap. The total number of node and element are 12,149 and 12,107 respectively.

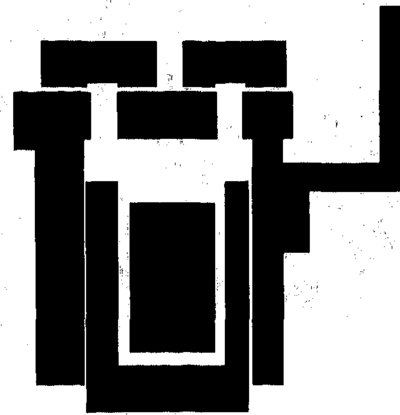


Figure 2: Axisymmetric 2-dimensional FEM Model

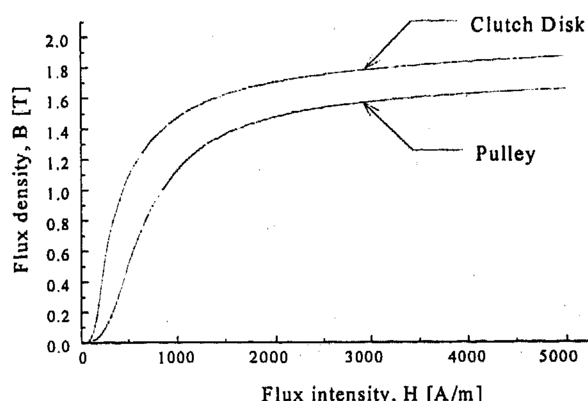


Figure 3: BH curves of the clutch disk and the pulley

The applied current of coil is assumed to be 2.9 A x 273 turns, and Figure 3 shows the BH curves of disk and pulley measured by the Hysteresisgraph system. ANSYS, a FEM solver was used to analyze the magnetic field of the magnetic clutch. Because it is relatively easier to construct than 3-dimensional model, this model was first investigated to see the role of each component and then 3-dimensional model was constructed without including the pulley groove and the bearing which do not affect the magnetic field in the airgap.

Because the magnetic flux flows within every quarter when the bridges of the disk face the bridge of the pulley exactly, the magnetic field of the whole magnetic clutch can be obtained by analyzing only one-quarter of the magnetic clutch as shown in Figure 4, with Neumman boundary condition at each cross-section. It was built with 8 nodes brick element and the total number of elements and nodes are 12,976 and 49,101 respectively. Again, ANSYS was used to analyze the magnetic field with the generalized scalar potential method. It takes about 4500 CPU sec with Cray YMP 4E for one position to solve the nonlinear magnetic field and to calculate the magnetic force by the principle of virtual work.

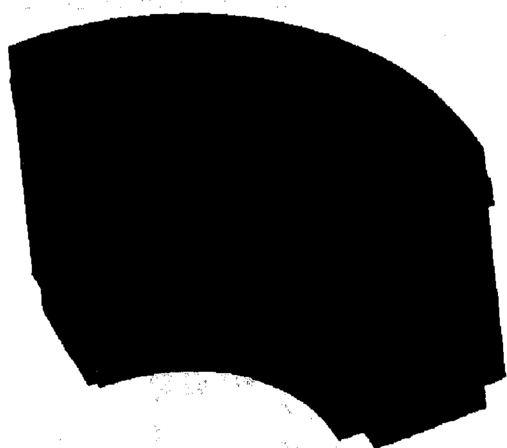


Figure 4: 3-dimensional FEM model of the magnetic clutch

EXPERIMENTAL SETUP - Figure 5 shows the experimental jig to measure the magnetic force of the magnetic clutch with the universal tension machine (UTM). The jig was made by aluminum to prevent the magnetic flux leakage. The pulley and the disk are attached to the jig and the aluminum sheets (0.3-0.7mm) are inserted between the disk and the pulley to measure the magnetic force for the different thickness of the airgap existing between the disk and the pulley, because the permeability of aluminum is almost same as that of air.

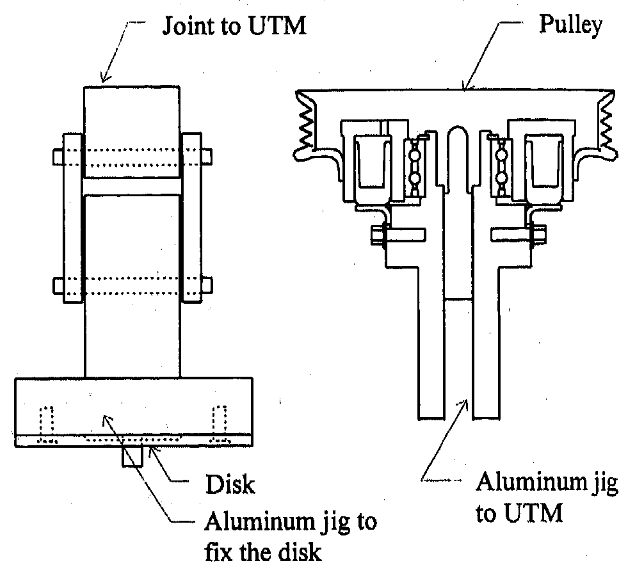


Figure 5: Experimental setup

RESULTS AND DISCUSSION

The geometry of magnetic clutch is so complicated that it must be simplified for the construction of a FEM model. The effect of pulley groove was investigated by analyzing the axisymmetric 2-dimensional FEM models for a case with a pulley groove and another case without it, when it has 0.6mm airgap. Both cases resulted in the same magnetic field in the airgap and the same magnetic force of 246.3 N, because the pulley groove plays the role of a magnetic open circuit which does not affect the total magnetic field along the airgap.

Figure 6 shows the variation of the magnetic force due to the existence of the bearing. The magnetic force was measured in the experimental setup as shown in Figure 5. The magnetic force with the bearing is almost the same as the force without it, because the total magnetic flux in the airgap is not significantly affected even though there may be minor change of the magnetic flux near the bearing. Based on the above observation, the pulley groove and the bearing were eliminated in the FEM model.

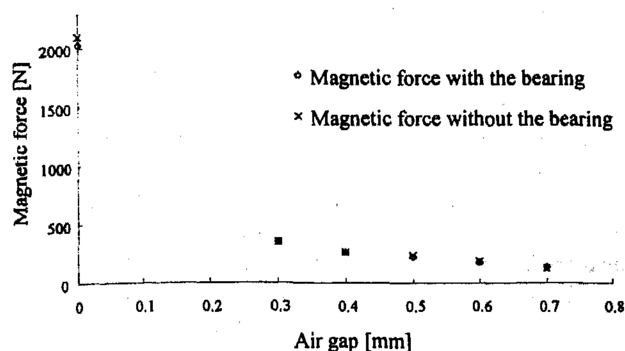


Figure 6: The variation of the magnetic forces due to the existence of the bearing

Figure 7 shows the calculated magnetic force and the measured magnetic force due to the different airgap thickness. The magnetic force is increased rapidly as the airgap is decreased. The discrepancy between the simulation and the experiment becomes larger as the airgap is decreased. The discrepancy between the axisymmetric 2-dimensional simulation and the experiment is relatively larger. But 3-dimensional model fairly decreases this discrepancy. The main difference between the axisymmetric 2-dimensional and 3-dimensional model is the bridges which connect poles in the pulley and the clutch disk as shown in Figure 1. The remaining discrepancy would be from the factors which are not included in 3-dimensional FEM model such as the rivet holes in the disk bridges, convexity of disk and surface roughness.

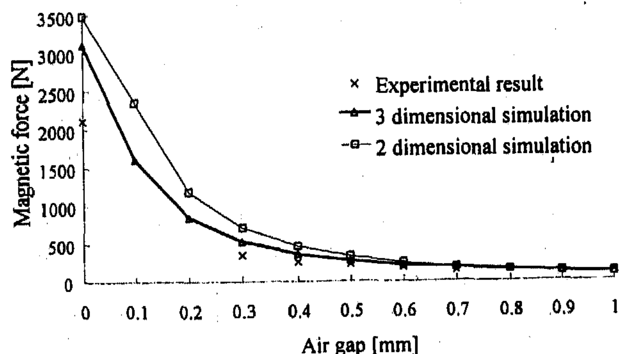


Figure 7: The calculated and the measured magnetic force

Figure 8 shows the contour plot of magnetic flux density in the pulley and the disk (0.6 mm airgap). High magnetic saturation is observed in the bridges of the pulley and the disk. Magnetic flux easily flows through the bridges not across the airgap because the permeability of iron is several hundred times greater than that of air. This results in the reduction of magnetic force because the bridges reduce the effective axial magnetic flux.

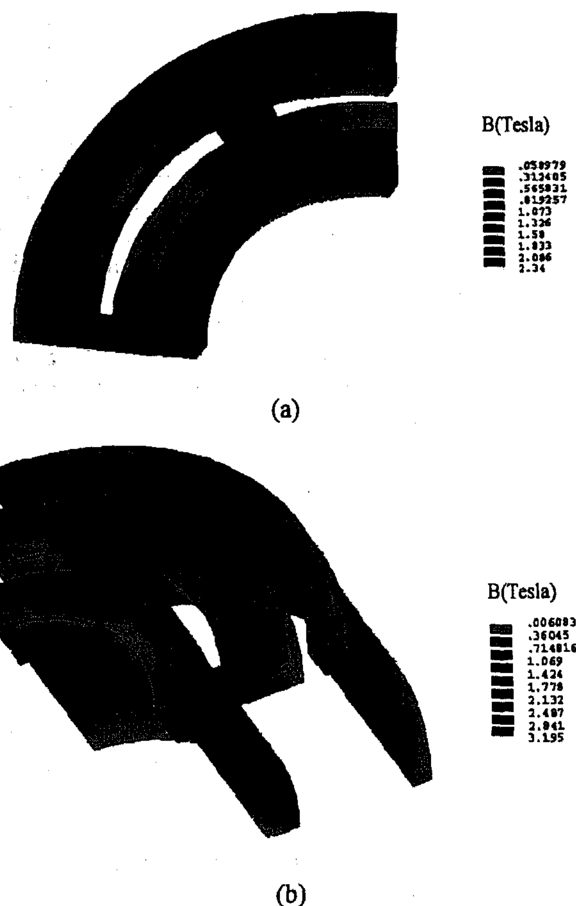


Figure 8: (a) The contour plot of the magnetic flux density in the disk
(b) The contour plot of the magnetic flux density in the pulley

Figure 9 shows the variation of the magnetic flux density in the bridges as the variation of the airgap. Even though the magnetic flux in the bridge is slightly decreased as the airgap is decreased, there are still high saturation of magnetic flux due to the large permeability of iron.

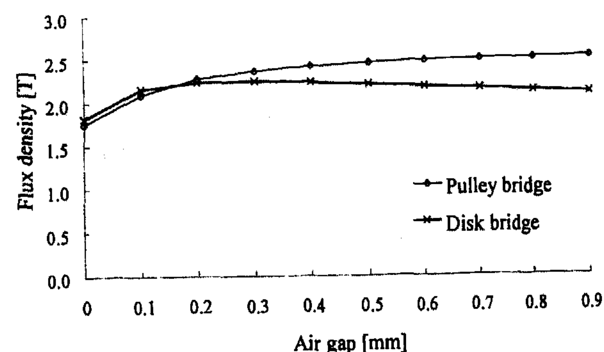


Figure 9: The variation of the magnetic flux density in the bridges

Figure 10 shows the variation of the magnetic force due to the variation of the cross-sectional area of the bridge. The magnitudes of the magnetic force are decreased as the cross-sectional area of the bridge is increased in the order of no bridge, bridge with 0.25 mm undercut, and bridge without undercut. The undercut in the bridge is defined as the axial removal of bridge from the disk and pulley surface. 0.25 mm undercut of the bridge reduces the effective cross-sectional area by 5 % and reduces the magnetic force by 8-17%. In the bridge, the magnetic flux flows radially until the bridge is saturated. The effective axial magnetic flux across the airgap is decreased to the amount of the radial magnetic flux flowing in the bridge, so that the attractive magnetic force is decreased.

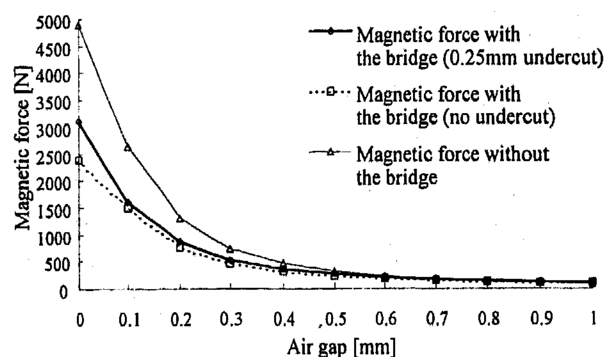


Figure 10: The variation of the magnetic force due to the area of bridge

The pulley of the magnetic clutch has six bridges inside (between the 1st and the 2nd pole) and four bridges outside (between the 3rd and the 4th pole). The effect of the location of bridges is investigated by analyzing another design of pulley with four bridges inside and six bridges outside. Figure 11 shows the magnetic force for both models. They produce almost the same magnetic force irrespective of the location of bridges, because they have the same effective cross-sectional area of bridge.

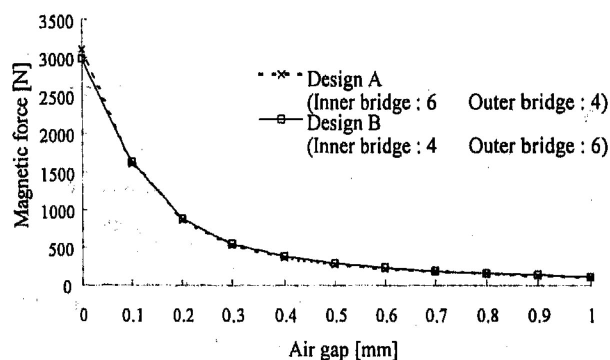


Figure 11: The variation of the magnetic force due to the location of bridges

CONCLUDING REMARKS

A method of analysis of the magnetic field and the magnetic force in the magnetic clutch has been presented from FEM, Maxwell stress tensor, the principle of virtual work and some experiments. Magnetic field is not affected by the pulley groove and the attractive magnetic force is also not significantly affected by the bearing because the total magnetic field in the airgap is not changed. The magnetic force is increased rapidly as the decrease of the airgap. The bridges in the disk and the pulley decrease the magnetic force because they are easily saturated and reduce the effective magnetic flux in the axial direction to the amount of the radial magnetic flux flowing in the bridges. Also the magnetic force is decreased proportionally as the increase of the cross-sectional area of bridge.

While the magnetic force allows the clutch disk to be engaged with the rotary pulley, the engagement occurs in a very short time, followed by impact with enormous impact force, and slipping between the disk and the pulley before firm engagement. Because the operation of the magnetic clutch is determined by the dynamic phenomena of the impact and the slipping as well as the characteristics of the attractive magnetic force, they should be thoroughly investigated in order to produce the magnetic clutch of high quality.

REFERENCES

1. Gyimesi, Miklos and Lavers, J. Douglas, "Generalized Potential Formulation for 3-D Magnetostatic Problems", IEEE Transactions on Magnetics, Vol. 28, No. 4, pp.1924-1929., 1992
2. Mayergoyz, I.D., Chari, M.V.K. and D'angelo, J., "A New Scalar Potential Formulation for Three-Dimensional Magnetostatic Problems", IEEE Transactions on Magnetics, Vol. MAG-23, No. 6, pp. 3889-3894, 1987
3. Jang, G. H., "Analysis of Magnetic Force and Torque in BLDC Motor", Ph.D. Thesis, University of California, Berkerly, 1993
4. Rizzoni, Giorgio, Principles and Applications of Electrical Engineering, Richard D. Irwin Inc.,1993
5. Woodson, H. and Melcher, J. R., Electromechanical Dynamics, Part I, II, III, The Robert E. Krieger Publishing Company Inc., New York, 1985
6. Hoole, S. Ratnajeevan, Computer-Aided Analysis and Design of Electromagnetic devices, Elsevier Science Publishing Co. Inc., 1989
7. Krause, Paul C., Wasynczuk, Oleg Electromechanical Motion Devices, McGraw-Hill Inc., 1989
8. Paul, C.R., Nasar, S. A., Unnewehr, L. E., Introduction to Electrical Engineering, 2nd Edition McGraw-Hill. Inc., 1992

9. Kraus, John D., Electromagnetics, 4th Edition, McGraw-Hill Inc., 1992

ABOUT THE MAIN AUTHOR

Gunhee Jang was born in Seoul, Korea, in 1961. He received the Ph. D. in mechanical engineering in 1993 form the University of California, Berkeley, USA. After he worked for Quantum Corporation, San Jose, USA, he has been with the Department of Precision Mechanical Engineering of Hanyang University since 1994. His major field of interest includes the analysis and the design of electromechanical device, high speed rotating machine and magnetically induced vibration. His address is as follows : Department of Precision Mechanical Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-ku, Seoul, 133-791, Korea, Email : ghjang@email.hanyang.ac.kr, Tel : 82-2-290-0431, Fax : 82-2-292-3786

970116

Simulation and Modeling of an A/C Rotary Vane Compressor

S. Takeshita
Zexel USA

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A simulation model and code suitable for rotary vane compressor are developed. The simulation features are the compression algorithm that makes extensive use of entropy tracking and the real R134a properties in the form of look-up table rather than a polytropic equation. Compression process is modeled as an isentropic process with mass flow into and out of the compression chamber due to leaks between compression chambers and flow through the valves. It was found, during the course of algorithm development, that valve back flow has a significant influence. The simulation output including torque, torque variance, heat capacity, coefficient of performance compare quite favorably with experimental measurements. To demonstrate how the simulation may provide useful insight for design and parts optimization, two simulation examples, one dealing with the effect of varying aspect ratio and the other with cylinder profile characteristics, are presented.

1. INTRODUCTION

It has become common practice to utilize computer simulation of a new product in order to reduce the development time. The simulation model of the product should contain the essential features that are important to the designer and often requires the incorporation of highly nonlinear behavior in order to be useful. This code is written to provide a simulation realistic enough to be a useful design tool, yet is easy and "light" enough even for a lap-top.

In previous simulations, particularly the ones with simplified algorithms, the compression process is described by a polytropic relationship between volume and pressure with a single polytropic coefficient [1]. The polytropic relation is derived from the treatment of working fluid as an ideal gas and the compression process as completely isentropic and leak-free compression. In other words, the property of the working fluid and the description of the compression process are coupled to yield the polytropic description.

Other simulation models are based on energy conservation, and require input of empirical parameters such efficiency. It is quite justifiable since the focus of the simulation is not the compression per se, but related phenomena such as heat transfer between working fluid, lubricant and components [2], and acoustic interaction between chambers or valve dynamics [3].

In this simulation, the compression process is treated as a combination of isentropic compression of the gas inside of the compression chamber and flow of gas into and out of the chamber with irreversible mixing. The thermodynamic property of the working fluid is given by the use of look-up table. Unlike the polytropic description of the process, the behavior of gas and compression process are encapsulated rather than interwound. The description of the compression becomes canonical through tracking of entropy and the addition of mass flow through gaps and valves.

Some of the parameters that are necessary as simulation input were not obtained, thus the simulation code is used to determine them by best-fitting a number of experimentally obtained pressure profiles. The code is then verified by comparison of output with measurement data other than those used in computing the parameters. Examples of how the simulation can be used in predicting the influence of design parameter are introduced to provide some insights into cylinder design.

2. DESCRIPTION OF PHYSICAL MODELING

The simulation code is written for rotary vane compressor with five vanes and cylinder with two fold point symmetry, namely Zexel KC_XX type compressors. The compressor is of variable capacity type. A brief cross-section diagram is shown in Figure 1.

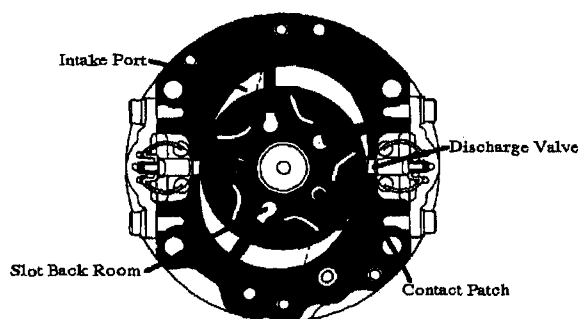


Figure 1. Cross-section of Zexel KC_XX Compressor

2.1. WORKING FLUID

A look-up table for R134a thermodynamic properties was constructed in an encapsulated subroutine form. It accepts any two of the seven state functions (T, P, x, v, u, h and s) and computes the other by interpolation and/or reverse interpolation. It also computes dependent properties such as heat capacities, acoustic speed, and viscosity. It provides encapsulated description of refrigerant behavior over a very wide region, including compressed liquid and saturated mixture. The table can be used in simulations of heat exchangers and of the entire system. The values of thermodynamic functions were taken out of Makita, et. al. [4]

It has been customary to model the refrigerant gas as a polyatomic ideal gas, that is, it obeys the ideal gas law, but its heat capacity is a function of temperature. In many works, this model is further simplified by assuming a constant ratio of specific heats, typically around 1.1, which is much more realistic than the specific heat ratio for monatomic ideal gas of 5/3. Nevertheless, in such scheme, the property of refrigerant and the characteristics of the compression path are intertwined.

The compression process is modeled as an approximately isentropic process, more precisely, an isentropic process with corrections due to gas flow in and out of compression chamber. The isolation of the gas property and compression process itself is a desirable feature to properly incorporate the corrections.

A good measure of the validity of the ideal gas approximation is the compressibility. If it is close to unity throughout the compression path, the ideal gas is a good approximation. In our preliminary investigation, we found out that this may not always be the case. In Temperature-Entropy diagram, a typical compression process shows itself to be very near the saturated gas line. While the compressibility stays close to unity when the discharge pressure is relatively low, the compressibility starts deviating from unity as the pressure becomes higher. The designed compression path may be in approximately ideal region, but the simulation must also be able to analyze operating conditions that are out of designed scope.

2.2. COMPRESSION

The compression process is modeled in two parts. Consider the compression process between time t_0 and $t_1=t_0+\Delta t$. The compression of the gas that is in the chamber at time t_0 is modeled by isentropic compression. The other part is the leak into and out of the chamber that takes place between time t_0 and t_1. The leak and flow through valve are described in the next section. For now, it is a certain amount of gas in a certain state that enters the chamber. The heat transfer, including that due to friction, is not yet accounted for this version of simulation.

The state of chamber at t_1 is computed from the state at t_0. The mass in the chamber increases by the amount of the net flow into the chamber.

$$m_1=m_0+\Delta m_{in}-\Delta m_{out} \qquad (1)$$

where $\Delta m_{in}=\Delta m_{in}$ (origin state, chamber state; flow geometry)
$\Delta m_{out}=\Delta m_{out}$ (destination state, chamber state; flow geometry)

The internal energy in the chamber can be approached in two ways. The first is the use of the first law of thermodynamics exclusively, and the second is the incorporation of the second law for the compression process.

The first law treatment is simple energy balance.

$$U_{1P}=U_0 + \Delta Q - \Delta W + \Delta m_{in}h_{in} - \Delta m_{out}h_{out} \qquad (2)$$

where U_{1P} is the total internal energy in the chamber at t_1,
$= m_1u_1$,
U_0 at t_0, $= m_0u_0$,
ΔQ is the heat flow,
$=0$ for this analysis,
ΔW is the work done,
$=$ P(chamber state)ΔV,
h_{ini} is enthalpy of the incoming gas,
$= h$ (origin state),
h_{out} is enthalpy of the outgoing gas,
$= h$ (chamber state)

This model is used in predictor part of the calculation.

The other treatment arises from the model of compression part per se being isentropic. The work to compress the gas comes mostly from the change in the chamber volume but also, in small part, comes from the gas that is entering the chamber through the leak. This formulation allows the removal of work variables.

$U_{1C}=U_{0C} + \Delta Q + \Delta m_{in}u_{in} - \Delta m_{out}u_{out}$ (3)

where U_{1C} is the total internal energy in the chamber at t_1,
$= m_1u_1$,
U_{0C} is the energy of the isentropically compressed gas,
$= m_0u(P_1, s_0)$,
$\Delta Q = 0$, as before,
u_{in} is internal energy of the incoming gas,
$= u$(origin state),
u_{out} is internal energy of the outgoing gas,
$= u$(chamber state)

This model is used in corrector calculation.

The derivation of corrector equation (3) arises from the modeling of compression process to be isentropic to gas present in the chamber from t0 to t1. The work done onto the gas present in chamber comes from the work of the chamber and the net of the work done by the gas incoming and outgoing.

$$\Delta W_{chamber} = P\Delta V = \Delta W(\text{gas compression}) - \Delta m_{in}P_{in}v_{in} + \Delta m_{out}P_{out}v_{out} \quad (4)$$

Noting that, for the isentropic compression, the energy relationship for the gas present in the chamber is

$$U_{1C}=U_{0C} + \Delta W(\text{gas compression}) \quad (5)$$

Combining equations (2), (4) and (5) yields equation (3) representing the second scheme. The second method automatically incorporates assumptions (4) and (5). As stated earlier, the compression is modeled as fundamentally an isentropic compression with flows from leaks and valves. The second scheme follows more closely this canonical description of the process.

2.3. LEAK

For rotary vane and scroll type compressors, the leakage plays a major role in determining the compressor thermodynamic characteristics. In modeling this characteristics, leakage through the contact patch into the suction plenum and sides of the vane into the following chamber are considered, as they are the dominant leak mechanism.

The model employed is Fanno flow with possible (and probable) choking at the exit of the gap. The model is further explained with Figure 2, which shows various ranges of Fanno flow. Typical Fanno flow without choking starts out with isentropic acceleration from the initial stagnation state. The gas then enters the gap, which provides friction resulting in the increase of entropy as well as drop of pressure. The gas isentropically decelerates to stagnation after exiting from the gap. This process is shown in the Figure as path 00->0->1->10.

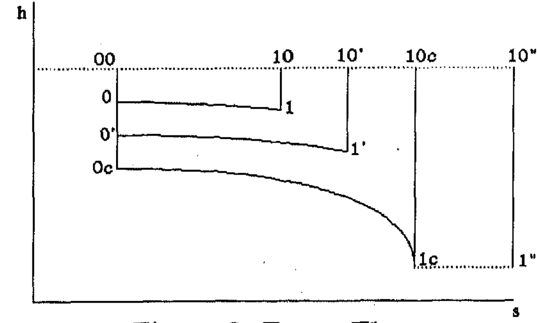


Figure 2. Fanno Flow

When the pressure difference is increased, a similar path is taken, but for higher flow density, it is shown in the figure as 00->0'->1'->10'. If the pressure difference is further increased, the flow will become critical at the exit of the gap This path is shown as $00\text{->}0_C\text{->}1_C\text{->}10_C$. The exit state 1_C is the point of maximum entropy on Fanno curve. If the pressure difference further increases, the flow up to the gap exit will remain the same and choking, which is the abrupt increase in entropy associated with abrupt drop in pressure, at the exit of the gap will occur. This is shown as $00\text{>}0_C\text{->}1_C\text{->}1''\text{->}10''$. For the most of flow across the contact patch of vane sides, the pressure difference is high enough for choking to occur. Thus, the flow density of the critical flow will be the flow quantity of the gap. The following relationship exists between the exit stagnation state and the flow density G, because the sum of enthalpy and kinetic energy is conserved.

$$h_{10C} = h_1 + \frac{1}{2}c_{s1}^2 = h_1 + \frac{1}{2}G^2v_1^2 \quad (6)$$

where c_{s1} is the acoustic speed at state 1,
v_1 is the specific volume at state 1.

An iteration scheme to obtain G from s_1 or vice versa is easily constructed from Equation (6).

While the energy conservation of Fanno flow provides a relationship between the flow density and entropy increment, we need yet another relationship to solve this problem with two unknown - the flow density and one of the state variables for the exit state. The other relationship is provided by examining how entropy increases by generation of lost work due to the friction with the gap walls. The entropy increase of the gas is associated with the generation of lost work due to the friction between the wall and the flowing gas. The lost work over the Δx segment of the gap is computed from the shear stress at the gap walls, as shown in Figure 3.

$$fdx=w_{Gap}(\tau(x, y=\frac{h_{Gap}}{2}) - \tau(x, y=-\frac{h_{Gap}}{2}))\, dx \quad (7)$$

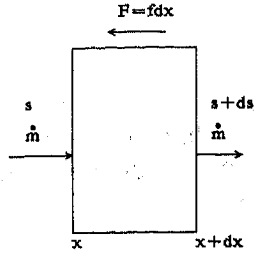


Figure 3. Entropy Increase in a Control Volume

where fdx is the frictional force generated within the differential segment
w_{Gap} is the width of the gap
τ is the shear stress, = μ ($\partial u/\partial y$)
μ is the viscosity

The lost work generation rate due to this friction force is

$$d\dot{L}W = fdx\ Vel = fdx\ vG \quad (8)$$

where $\dot{L}W$ is the total lost work generation rate,
$d\dot{L}W$ is that due to the differential segment,
Vel is the average velocity of the flow

Since the lost work generated contributes to increment of entropy equal to that divided by the temperature, the entropy increase that occur in the differential segment of the gap is given by the following.

$$dS = AG\ ds = d\dot{L}W/T = (fvG/T)\ dx$$
$$= AG\ (fv/AT)\ dx \quad (9)$$

where dS is the infinitesimal increase in the total entropy the segment,
A is the cross-sectional area of the gap $= h_{Gap}w_{Gap}$,
G is the flow density,
ds is the infinitesimal change in specific entropy

Thus, the increase in entropy over the length of the gap is

$$\Delta s = \int_0^{l_{Gap}} (fv/AT)\ dx \quad (10)$$

where l_{Gap} is the length of the gap

The relationship between entropy increase and flow density is derived from this equation. The following velocity profile is assumed for the derivation.

$$u(y)=u_{max}(1-(\frac{2y}{h_{Gap}})^{2n}) \quad (11)$$

where u(y) is the flow component along the gap,
y is the position within the gap,
y=0 being at the center of the gap,
h_{Gap} is the height of the gap,
n is the shape parameter,
4 to 5 for turbulent flow.

n=5 is used for this calculation. The entropy increase associated with the increment of lost work is the lost work divided by the temperature. The gap cross-section is also assumed to be constant through the length of the gap. Furthermore, the pressure profile is assumed to be linear. Finally, the relationship between the exit stagnation state (entropy) and the flow density G is given by

$$G= \frac{\Delta s\ h_{Gap}^2 SF(n)}{l_{Gap} < \frac{\mu v^2}{T} >_L} \quad (12)$$

where SF(n) is shape factor dependent on n,
$= (2n+1)/8n^2$
$<>_L$ indicates the average over the gap

The two relationships between the flow density G and entropy increase shown in Equations (6) and (12) are shown in Figure 4. Computations based on standard fixed point iteration can be performed to find out what the flow would be given the source state, gap dimensions, and sink pressure.

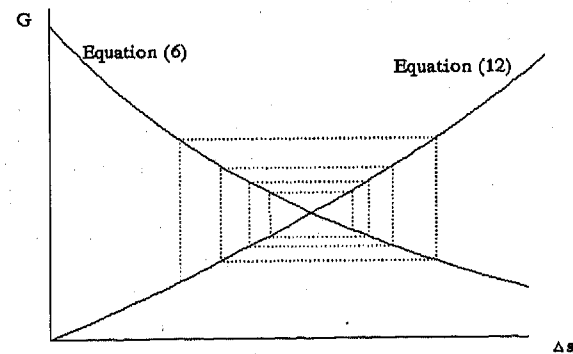


Figure 4. Relationship between G and Δs

The computations indicate that the choking would occur in most gap flow we are considering, that is, the pressure drop is quite small within the gap and the most of the pressure drop occurs at the exit of the gap. The gap flow density, therefore, is mostly a function of the flow origin state and the gap geometry. This also may provide an explanation for the findings that gap flow is better described using incompressible fluid model than compressible model without choking [5].

Unfortunately, this model still has two major flaws at the time of the writing. One is that the computation time is excessive; it costs as much computation time to compute one instance of gap flow as to compute the

entire cycle of compression chamber (with less sophisticated gap flow model). The other is that the computation is unstable in some regions on Δs-G map. To circumvent these problems, the computation is carried out for various combination of flow origin state and gap geometry for Freon 134a. An empirical formula created from the calculation results using linear regression in logarithm domain is used in the compressor simulation itself.

$$\ln G = a_0 + a_1 \ln l_{Gap} + a_2 \ln h_{Gap} + a_3 \ln P_{00} + a_4 \ln T_{00} \qquad (13)$$

where G is the flow density in kg/m²-sec
l_{Gap} is the gap length in m
h_{Gap} is the gap height in m
P_{00} is the origin stagnation pressure in MPa
T_{00} is the origin stagnation temperature in K
$a_0 = 20.69511$
$a_1 = -0.28968$
$a_2 = 0.659064$
$a_3 = 1.362197$
$a_4 = -1.22644$

2.4. VALVE

Valve characteristics that are important to the simulation are: (1) the forward flow from the compression chamber into the discharge plenum, and (2) the back flow. The forward flow characteristics determines the amount of excess compression and lost work, and is quite extensively studied. Back flow also plays an important role. During the preliminary investigation, it was found to be the most prominent correction to isentropic compression.

At this stage, the dynamic aspect of the valve behavior is not dealt with, since the simulation itself is that of steady state or constant rpm. Full dynamic response from the valve will be incorporated in the future for dynamic simulation with transient behavior. For both the forward and backward flow, the valve is modeled as a simple relationship between the mass flow rate and the pressure difference between the compression chamber and the discharge plenum.

The forward flow is modeled to be proportional to the 3/2 power of the pressure difference. The flow quantity is proportional to flow area and flow speed. The area is proportional to the lift of the valve, which is proportional to the pressure difference, and the flow speed is roughly proportional to the square root of the pressure difference. If the change in specific volume due to the pressure drop is not significant, these assumption correspond to the simplest model implied in Soedel [6].

$$dm/dt = C_F(\Delta P)^{3/2} \qquad (14)$$

where C_F is the constant for each valve

This hypothesis is confirmed by calculating the forward flow from the measured pressure data and plotting it against the pressure difference. Data from different operating condition (i.e. different rpm and discharge pressure) seem to line up the best for the power of 3/2. Of course, the coefficient can be derived from direct measurement.

The back flow is modeled differently, as the mechanism is quite different. The flow is assumed to be similar to the gap flow, where the gap closes as the pressure difference increases.

$$dm/dt = C_B(1-\Delta P/P_{CL})\ \Delta P^{1/2} \qquad (15)$$

where C_B and P_{CL} are valve dependent constants

The flow speed, again, is assumed to be proportional to the square root of the pressure difference. The better sealing of valve under higher pressure difference is represented by the term $(1-\Delta P/P_{CL})$. The coefficients can be based on actual measurement. In this simulation, these coefficient are determined by best fitting to the actual pressure data.

Another dynamic effect of the valve is the inertial delay of the closure, which causes the back flow as the vane crosses the discharge port. This is simulated by one time mass dump.

2.5. VANE AND ROTOR FORCE AND TORQUE LOADING

The vane loading is shown in Figure 5. The pressure in the back room of the vane is assumed to be the average of the two chambers adjacent to the vane. The friction between the vane and the rotor, or the friction between the vane and the cylinder wall are assumed to be dry friction characterized by one coefficient of friction.

The torque is computed by loading on the rotor

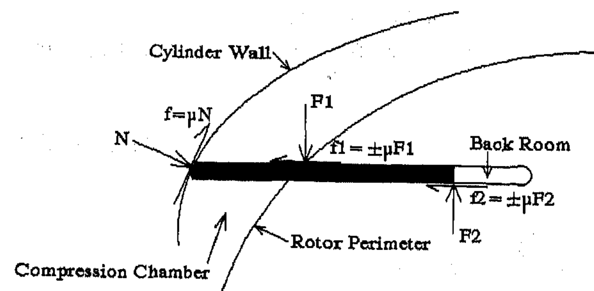


Figure 5. Vane Loading
(Pressure loading not shown)

coming from the vane and the pressure in the back room.

2.6. CYCLE EFFICIENCIES

Efficiency values that are of interest are volumetric efficiency of the compressor and coefficient of performance (COP) or the beta value for the entire cycle.

The volumetric efficiency is defined to be net mass flow rate divided by the product of compressor speed, compression chamber volume at the end of suction process and the density of the gas. In our case where there is no significant suction loss between the suction plenum and compression chamber, it is computed as the net mass that is discharged into the discharge plenum during one cycle of compression divided by the mass of gas that is present in the compression chamber at the start of compression, that is, when the compression chamber is physically separated from the suction plenum.

COP is the ratio of the heat absorbed in the evaporator divided and the work consumed by the compressor. A typical refrigeration cycle is shown on P-h diagram in Figure 6. Let subscript 1 denote the compressor inlet and evaporator outlet, 2 the compressor outlet and condenser inlet, 3 the condenser outlet and throttle valve inlet, and 4 the throttle valve outlet and evaporator inlet. Then, COP should be given by the following formula.

$$COP_{Experiment} = q_L/w_C = (h_1-h_4)/(h_2-h_1) = (h_1-h_3)/(h_2-h_1) \quad (16)$$

where q_L is heat absorbed per unit mass of refrigerant

w_C is the compressor work per unit mass of refrigerant flow

Rather than use the exact formula above, which involves the simulation and measurement of the entire cycle, we have chosen to use an approximate version in which the effect of pressure drop of the condenser is approximated by 5°C of subcooling. Thus, the cycle on which our COP calculation is based on is shown in the Figure as 1-2-3'-4'-1. Thus, the COP of our calculation is given by the following equation.

$$COP_{Simulation} = q'_L/w_C = (h_1-h_{4'})/w_{C'} = (h_1-h_{3'})/w_{C'} \quad (17)$$

where q'_L is the estimate of heat per unit mass of refrigerant based on 5°C subcooling

$w_{C'}$ is work required on the part of compressor per kg of discharged refrigerant, from the computation

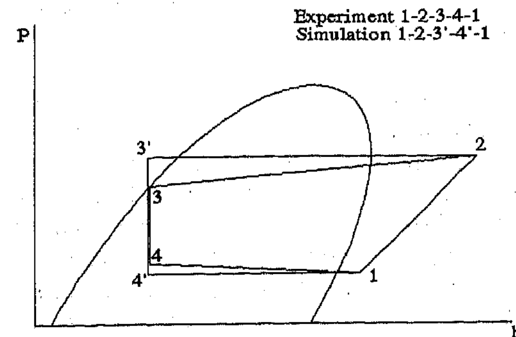


Figure 6. P-h Diagram for Refrigeration Cycle

3. DESCRIPTION OF NUMERICAL MODELING

3.1. INPUT AND OUTPUT

Two versions of the simulation were written; interactive and batch. In the interactive version, the input procedure is menu driven and the input data file for the batch version is organized in a fashion closely resembling the interactive menu. There are four menus for entering Design Parameters, Tolerance Parameters, Operation Parameters, and Computational Parameters. The Design Parameters menu contain items such as geometry of the cylinder and valve characteristics. The Tolerance Parameter menu contains various gap dimensions. The Operation Parameters menu contains such items as speed (in rpm), discharge pressure, and suction state. The Computational Parameter menu contains such items as the size of discretization of angular position, iteration repetition number and error criteria for convergence, and name of output file. Further details of each of the input variables are given in the appendix.

An operator can readily change the design variable through the menu. The batch version is used for computing multiple instances, and utility to create the batch input files through menu driven procedure was also written.

Each instance of simulation creates three files. The first is the header file, which is a duplicate copy of what the input parameters are. The second file is the pressure profile file, listing the chamber volume, pressure, temperature, mass, and mass flows through contact patch, vane sides, and discharge valve as a function of angular position. The third is the torque output. The first table contains the force loading on a vane and torque due to the loading on the vane as functions of angular position. At the end of the file, the overall thermodynamic performance parameters such as discharge temperature, volumetric efficiency and COP are also listed. The second set contains the torque loading from all five vanes (staggered by 72 degrees) and the combined torque as a function of angular position. At the end of the file, the total average torque and peak-to-peak variance is listed. Sample of each type of output files is given in Appendix.

3.2. PREDICTOR-CORRECTOR FOR PRESSURE CALCULATION

Predictor-Corrector scheme is used to compute the state of compression chamber. The canonical description of the compression process, i.e. the scheme represented in equation (3) is put in implicit form. The explicit form of the first law description represented in equation (2) is used to provide the initial guess to start the iteration loop. For the $\Delta\theta$ of 3 degrees and error ceiling of 0.001, in most part, the iteration converges within two tries.

The iterative method was chosen over an explicit scheme such as Runge-Kutta because of the computational instability that may develops shortly after the maximum pressure is reached. When the iterative computation becomes unstable, the bisection method is used. With the implicit iterative method, one can easily detect whether the iteration is converging or not, therefore, can judge whether to use bisection method for the next calculation.

3.3. ITERATIVE LOOP FOR OVERALL PROCESS

In order to compute the state at the next angular position, one needs to know the state of the chamber that is 72 degrees ahead and that of the discharge plenum, i.e. the source of leak and back flow into the chamber as shown in Figure 7a. But in the pressure calculation sequence, these are calculated later. To simultaneously solve for the chamber for every angle and discharge plenum state requires large computational power, and is inefficient.

This simulation uses iterative approach here as well. The state of chamber ahead and discharge plenum from the ith iteration loop is used to compute the state of chamber in the i+1st iteration loop. In the very first loop, all of the leaks are treated as if they are coming from the discharge plenum, whose state is estimated from the initial guess of the discharge temperature. This scheme is briefly represented in Figure 7b.

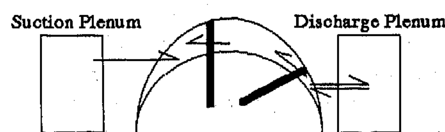


Figure 7a. Material Exchange between Chambers

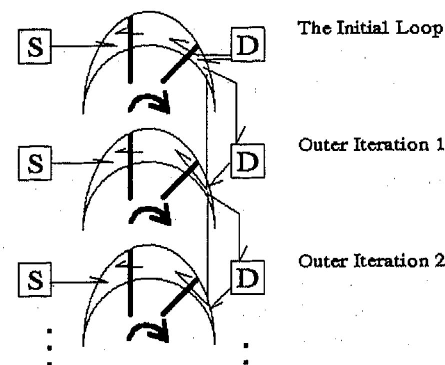


Figure 7b. Outer Iteration Loop

3.4. PARAMETER MATCHING AND VERIFICATION OF THE SIMULATION CODE

The important characteristic, such as the forward flow coefficient and back flow coefficients were not independently available. As a result, the technique of experimental design was used to find the values for these coefficients to best fit the compression chamber pressure profile data. The values determined by the method are used in subsequent calculations.

The operating conditions for the pressure profile data are given in Table 1 below. Figures 8a through d show the pressure profiles from the data set.

Table 1. Operating conditions for pressure profiles used in tuning.

Data Name	Speed (rpm)	Suction Pressure (Mpa)	Discharge Pressure (Mpa)
CV_10	800	0.2	1.05
CV_12	1800	0.2	1.05
CV_15	800	0.2	2.0
CV_17	1800	0.2	2.0

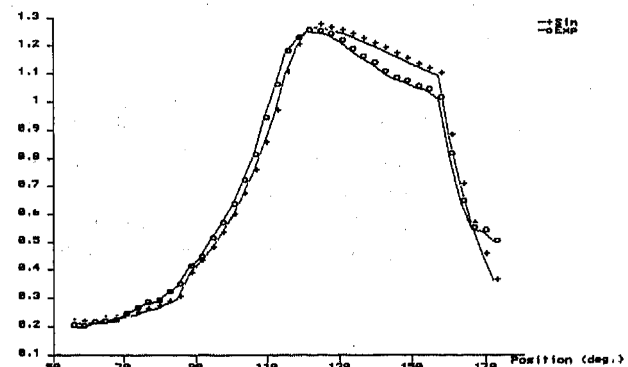
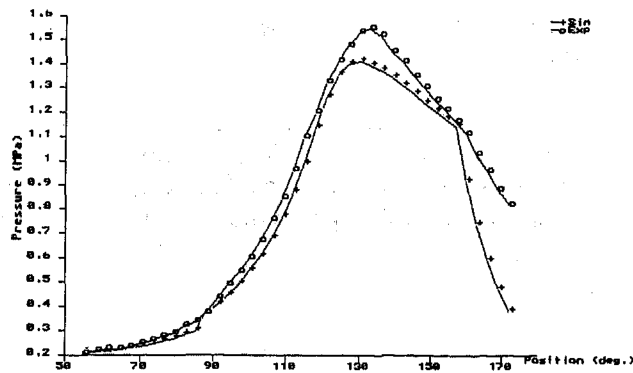
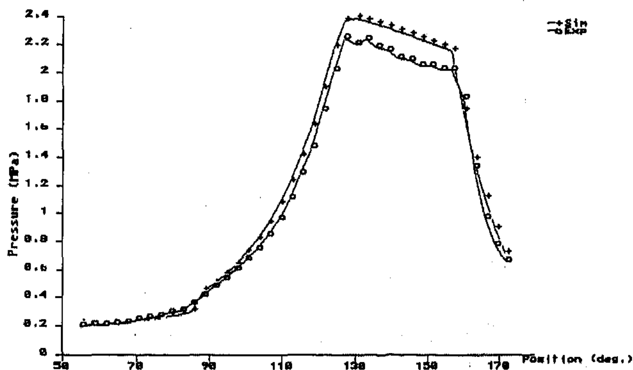
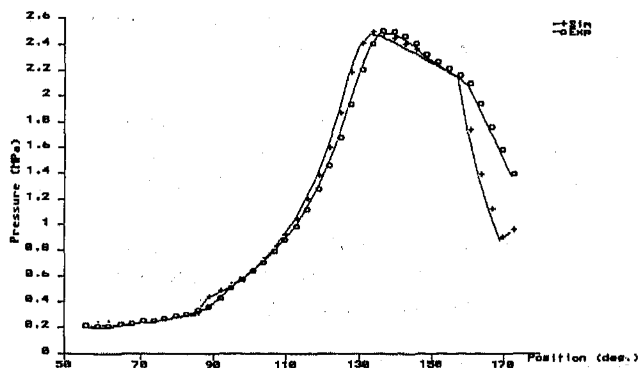


Figure 8a. Pressure Profile for CV_10


Figure 8b. Pressure Profile for CV_12


Figure 8c. Pressure Profile for CV_15


Figure 8d. Pressure Profile for CV_17

4. COMPUTATIONAL RESULTS

With the model now verified, design parameters can now be examined for their influence. Two examples are presented. The first example is the effect of the aspect ratio to torque, torque variance and compressor efficiency. The second example is the cylinder shape where the manner of vane extension and retraction are altered.

4.1 EFFECT OF ASPECT RATIO OF THE CYLINDER

The space available in the engine compartment for A/C compressor continue to shrink in size. In particular, the emphasis is on a slimmer compressor. Therefore, it is important to find out what effect this change in the aspect ratio will have, and what range of aspect ratios are reasonable.

4.1.1. Description Of Design Parameter Varied

The aspect ratio of unity is given to the current design. The diameter of the compressor is increased by the aspect ratio a, while the length is reduced by the square of the aspect ratio to keep the nominal volume constant.

4.1.2. Results

There is very little change in torque and torque variance, as shown in Figure 9a. This agrees with a simple energy calculation. The efficiency of the compressor, however, decreases as the aspect ratio is reduced, because the effect of leak through contact patch grows as the aspect ratio is reduced. The effect of the leak is easily seen in the pressure profile. For the low aspect ratio, the right "shoulder" of the pressure profile is reduced, indicating the stronger presence of mechanism to reduce the pressure of compression chamber after it reaches high pressure state, as shown in Figure 9b.

Corresponding to the reduction in efficiency, the rise in temperature of the discharge state is also anticipated.

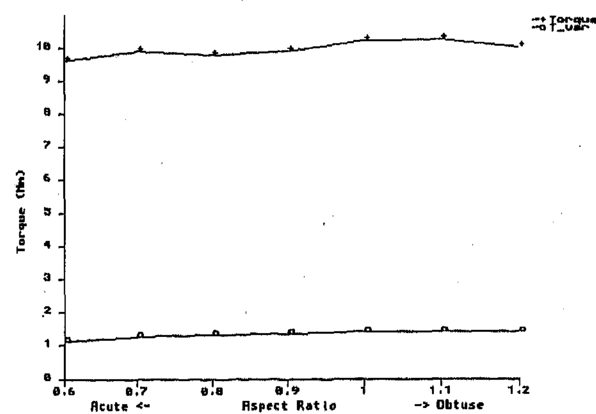
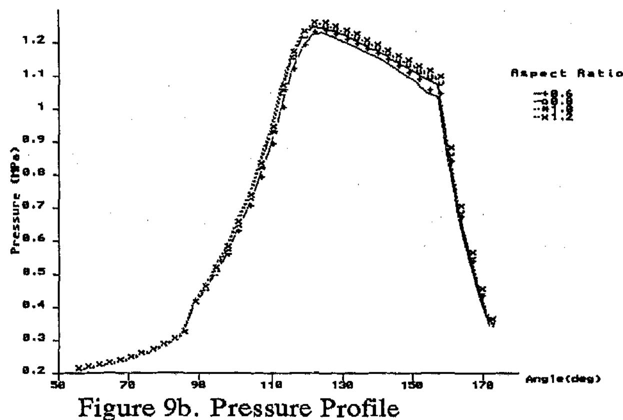


Figure 9a. Torque v. Aspect Ratio


Figure 9b. Pressure Profile

4.2 EFFECT OF VANE EXTENSION AND RETRACTION SCHEME

It is important to choose the cylinder shape rationally. To aid this, the cylinder shape is characterized by a number of parameters. The parameters then can be varied, according to design of experiment schemes, to estimate what combination of the parameters will yield the most desirable effect.

Here, the desired effect is the reduction of torque variance and improved efficiency.

4.2.1. Description of Design Parameters Varied

The cylinder shape is parametrized by the vane excursion as a function of rotor angular position. The excursion function is C^2, that is, its second derivative is continuous. The parameters that are varied are the beginning position of the maximum excursion of the vane, beginning position of the retraction of the vane, and the skew parameter for the extraction/retraction path. The maximum extension start position and the skew parameter determines how the vane extends, and the retraction start position and the skew parameter determines how the vane is retracted. Further description of the vane position is given in the appendix.

For the results in this paper, the maximum extension of the vane is held constant.

4.2.2. Results

The results indicate that the torque variance and efficiency are most greatly affected by how the vane retracts. The manner with which the vane extends does not have a major effect on torque variance, etc. Specifically, the position of vane retraction start has the greatest bearing on torque variance. The torque variance is determined by the height and width of torque profile from each vane. When the retraction starts later, the peak becomes higher and narrower, resulting in higher variance.

The simulation indicates that the vane can be extended rather quickly, giving the possibilities of increasing the length of the contact patch. Increasing the length of the contact patch may result in improved efficiency.

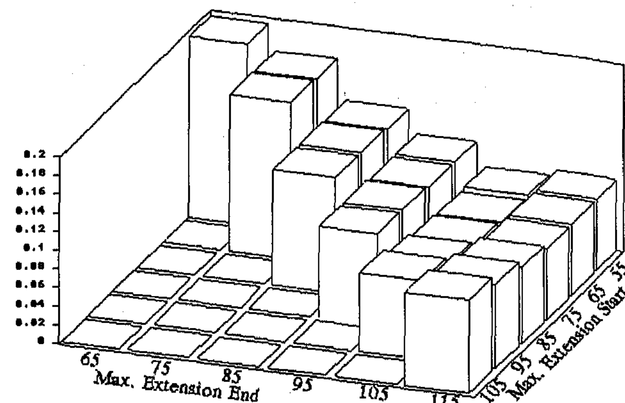


Figure 10. T_var/T_av v. Cylinder Profile Parameters

5. CONCLUSION

The compressor simulation code is constructed for steady state (constant rpm) operation. The simulation uses "real" R134a property and use of the first law with the emphasis on entropy accounting, rather than the conventional polytropic compression curve.

This simulation provides solid foundation for future expansion into dynamic and transient compressor simulation and overall cycle and system simulation.

Simulation results indicate that it is a useful tool in optimizing the compressor design parameter, such as aspect ratio and cylinder shape profile.

6. ACKNOWLEDGMENT

Author wishes to thank Mr. Shitara, Mr. Enokido and Zexel Compressor Design and Experiment Group for providing compressor design detail and performance test in formation; Mr. Butts, Mr. Sumikawa, Mr. Bearnt, Mr. Stricklen, Mr. Kato and Mr. Zhou at Zexel USA for valuable discussions over compressor engineering; Dr. Kobayashi and Dr. Miyamoto at Zexel R&D Center for guidance on fluid flow and leak mechanisms.

7. REFERENCES

1. Muta, S.
   Rotary Vane Compressor Simulation
   Zexel Internal Technical Report 54-CPV-90

2. Fukuta, M., Yanagisawa, Y., Shimizu, T., Suzuki, Y.
   Mathematical Model of Vane Compressors for Computer Simulation of Automotive Air Conditioning Cycle, JSME Int'l J., Series B, Vol 38, No. 2 pp199

3 Soedel, W.
   Mechanics, Simulation, and Design of Compressor Valves, Gas passages and Pulsation Muffleres. 1992

4. Makita, K (chief ed.), et. al., Japan Ass'n of Refrig, Japan Freon Gas Ass'n
   Thermophysical Properties of Environmentally Acceptable Fluorocarbons, 1990

5. Ishii, N., Bird, K., Sano, K., Oono, M., Iwamura, S., Otokura, T.
   Refrigerant Leakage Flow Evaluation for Scroll Compressors
   1996 Int'l Compressor Compressor Engineering Conference at Purdue

6. Soedel, W.
   Design and Mechanics of Compressor Valves, 1984

8. APPENDIX

8.1 SIMULATION INPUT VARIABLES

Design Parameters

1. Cylinder profile file
   Path and name of the cylinder profile data file. ASCII file with angle and r as a function of the angle. One pair of entry per line.
2. Rotor radius
3. Cylinder thickness
4. CP start angle
   Contact patch start
5. CP end angle
   Connect patch ends
6. Vane offset
   Distance between the center of the rotor and the line of vane motion
7. Vane Thickness
8. Vane off-center
   Distance between the center of the vane tip arc and the center of the vane.
9. Vane tip radius
10. Vane length
11. Vane mass
12. Vane-Cyl frict. coeff.
13. Vane-Rot frict. coeff.
14. Valve fwd. coeff.
    m_dot = CF x delta_P^3/2
15. Valve bwd. coeff.
    m_dot = CB x (1-delta_P/PCL) x delta_P^1/2
16. Valve bwd. P0
    PCL in the equation above.
17. Valve bwd. dump
    delta_m at port start position
18. Port Volume
19. Port Start

Tolerance Parameters
1. CP radial clearance
2. CP leak fudge   default 1.0
3. Vane length clearance
4. Vane tip gap width   Set to zero for this version
5. Vane side clearance
   Compression chamber to slot back room leak. Not addressed in this version
6. Vane Leak Fudge   default 1.0

Operation Parameters
1. Compressor RPM
2. Suction Port Closure   default max. intake position
3. Discharge Pressure   MPa
4. Discharge Temperature   K, Initial guess
5. Suction Pressure   MPa
6. Suction Temperature   K

Computation Parameters
1. Del_theta
   Angular step (3deg for the calculations shown in this paper)
2. Inner Iteration Max.   default 10
3. Inner Itr. Epsilon   default 0.001
4. Bisec Iteration   default 10
5. Bisec Search Bound   default 0.2.
   May need to be increased for high discharge pressure or high rpm calculation
6. Outer Iteration Max.   default 20
7. Outer Itr. Epsilon   default 0.01.
   Measures the convergence of the flow from a compression chamber to the following chamber when the latter has reached the discharge port position
8. Output File Name Stem
   Output file name. Default is based on the clock reading so that the output form the earlier calculation would not be overwritten

8.2 EQUATION OF VANE EXCURSION

The vane motion has five stages: retracted stage (prior to the end of the contact patch), where the position stays at the rotor radius, the extension stage where the vane moves from the rotor radius to maximum radius, maximum extension stage where the vane is at

the maximum radius, retraction stage, where the vane retracts from the maximum radius to rotor radius, and retracted stage (after the start of the contact patch). The first retracted stage ends when $\theta \leq \theta_a$, the extension stage is when θ is between θ_a and θ_b, the maximum extension stage is when θ is between θ_b and θ_c, the retraction stage is when θ is between θ_c and θ_d, and the second retracted stage starts when $\theta \geq \theta_d$. Let E_{max} be the maximum excursion,

$$\text{excursion}(\theta) = \begin{cases} 0, & \theta < \theta_a \\ E_{max}(f_0(\xi) + S \cdot f_1(\xi)), & \theta_a < \theta < \theta_b \quad \text{where } \xi=(\theta - \theta_a)/(\theta_b - \\ E_{max}, & \theta_b < \theta < \theta_c \\ E_{max}(f_0(\xi) + S \cdot f_1(\xi)), & \theta_c < \theta < \theta_d \quad \text{where } \xi=(\theta_d - \theta)/(\theta_d - \theta_c) \\ 0, & \theta > \theta_d \end{cases}$$

where

$$f_0(\xi)=10.0\xi^3 - 15.0\xi^4 + 6.0\xi^5$$
$$f_1(\xi)=-\xi^3 + 3.0\xi^4 - 3.0\xi^5 + \xi^6$$

For this calculation, E_{max} of 8.5mm, θ_a of 8 deg., and θ_d of 172 deg. were used, while θb, θc and the skew parameter S were varied.

8.3 SAMPLE OUTPUT FILES

Pressure Output Format (Section 4.2)

```
sim_098595.pres
See the header file for condition information

Angle       Volume          Pressure   mass           Temperature dm_cp          dm_72          dm_v

55.000000   1.330373e-05    0.210000   1.334266e-04   273.149994  0.000000e+00   0.000000e+00   0.000000e+00
58.000000   1.323005e-05    0.222148   1.279202e-04   277.267731  0.000000e+00   4.493632e-06   0.000000e+00
61.000000   1.307934e-05    0.235784   1.321875e-04   281.110077  0.000000e+00   4.367390e-06   0.000000e+00
64.000000   1.185539e-05    0.250410   1.362016e-04   284.720551  0.000000e+00   4.014149e-06   0.000000e+00

                            ......

169.000000  1.571560e-07    0.447445   4.735728e-06   315.445546  4.763376e-07   1.536172e-06   0.000000e+00
172.000000  1.570796e-07    0.358131   6.115806e-06   306.103394  3.620016e-07   1.762079e-06   0.000000e+00

Discharge state temp 348.706726
Volumetric efficiency 0.455928
Isentropic efficiency 0.618185
Heat capacity 1.134107
```

Torque Output (Used for Section 4.2)

```
sim_098595.torq
See the header file for condition information

Angle       Torque          Normal          Rotor           Root
                            ......

91.000000   3.695436e+00,   4.079009e+00    1.781032e+02,   7.646142e+01
94.000000   4.312363e+00,   4.553234e+00    2.079603e+02,   8.929834e+01
97.000000   5.018242e+00,   5.629814e+00    2.419070e+02,   1.038354e+02
100.000000  5.807233e+00,   8.605466e+00    2.786978e+02,   1.191164e+02
103.000000  6.648131e+00,   1.378350e+01   3.158032e+02,   1.334575e+02
```

970118

Seat Belt Limitations in Collisions that Involve No Compromise of the Passenger Compartment

J. Hill and G. M. Mackay
Birmingham Accident Research Centre

S. Henderson
Rover Group Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This is a field study of cars that sustained very little passenger compartment intrusion in frontal crashes. It has been possible to define the key limits on seat belt effectiveness in these, the most common type of crash.

Data are taken from the in-depth investigations by the Co-operative Crash Injury Study and Rover Group in the U.K. Vehicle crashworthiness is assessed in detail and then matched with medical information from coroners reports, hospital records or questionnaires sent to survivors.

The study describes and compares important injury mechanisms such as driver head to steering wheel contacts, front passenger torso injuries caused by seat belt loads and front occupant leg injuries. 14% of occupants were rated two or above on the Maximum Abbreviated Injury Scale.

It is realistic to design seat belt systems for use in passenger compartments that can be expected not to deform in the more common frontal crashes. The discussion is concerned with the importance of these results when developing safety systems in that context.

INTRODUCTION

Looking to the near future we can expect new passenger cars to have ever increasing standards of structural front-crash performance. Present-day knowledge of design and testing can allow the passenger survival space to remain uncompromised in many crashes at impact speeds at least equal to current and proposed legislative impact tests.

Much research and development is presently concerned with making the most effective use of occupant interior space. This work focuses on seat belt supplementary technologies such as pretensioning devices and airbags. There is also a growing desire to develop dynamic, integrated restraint systems that will be effective over a wide variety of occupant characteristics and impact conditions. The specifications of these new technologies will draw heavily on field research into road traffic accidents.

This paper is a study of field cases where the crash effectiveness of seat belts was not compromised by any disruption of passenger interior-space. Results should then be directly relevant to restraint system design and modelling where the compartment is often considered undamaged.

In this paper the term "intrusion" refers to deformation of the passenger compartment measured inside towards the occupants. Vehicles chosen for this study did not have passenger compartment intrusion. They were often from "low-energy" crashes with injuries of relatively minor severity. There are, however, several good reasons to take interest in these cases. Trivial injuries caused by internal vehicle contacts provide information about occupant kinematics. Head strikes, for example, may only have caused external injuries, but they should be taken in to account because more severe impacts may cause internal injuries in otherwise similar conditions. Occasional, more severe, injuries can define important limitations to the effectiveness of current restraint systems, perhaps in terms of occupant or vehicle characteristics.

Horsch (1987) has shown that the relatively high frequency at which minor injuries occur relative to more serious or fatal crash outcomes, means that these injuries are the much more costly consequence to society as a whole. This is another reason why this study was considered important. Neck strains, for example, were one of the most numerous consequences of crashes in this study. The Abbreviated Injury Scale (AIS; AAAM, 1990) rates neck injuries as "minor" in terms of their relative severity, but these injuries are known to have both long-term and costly consequences.

ANALYSIS AND RESULTS

It is useful first to consider injuries occurring in frequent, "low-energy" impacts that involve current model cars. The Rover Group Accident Injury Study was commenced in August 1995 to examine vehicles involved in collisions that are owned and operated by Rover Group. The total fleet consists of approximately 20,000 vehicles, the majority of which are under 12 months old. The vehicles are driven by employees and families of employees of Rover Group and its associated companies. Medical details are obtained from

hospitals and questionnaires sent to the car occupants. All vehicles with significant damage (cost of labour to repair > £1000) are investigated.

The sample size is small at present (72 vehicles involved in frontal impacts with a total of 110 occupants), and mostly minor injuries have so-far been reported (33% of occupants rated AIS 1). In addition, two occupants were rated MAIS 2. One of these was an unrestrained driver who suffered loss of consciousness; the other was a restrained rear seat occupant who suffered two fractured ribs. Table 1 has been included to show the incidence of minor injuries in this accident sample that essentially represents typical, "every-day" crashes, involving current model vehicles. Injuries due to seat belt loads, neck strains and impacts by occupants on forward structures form the basis of this paper.

Injury Category	No Injuries	Seat Belt Bruising (AIS 1)	Neck Strain (AIS 1)	Impact Injuries (AIS 1)
Frequency	*71*	*9*	*22*	*24*
Rate	*65%*	*8%*	*20%*	*22%*

Table 1: Injuries sustained by 35 occupants from the Rover Study *Some occupants sustained more than one type of injury described above*

The Co-operative Crash Injury Study (CCIS) database was used for the rest of this study (Mackay, 1985; Hassan, 1996). The CCIS was very suitable for this purpose because it represents a full range of injury severities. It has been underway in the UK since December 1983. In-depth vehicle examinations are made at garages or scrap yards several days after road accidents occur. Medical details of car occupants are obtained from hospital records, coroners reports and questionnaires sent to survivors. Accidents are sampled according to a plan that ensures that a full range of accident injury severities are obtained equally from crashes occurring in rural and urban areas. The sample plan favours vehicles that contained fatal and seriously injured car occupants (rated according to the UK DoT system). Less-severe and non-injury cases are, however, very well represented. This was an important requirement of the present analysis requiring vehicles essentially from "low-energy" and low-injury severity crashes.

The current study makes use of records for accidents that occurred from April 1992 to December 1995. Accident records from before 1992 are not presently compatible with later cases.

For this study, a sub sample of seat-belted front occupants was selected from vehicles having sustained a front impact with no intrusion. Injuries of AIS 2 and above (AIS 2+) were initially studied. Secondly, all injury (AIS 1+) details were summarised to isolate the main types of injury. Occupant and crash characteristics were then compared.

FRONT IMPACTS WITH PASSENGER COMPARTMENTS UNCOMPROMISED - Occupants were only included for study if their vehicles met the following criteria:-

1. The most severe (or high energy) impact was frontal,
2. Secondary impacts, or roll-overs, were permitted if less severe than an initial front collision,
3. Zero intrusion of the passenger compartment measured inside the vehicle, towards the occupants, from any direction,
4. No steering wheel displacement.

Figure 1 shows that there were 1200 vehicles without any intrusion from front impacts.

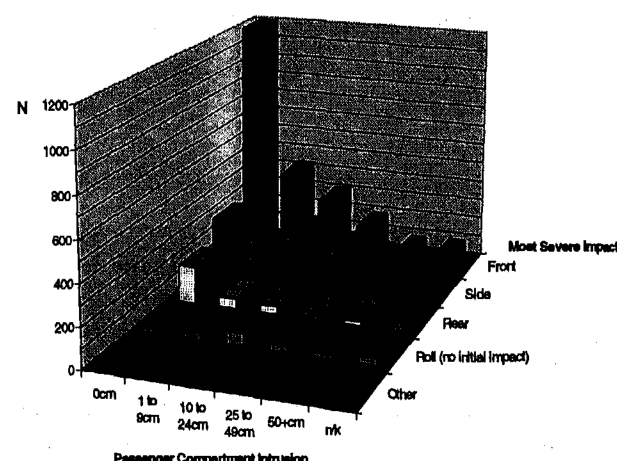


Figure 1: Intrusion vs. Impact Type for All Vehicles on the CCIS Database

FRONT SEAT OCCUPANTS - The major part of this study involves occupants selected according to the following criteria:-

Drivers and outboard-seated front seat passengers (FSPs),

Seat-belted using an emergency locking retractor, three-point (lap & diagonal) system,

Appropriately belted: occupants were excluded if known to be using belts in an inappropriate manner, if they were small children, or if a child restraint system was used. After consideration of work by Stalnaker (1993), small children were said to be occupants whose mass was < 35 kg, or if mass unknown < 13 years.

CCIS SUB-SAMPLE - Filtering occupants, as described above, on the CCIS database produced a sub-sample of 766 drivers and 188 FSPs. These can usefully be compared with all front occupants currently on the CCIS database. Crash severity was assessed from vehicle damage dimensions using the CRASH3 programme (NHTSA, 1982). It can be seen that a relatively "mild" sub-sample of cases has been selected, encompassing a less severe range of ETS (Figure 2, over) and occupant Maximum Abbreviated Injury Scale (MAIS, Figure 3). Two hundred and fifty nine occupants were not injured at all (MAIS 0). As might be expected, the sub-sample contained more females (Figure 4), FSPs, and older people (Figure 5).

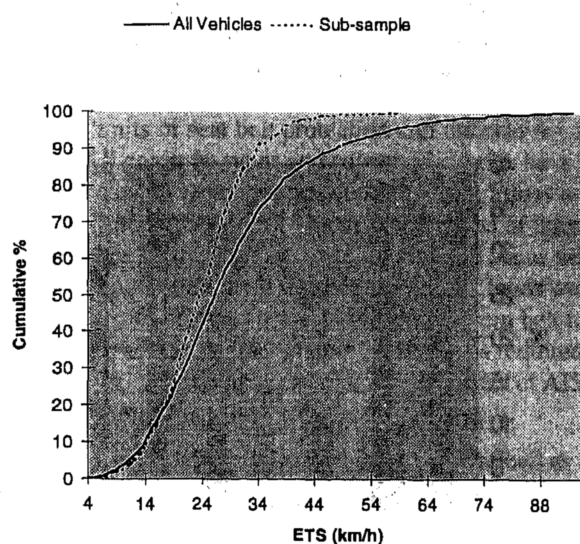


Figure 2: **ETS Distributions for All CCIS Vehicles vs. the Current Sub-sample**

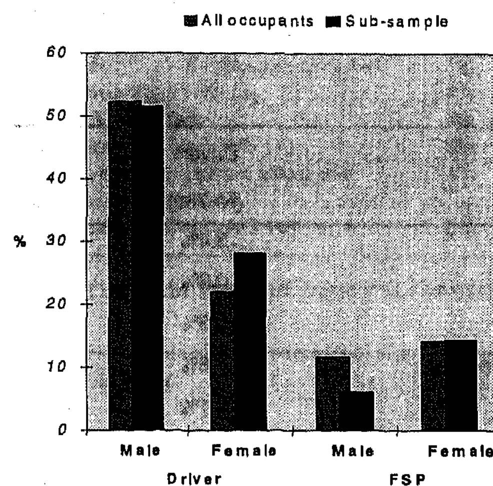
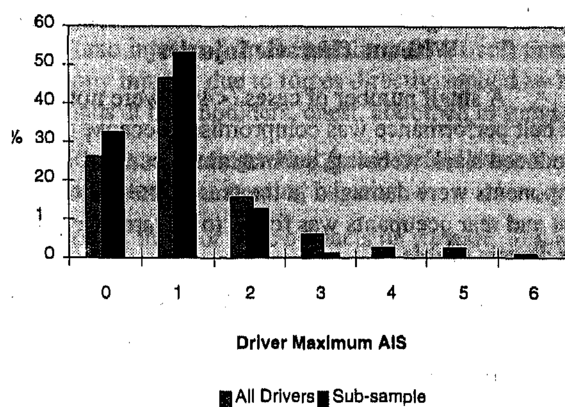


Figure 4: **Occupant Type Distributions Showing All CCIS Drivers and FSPs vs. the Current Sub-sample**



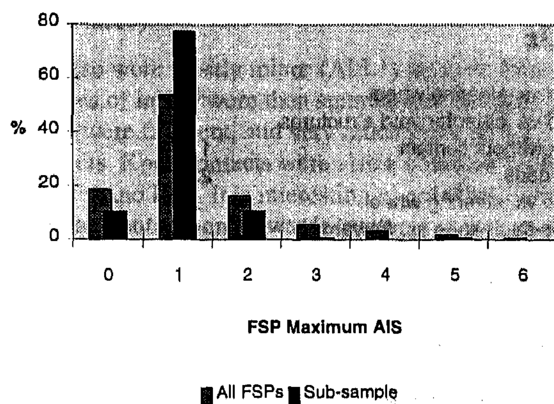
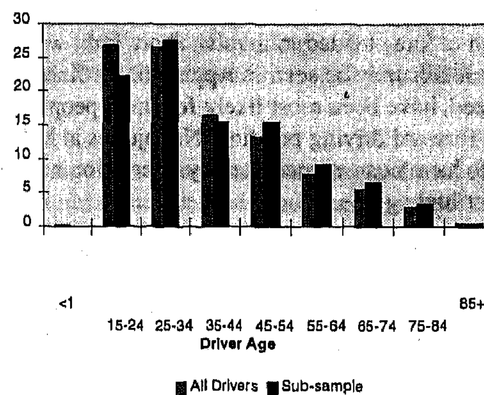


Figure 3: **Maximum AIS Distributions for All CCIS Occupants vs. the Current Sub-sample**



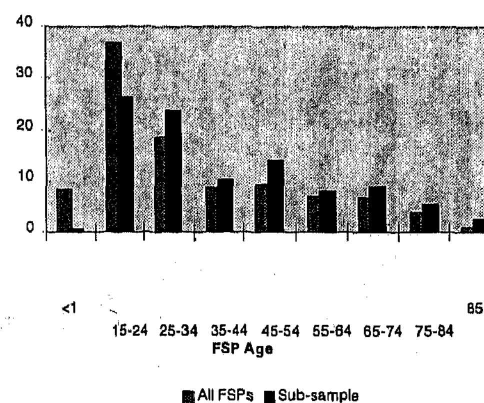


Figure 5: **Age Distributions for All CCIS Occupants vs. the Current Sub-sample**

Despite the relatively mild nature of these frontal impacts 13.8% of occupants were injured with MAIS 2+ and 1.6% (15 people) with MAIS 3+. The body regions injured at AIS 2+ are shown in Table 2. Main causes of concern are the thorax and, for drivers, the head.

	Drivers (N=766)	FSPs (N=188)
Head	3.4%	0.5%
Face	0.3%	
Neck	0.3%	
Upper Extremities	1.6%	1.1%
Thorax	6.1%	6.4%
Abdomen	-	1.1%
Spine (T & L)	0.3%	1.1%
Pelvis, Thigh & Knee	1.6%	1.6%
Lower Leg	1.2%	1.1%

Table 2: AIS 2+ Injury Rates

There were two neck injuries and several instances of thoracic/lumbar spine injuries. Injuries to female drivers at pelvis, thigh or knee tended to involve short, light-weight occupants (details may be seen in Appendix A). Knee contacts would, indeed, have been most likely for small people requiring a forward driving position. No injuries at MAIS 2+ were seen to have been exacerbated by interaction with other occupants or luggage.

It was possible to consider the interaction of various occupant and vehicle factors including occupant age, weight and height data together with seat belt performance and interaction between occupants. (Appendix A provides full details of occupant characteristics comparing all occupants (MAIS 0+) against AIS 2+ injuries at each body region.) Older occupants are more prone to an AIS of 2+ (Figure 6).

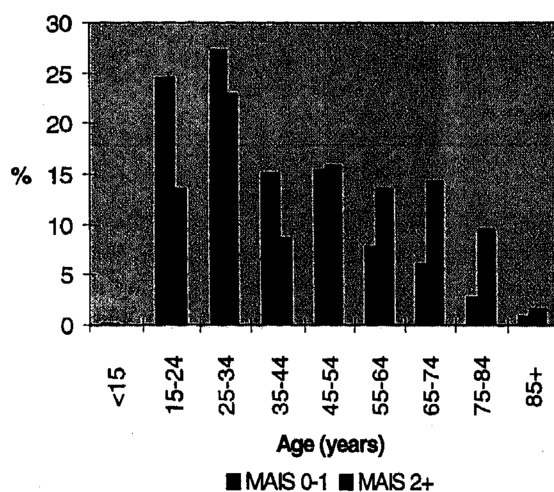


Figure 6: Age Distributions for All Front Seated Occupants With and Without Injuries Rated MAIS 2+

Advanced age was chiefly a factor affecting the thorax particularly considering female FSPs. Female FSPs were present over a very wide age range, but the majority (>75%) who suffered thoracic injuries were 65 years or older (Figure 7).

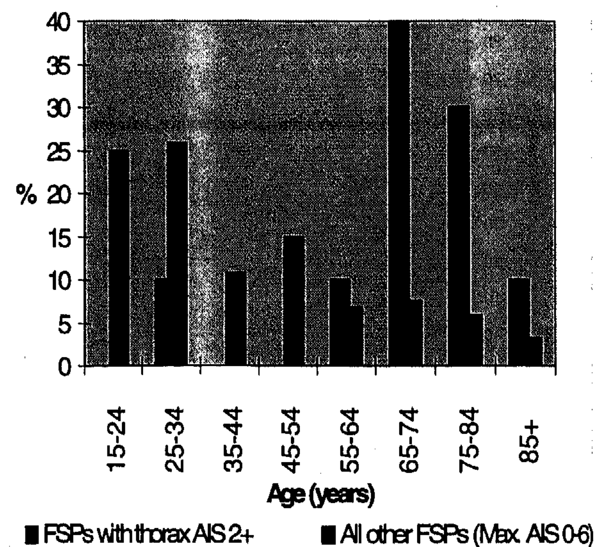


Figure 7: Age Distributions for Female FSPs With And Without Thoracic Injuries

A small number of cases (<4%) were noted where seat belt performance was compromised because clothing had introduced slack webbing, submarining occurred or seat belt components were damaged in the crash. Interaction between front and rear occupants was found to be rare affecting only four drivers (Table 3).

	Drivers (N= 766)	FSPs (N= 188)
Interaction between occupants/luggage	4	-
Seat belt compromised	30	8
AIS 1+		
Head vs. steering wheel	58	-
Head vs. other forward structures	6	8
Neck without contact	174	73
Belt loads	232	100
Torso vs. steering wheel	20	-
Pelvis-thigh-knee vs. forward structures	91	25
Legs vs. facia or footwell	43	15
AIS 2+		
Head vs. steering wheel	12	-
Head vs. other forward structures	1	1
Neck without contact	-	-
Belt loads	35	14
Torso vs. steering wheel	8	-
Pelvis-thigh-knee vs. forward structures	7	1
Legs vs. facia or footwell	7	1

Table 3: Types of Injury and other Factors

MAIN TYPES OF INJURY - The next objective of this study was to summarise common injurious occupant contacts on forward structures (AIS 1+), and other injury mechanisms. AIS 1 injuries arising from various forward contacts were worthy of detailed attention because they can describe the limits of seat belt protection and indicate occupant kinematics. All common types and causes of injuries have been described by several recent reports (ETSC, 1993; Fildes et al., 1991; Mackay and Hill, 1995). The principal types of injury and contributory factors that are widely known to occur were counted, as shown in Table 3. Injuries due to belt loads were extremely common. However, because injuries from belt loads are inevitable considering the magnitude of forces required to restrain people in frontal impacts, only such injuries at AIS 2+ were judged appropriate for further consideration.

Main types of injury were then briefly defined as follows:-

1. Head/face vs. steering wheel,
2. Head/face vs. other forward structures (includes contact on A pillars, windscreen, bonnet, header rail & facia),
3. Neck non-contact injuries, (includes soft tissue & bony injuries caused by decelerative forces),
4. Torso injuries form belt loads (AIS 2+ soft tissue or bony injuries due to forces directly applied by belt loads at the shoulders, chest, abdomen or pelvis),
5. Torso vs. steering wheel (some belted drivers contacted their steering wheels),
6. Pelvis, thigh or knee contacts on forward structures (caused by contacts on facia or steering column, but not steering wheel, and may also include indirect pelvic /thigh injuries due to knee contact on the facia),
7. Legs vs. facia or footwell (any injuries due to contact of extremities below the knee, on facia or footwell surfaces).

These were mostly minor (AIS 1) injuries. Rates for the main types of injury were then summarised in Table 4. Neck strains were frequent, and very much more so for front seat passengers. Knee contacts were also a common problem for both driver and FSP. It is interesting to note that FSPs sustained head contacts on forward structures at the same rate as drivers on steering wheels. It should be borne in mind that these results relate to a variety of frontal impacts some including multiple collisions or roll-overs. Accident reconstructions are carefully conducted and these results have been calculated where researchers concluded that occupant contacts occurred on frontal structures. Table 5 was produced to clarify this situation.

Table 5 shows injury rates for single frontal impacts from straight ahead of the vehicle (12 o'clock direction of force). Now the results suggest that drivers contacted their steering wheels more often than FSPs strike forward structures. Small increases in belt loading rates over results from Table 4 may well be significant as they indicate increased injury probability under more direct frontal loads.

Main Types of Injury	Drivers (N=440)	FSPs (N=109)
Head vs. steering wheel (AIS 1+)	8%	-
Head vs. other forward structures (AIS 1+)	1%	8%
Neck without contact (AIS 1+)	24%	40%
AIS 2+ due to belt loads	5%	5%
Torso vs. steering wheel (AIS 1+)	3%	-
Pelvis-thigh-knee vs. forward structures (AIS 1+)	13%	14%
Legs vs. facia or footwell (AIS 1+)	6%	8%

Table 4: The Main Types of Injury in Frontal Impacts

Main Types of Injury	Drivers (N=440)	FSPs (N=109)
Head vs. steering wheel (AIS 1+)	7%	-
Head vs. other forward structures (AIS 1+)	4%	4%
Neck without contact (AIS 1+)	21%	43%
AIS 2+ due to belt loads	6%	6%
Torso vs. steering wheel (AIS 1+)	3%	-
Pelvis-thigh-knee vs. forward structures (AIS 1+)	12%	13%
Legs vs. facia or footwell (AIS 1+)	5%	9%

Table 5: 12 O'clock Single Impacts No Belt-compromise or Occupant Interaction

Age was seen to be an important factor regarding torso injuries from belt loads and steering wheel contacts for all occupants. Typical results are shown in Figure 8 and Figure 9 for injuries caused by seat belt loads and steering wheels respectively.

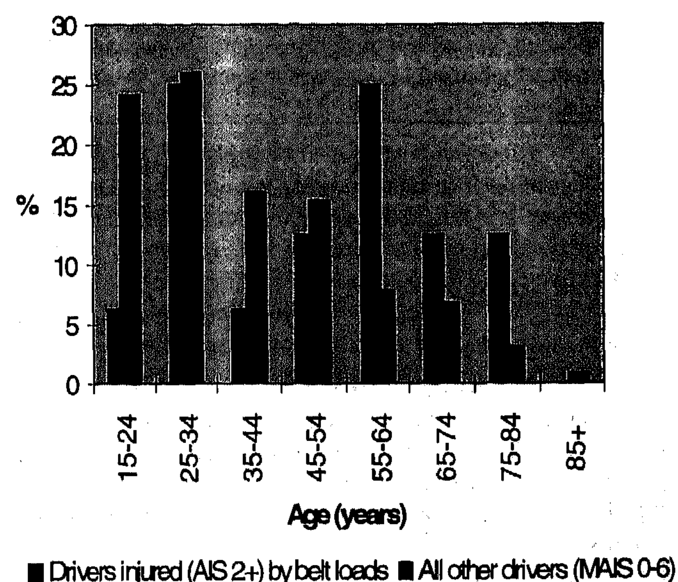


Figure 8: Age Distributions for Male Drivers With and Without Injuries to the Torso Caused by Seat Belt Loading

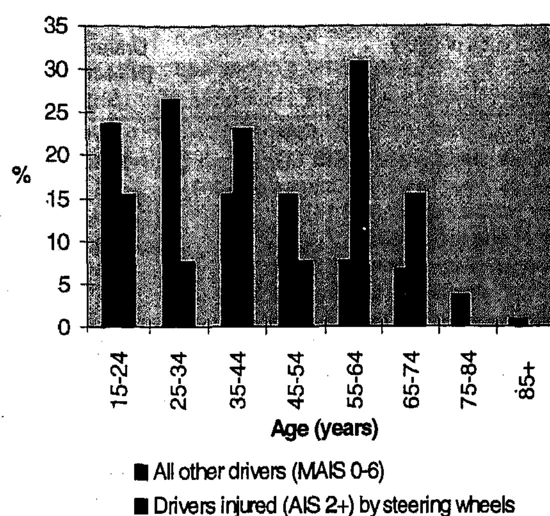


Figure 9: Age Distributions for Male Drivers With and Without Injuries to the Torso Caused by Contact with Steering Wheels

DISCUSSION

The main types of injuries to belted front seat occupants were identified for frontal crashes. These may be said to define limitations on seat belt effectiveness when the passenger compartment remains undeformed, and there is no steering wheel displacement.

The data gives a picture of car crashes on UK roads in the 1990s. All types of passenger car and years of registration are adequately represented, however the CCIS database is deliberately biased towards more severe impacts and many types of vehicles will be based on old designs. The Rover study, on the other hand, represents the "every-day" crash situation for current model cars often with supplementary restraint systems fitted. In the Rover subset the early data would suggest that minor injuries continue to occur and these are typically caused by forward contacts, belt loads and neck strains.

More frequent fitment of airbags may reduce the number of neck strains occurring to both drivers and FSPs in the future. It is hoped that future restraint systems will be tailored dynamically to occupant and crash characteristics, the objective being to make the best possible use of the occupant interior space (Mackay, 1994). Benefits should then extend to more severe impacts than those discussed here.

This paper is a first assessment of recent CCIS data. It is hoped that it will provide early clues for those designing restraint systems. A major challenge, and priority, could be to provide advanced load limiting, selectively for older people, without compromising protection through increased ride-down.

ACKNOWLEDGMENTS

The Co-operative Crash Injury Study is funded by the Department of Transport, Ford Motor Company Limited, Nissan Motor Company, Rover Group, Toyota Motor Company and Honda Motor Company. The project is managed by the Transport Research Laboratory. Grateful thanks are extended to Katie Stabler for her assistance in the preparation of this manuscript, and everyone involved with the CCIS data collection process.

REFERENCES

American Association for Automotive Medicine. The Abbreviated Injury Scale 1990 Revision. Association for the Advancement of Automotive Medicine; 1990.

ETSC. Technical Annexes. Reducing Traffic Injuries Through Vehicle Safety Improvements - The Role of Car Design. E.T.S.C.; 1993.

Fildes, B. N.; Lane, J. C.; Lenard, J.; Vulcan, A. P. Passenger Cars and Occupant Injury. Federal Office of Road Safety; 1991. ; CR95.

Hassan, A. M.; Hill, J. R.; Parkin, S.; Mackay, G. M. Secondary Safety Developments : Some Applications of Field Data. Autotech 1995; I Mech E; 1995.

Mackay, G. M.; Galer, M. D.; Ashton, S. J.; Thomas, P. D. The Methodology of In-depth Studies of Car Crashes in Britain. International Congress and Exposition; 1985; Detroit, Michigan, U.S.A. : S.A.E.; 1985:365-390.

Mackay, M. Smart Seat Belts - Some Population Considerations Applied to Intelligent Restraint Systems. International Congress & Exposition; 1994 Feb; Detroit, Michigan. Warrendale, PA: S.A.E.; 1994; 940531.

Mackay, G. M.; Hill, J. R. The Limitations of Current Seatbelts In Europe - Some Population Considerations. The Journal of Trauma: Injury, Infection and Critical Care. 38:533-537; 1995.

NHTSA. CRASH3 User's Guide. U.S. Department of Transportation, National Highway Traffic Safety Administration; 1982.

Stalnaker, R. L. Inconsistencies in State Laws and Federal Regulations Regarding Child Restraint Use in Automobiles. Child Occupant Protection. S.A.E.; 1993: SP-986 933087: 51-69.

APPENDIX A: OCCUPANT CHARACTERISTICS & INJURIES RATED AIS 2+

Various occupant and vehicle factors are tabulated. Age, weight and height data are compared for all occupants (MAIS 0+) against AIS 2+ injuries at any given body region. Inter-quartile ranges are given in square brackets. Alternatively, values were listed when a cell contained less than six known values.

		Male		Female	
		Driver	FSP	Driver	FSP
All occupants	age	[25 to 51 yrs]	[20 to 50 yrs]	[26 to 48 yrs]	[25 to 61 yrs]
(MAIS 0+, n=954)	weight	[70 to 86 kg]	[67 to 85 kg]	[56 to 71 kg]	[55 to 71 kg]
	height	[1.7 to 1.8 m]	[1.7 to 1.8 m]	[1.6 to 1.7 m]	[1.6 to 1.7 m]
Head	age	[21 to 60 yrs]		[25 to 70 yrs]	67 yrs
(AIS 2+, n=27)	weight	[69 to 80 kg]		55,68,83 kg	
	height	[1.7 to 1.8 m]		1.6,1.7,1.7,1.7 m	
	belt compromised interaction	n=1			
Face	age	34 yrs		29 yrs	
(AIS 2+, n=2)	weight	49 kg			
	height	1.6 m			
	belt compromised interaction				
Neck	age	28,51 yrs			
(AIS 2+, n=2)	weight	86,92 kg			
	height	1.8,1.8 m			
	belt compromised interaction				
Upper extremities	age	[34 to 68 yrs]		23,33,48,56 yrs	25,53 yrs
(AIS 2+, n=14)	weight	[74 to 95 kg]		54,64,76,116 kg	63,91 kg
	height	[1.7 to 1.8 m]		1.6,1.6,1.7,1.7 m	1.7,1.7 m
	belt compromised interaction				

		Male		Female	
		Driver	FSP	Driver	FSP
All occupants	age	[25 to 51 yrs]	[20 to 50 yrs]	[26 to 48 yrs]	[25 to 61 yrs]
(MAIS 0+, n=954)	weight	[70 to 86 kg]	[67 to 85 kg]	[56 to 71 kg]	[55 to 71 kg]
	height	[1.7 to 1.8 m]	[1.7 to 1.8 m]	[1.6 to 1.7 m]	[1.6 to 1.7 m]
Thorax	age	[34 to 65 yrs]	55,63 yrs	[29 to 54 yrs]	[65 to 81 yrs]
(AIS 2+, n=59)	weight	[67 to 89 kg]	68 kg	[56 to 71 kg]	54,57,60,76 kg
	height	[1.7 to 1.8 m]	1.7 m	[1.6 to 1.6 m]	1.5,1.5,1.7 m
	belt compromised interaction	n=3		n=2	n=1
Abdomen	age				31,32 yrs
(AIS 2+, n=2)	weight				60,60 kg
	height				1.5,1.7 m
	belt compromised interaction				n=1
T. & L. Spine	age	18 yrs		74 yrs	23,39 yrs
(AIS 2+, n= 4)	weight	79 kg		67 kg	65,77 kg
	height	1.7 m		1.5 m	1.7,1.8 m
	belt compromised interaction				
Pelvis,Thigh or Knee	age	[39 to 62 yrs]	55 yrs	[25 to 50 yrs]	26,77 yrs
(AIS 2+, n=15)	weight	75,76,79,102 kg		46,57,57,57 kg	79 kg
	height	1.7,1.7,1.8,1.8 m		1.5,1.5,1.6,1.6 m	1.6 m
	belt compromised interaction	n=1			
Lower leg	age	28,38,48,50,59 yrs	55 yrs	28,46,74,81 yrs	67 yrs
(AIS 2+, n=11)	weight	73,79,84,113 kg		49 kg	48 kg
	height	1.7,1.7,1.8,1.9 m		1.5 m	1.5 m
	belt compromised interaction	n=1			

970121

Glazing Effects of Door or Frame Deformations in Crashes, Part 2

Herbert Yudenfriend
My-Lite Corp.

Carl C. Clark
Safety Systems Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT:

The non-uniformity of temper of 27 motor vehicle door windows as measured by the polarization method is presented, illustrated by two photographs of windows between crossed polarizers. The tempered glass fracture characteristics of 11 Geo Prism door windows are illustrated by two video frames of the windows after fracturing by door deformation slowly applied at bumper level. The speed of some moderate sized fragments was measured to exceed 60 km/h, due to strain relief rather than contact by an intruding object. Glazing history is briefly summarized, with our support of the NHTSA conclusion that more than 1300 lives per year could be saved cost effectively by replacing tempered glass by an "advanced glazing."

INTRODUCTION:

In the first paper [1] of this continuing study, 12 front door windows for the 1990 Geo Prism, of tempered glass 3 mm thick, were slowly strained by a 12.7 cm diameter horizontal roller pressing on the outside of the glazing between two horizontal rollers 50 cm apart on the inside of the glazing. After counter-bending the glass to failure, the tempered glass shattered, with the pieces flying primarily in the direction of compression relief by the bending, "inside.". The cracks in the glazing were not in a square array, as occurs approximately when tempered glass is broken without being strained, but were angled, with some cracks not reaching the glass surface on the compression side. The consequence was that pieces with cracks not yet separated, called clusters, were larger and heavier and sharper, many with points, or many as slivers rather than the "cube-like" pieces of the unstrained shattered tempered glass.

In this Part 2, the non-uniformity of temper of glazing for a number of different vehicles, and the conditions of shattering of tempered glass windows when the deformation force was slowly applied at bumper level on the door, were studied, and the safety consequences of the "shatter and fly" characteristics of tempered glass as these affect laceration and ejection injuries are discussed.

MATERIALS AND METHODS:

This Part 2 study has two experimental sections. The first section deals with the variation of temper over the surface of 27 door windows of six make and model year vehicles, as determined by the birefringence method. Glass with a higher compression has a higher refractive index. The rapid cooling of the surfaces of the glass, in making tempered glass, puts the surfaces in greater compression (the glass shrinks), than the interior. Because of this the refractive index of light polarized parallel to the surface is greater than the refractive index of light polarized perpendicular to the surface. That is, the anisotropic compression of the glass produces an anisotropic refractive index or birefringence. The surface compression of the two sides of the glass is not perfectly uniform. As a result, looking at a piece of tempered glazing between crossed polarizers, or even looking through polaroid glasses at the glazing in sunlight reflected off a surface (and hence partially polarized) reveals the glazing to be light or birefringent in some areas (Figures 1 and 2). Indeed, this is the method to distinguish tempered glass, which is birefringent, from annealed glass, which is more nearly isotrophic or non-birefringent. Looking at a fragment of the glass on edge between crossed polarizers, as with a polarizing microscope, reveals the greatest birefringence of the glass, with the magnitude of the birefringence a measure of the compression level or extent of tempering of the surface. A laser beam Grazing Angle Surface Polarimeter (GASP) [2] was used to measure the birefringence along and perpendicular to the surface and so to measure the glass temper without having to examine only the edges of the glass. Three to five points were measured on each window using points that were brightest (most birefringent) and darkest (least birefringent) on the transmission views of the entire glazing between crossed polarizers.

The second experimental section of this study involved successively mounting 11 driver side doors of the 1989 Geo Prism on a frame. The hinges were attached to the frame, but the latch side of the door was not supported. The bottom of the door is supported by a box beam simulating the rocker panel, but the top of the door is supported with a steel

member that could pivot as the back of the door deformed. A rigid bumper was then pressed with a hydraulic ram against the door at a typical bumper height then advanced slowly until the tempered glass door window shattered, with the ram traveling up to 35 cm in more than 20 seconds. Different tests had different bumper horizontal angles with respect to the door (0 to 30 degrees), and different extent of overlap (left edge of the bumper at the front of the door to 30 cm rearward), simulating particularly side impacts with B pillar collapse. Two Super VHS video cameras were used, with shutter speeds of 0.008 seconds, at 30 frames per second. The views of the two cameras of the assembly and scatter patterns were from the top ("vertical view") and from the side ("horizontal view."), allowing parallax correction. High intensity lighting was used. White lines drawn six inches apart provide a scale in the camera views.

RESULTS

Using the GASP device, the variation of compression across the surfaces of 27 new original equipment manufacturer tempered door glasses from six different popular car models (Table I) was measured. The National Automotive Glass Specification (NAGS) catalog [3] number of the windows is given, along with the number of windows measured of each make/model. Also listed are the years that the same window part was utilized for the stated make/model. The highest and lowest temper of measurements within the make/model set of windows, and the highest and lowest temper measured within one window with the lowest temper value within the make/model set, are presented. Also given are other years for makes and models using the same window part number. The significant variation of temper over a single pane of glass, and the variation over different panes of glass, leads to a variation of strength of the glazing and a variation of the size of the particles and clusters when the glazing shatters.

For the door deformation tests, a selected frame of a top and of a side view are presented as Figure 3 and Figure 4, showing the diversity of sizes of the particles and clusters. Following some of the fastest single clusters that were big enough to be seen in successive frames at 30 frames per second allows an estimation of the speed of over 60 km/h, due not to the motion of the contacting bumper but to the abrupt release of the surface energy of the glazing. The smaller particles were moving faster, but a higher camera frame rate will be required to measure their speed. When the glass first shattered, more than half of the glass weight fell down on the outside of the window. In the different tests, from less than 20 % to about 40 % of the glass flew inward, with three tests with very little glass flying inward. We note the many road crashes with shattered tempered glass which have less door and window frame deformation than was required in these slow deformation tests. The significance of shock loading in dynamic crashes needs further study.

It is interesting to note that many of the larger clusters flying toward the inside break apart as they travel even 30 to 60 cm through the air, and so their size will not be judged by looking at particles of glass scattered throughout the vehicle after the crash. In other tests of shattered 6 mm tempered automotive glass, a cluster fragment weighing 62 grams was collected. Note that the head and body are typically less than 30 cm from the near side glazing. If the glazing is accelerated by an intruding object, its speed can be even higher. This combination of the greater weight than 4.25 grams of the uncracked fragment of a shattered tempered glazing specified in Federal Motor Vehicle Safety Standard 205 [4], and the high speeds, and the pointedness and sharpness of the fragments released under strained conditions [1], explains the observed penetration of tempered glass fragments not just into the skin but through the skull and into the brain [5]. In this case, Richard Johnson was found by computer tomography and an operation for removal to have a 6.2 by 2.5 cm triangular piece of tempered glass in the left parietotemporal region of the brain.

DISCUSSION:

In this section we discuss the death and injury consequences, involving laceration and ejection, of tempered glass characteristics, and the alternative glazings that can reduce these deaths and injuries.

A few years ago, the National Accident Sampling System data collection was modified to pay more attention to glazing injuries and their sources. A special request to NHTSA for 1995 data produced the following (rounded off) for occupant lacerations in crashes of light passenger vehicles in tow-away police reported crashes:

windshield contact	159,000
side window contacts	8,800
rear window contacts	903
flying glass	188,000

The lacerations due to flying glass are due to fragments of the shattered tempered glass flying throughout the occupant compartment, causing lacerations without the occupant moving onto the pre-crash glazing areas. Total lacerations were 1.1 million in 1995 crashes. About 35 % of these were due to glass, with the majority of the glass lacerations caused by flying tempered glass.

In most of the side and rear window glass contacts resulting in lacerations, the tempered glass shatters, leading to partial or complete ejections. Ejections also occur if the tempered glass is shattered by crash forces prior to the body reaching the window area. . On the other hand, contact with lacerations with the laminated glazing of the windshield typically involves cracking of the windshield, but with fewer ejections, and the windshield today is rarely opened out by crash forces prior to body contact, allowing ejection without glazing contact..

Ejection data was also requested, from NASS 1995, ratio-weighted. For the known ejections and routes:

	partial	complete
windshield	1,203	728
side windows	14,624	12,149
rear window	1,424	5,593

Partyka [6], correcting for the estimates by the police of partial ejections, low by about 50 %, concluded that the

deaths due to ejection were near 10,000 per year. Malliaris concluded from his studies that 75 % of ejections with death were through glazing areas, and 85 % of the ejections through glazing areas involved the shattering of the glazing by crash deformation forces prior to the occupants reaching the pre-crash glazing surface [7]. About 20 to 25 % of passenger car occupant deaths involve ejection, and about 40 percent of light truck, van, and sport/utility vehicle occupant deaths involve ejection, and this proportion has been almost the same for 15 years [8], while belt restraint use increased from a few percent to over 65 % [7]. Increased belt use over the years has not reduced ejection deaths. Although the belted occupant has a small percentage chance to fully eject, deaths and injuries result from partial ejections in side impacts and rollovers.

Laminated glazing was used for windshields and side and rear glazing from the early 1930's to 1962, with a transition to the use of the less expensive tempered glass for the side and rear glazing from the late 1930's, but particularly in the period of 1957 - 1962. The laminated windshield continued to be used, and was greatly improved in 1966 when the High Penetration Resistant (HPR) windshields, with 0.76 mm of polyvinyl butyral as the plastic interlayer, replaced the earlier windshields with 0.38 mm of PVB as the plastic interlayer [9], greatly reducing deaths with windshield ejection, and significantly reducing windshield lacerations.

The change from laminated to tempered side and rear windows (by 1962) and the change to HPR laminated windshields (1966) both occurred before the creation by Congress of the National Highway Safety Bureau (November 1966). This agency was created to require the vehicle manufacturers to provide certain specified minimum levels of safety in their products. The initial regulations primarily put the force of law behind the various industrial standards, with Federal Motor Vehicle Safety Standard 205 [4] adopting the American Standards Association (ASA) Z-26.1 Standard. This standard has remained essentially unchanged with regard to the absence of direct tests for laceration, head impact loads and ejection. Indeed, the initial drop tests of head penetration using a 22 pound headform and large glass pieces (2 x 3 feet) was replaced by drop tests with a 5 pound steel ball impacting on small (one foot square) glass pieces, based on correlation of results with tempered glass [8]. The correlation is less adequate with laminated (including glass plastic) glazings. The 22 pound headform provided some indication of ejection resistance, whereas the 5 pound steel ball test is primarily a test of glazing fracture characteristics.

Glass-plastic glazing, a form of advanced glazing, was invented in the United States in the 1960s but first commercialized in production motor vehicles in France [10 - 12]. This glazing is produced in two versions. The first is one or more layers of glass laminated with one or more layers of plastic, with one layer of plastic on the inside surface of the glazing in the vehicle. The most notable application of this system was the anti-laceration windshield, for which a plastic "layer" (which in different formulations may have differing plastics in multiple layers) was put on the inboard surface of an HPR windshield [13], used on a number of General Motors production vehicles between 1984 and 1987 [14]. The second type consists of an inboard plastic "layer" generally on a single layer of tempered glass [15]. The inboard plastic layer provides a smooth surface for body contact, and so essentially eliminates lacerations. With the two-ply tempered glass-plastic glazing, the inboard plastic layer will prevent shattered tempered glass pieces from flying into motor vehicle occupants, and will greatly reduce contact lacerations, and will provide a "safety net" over the window area with the potential to almost eliminate the ejection of children and significantly reduce the ejection of adults. This is particularly true when the plastic layer is bonded to the window frame or supported in an encapsulation trapped within the window channels for a moveable window, initially the "T-edge" concept [16, 17], with the glazing able to prevent ejection of an 18 kg headform moving at 32 km/h.

The NHTSA is responding to the injury reduction possibilities of the "advanced glazings" by extending the earlier research on glazing [18 - 22], with indications that more than 1300 lives per year could be save, in a cost-effective manner, with the replacement of tempered glass by advanced glazing. A regulation for the change has not yet been issued.

A number of European and Japanese manufacturers have or are about to change from the use of tempered glass to the use of laminated glazing in some of their side and rear and roof windows, including BMW, Audi (A-8 model), Volvo (1998 station wagon), Toyota, Nissan, and Honda. These changes are presented as improving noise or security characteristics, but they also improve occupant safety, and we hope to see this aspect advertised and prevailing throughout the industry. We note with great interest the Swedish effort to set a "zero goal" for motor vehicle safety [23], and recommend to them that glazing safety be part of this effort. We plan to continue our studies of the deformation conditions that shatter tempered glass, and their effect on glass-plastic glazing. We also plan to examine the change of stress to shatter tempered glass as it ages, particularly with the accumulating scratches and other surface damage with use, as reported earlier [24].

SUMMARY AND CONCLUSIONS

We have presented measurements of the non-uniformity of temper, as determined by the polarization method, of 27 door tempered glass windows. We have presented tempered glass fracture characteristics when 11 car doors were deformed at bumper height, with small fragments and large clusters flying into the occupant compartment, some with speeds above 60 km/h due to strain relief and not due to intruding object contact. We review the laceration injuries and ejection deaths with present vehicles with tempered glass side and rear windows, and means to reduce these injuries and deaths by laminated or glass-plastic glazing.

We conclude that tempered glass has dangerous characteristics of non-uniformity of temper, shattering under strain conditions into large and/or sharp fragments which

"fly" into the occupant compartment, and no ejection reduction capability once cracked.

ACKNOWLEDGEMENT:

We thank Henry Chamberlain and the staff of Automotive Glass Engineering, Kansas City, for assistance in the experimental work of this study.

REFERENCES:

1. Herbert Yudenfriend (My-Lite Corp.), Glazing Effects of Door or Frame Deformations in Crashes, SAE 960105, 1996.

2. Alex S. Redner, Measuring Stress in Heat Treated Glass, Glass Magazine, pages 56 - 61, March, 1991.

3. National Automotive Glazing Specifications, Inc. (NAGS), NAGS Catalog, 1995.

4. National Highway Traffic Safety Administration, Federal Motor Vehicle Safety Standard 205, Glazing Materials, United States Code of Federal Regulations, Title 49, Part 571.205, U.S. Government Printing Office, Washington, DC (updated annually).

5. Johnson v. General Motors Corporation, Case No. 95C - 1572, Circuit Court, Davidson County, Tennessee.

6. Susan Partyka, (NHTSA), Occupant Ejections from Light Passenger Vehicles, NHTSA Staff Paper, 1991. Available in NHSA Docket 88-06-N03, Document 044.

7. Data Link, Inc. (including A.C. Malliaris), The Ejection Experience of Light Vehicle Occupants, A Report Prepared for the Crashworthiness Staffs of the National Highway Traffic Safety Administration, 1992.

8. National Highway Traffic Safety Administration, Fatal Accident Reporting System (FARS), Annual, and the National Accident Sampling System (NASS), Annual..

9. J.C. Widman (Ford Motor Company), Recent Developments in Penetration Resistance of Windshield Glass, SAE 650474, 1965.

10. Otto Jandeleit (Vereinigte Glaswerke, West Germany) and Roger Orain
(Saint-Gobain Industries, France), Performance of a Non-Lacerative Windshield in Case of Accident and Under Environmental Conditions, SAE 770246, 1977.

11. Roger Orain (Saint-Gobain Vitrage, France), Letter to Adele Spielberger (NHTSA) re the production of Securiflex and Securiflex-2 glass plastic glazings, September 2, 1987.

12. Jean-Louis Bravet (Saint-Gobain Vitrage) and Koji Kurita (Asahi Glass Co.): Bilayer Glazing: A Concept with Improved Solutions for Tomorrow's Automobiles, SAE 950051, 1995.

13. Thomas C. Walbridge (Executive Vice President, Marketing and Sales, Libby- Owens-Ford Co.), Anti-Lacerative Windshields, An Overview of Safety Development Technologies, U.S. Glass, Metal, and Glazing, pages 64 - 67, September-October, 1988.

14. Robert Rogers (GM Director of Automotive Safety Engineering), Letter to Adele Spielberger (NHTSA) re the production of "Inner Shield" glass plastic glazing, and serious interest in the two ply version, June 15, 1987.

15. James E. Greene (Calspan Corporation), Occupant Survivability in Lateral Collisions. volume 1 (of 2; volume 2 contains test data), NHTSA Report DOT HS 801 801, January, 1976.

16. Carl C. Clark and Peter Sursi, (NHTSA), The Ejection Reduction Possibilities of Glass-Plastic Glazing, SAE 840390, February, 1984.

17. Carl C. Clark, (NHTSA), and Peter Sursi, (Automotive Safety Testing, Inc.), Rollover Crash and Laboratory Tests of Ejection Reduction by Glass-Plastic Side Windows and Windshields, SAE 890216, February, 1989.

18. Clarke B. Harper and Susan C. Partyka, (NHTSA), A Review of Motor Vehicle Glazing-Related Ejection Injuries, SAE 930740, 1993.

19. Clarke Harper and Stephen Summers (Co-Chairs of the NHTSA Advanced Glazing Research Team), and Paul Atelsek, Stephen Duffy, Margaret Gill, Lillvian Jones, John Lee, Linda McCray, Dinesh Sharma, Robert Sherrer, and Donald Wilke (other members of the Team), with assistance from Susan Partyka and John Winnicki, Ejection Mitigation Using Advanced Glazing, A Status Report, NHTSA, November, 1995.

20. William S. Ward (Management Engineering Associates), Alternative Glazing Cost Study, Report No. MEA 95-04502, Final Report, NHTSA Contract DTNH22095-C-02201, September, 1995. (Available from Mr. Clarke Harper, NHTSA Office of Rulemaking, and NHTSA Docket No. 95-41-GR-001).

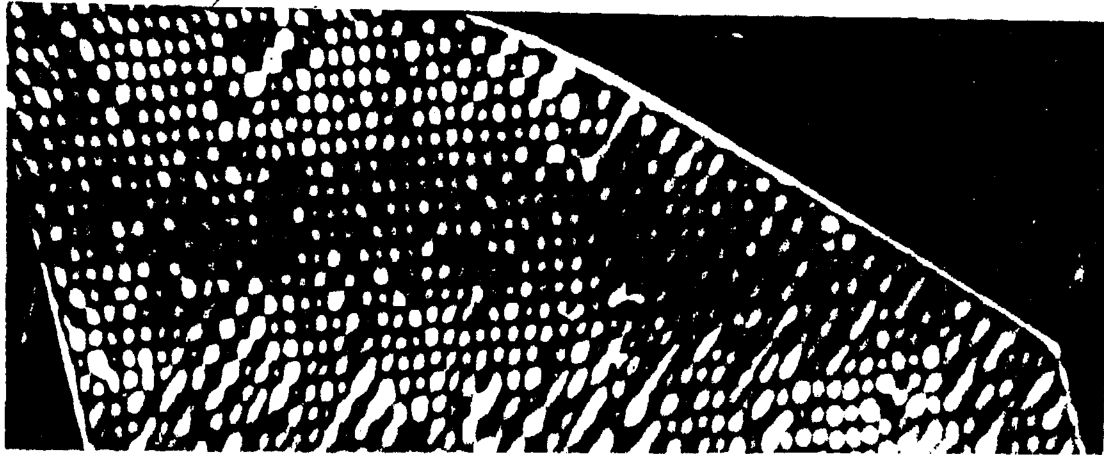
21. Stephen Sommers, Glen C. Rains, and Donald Wilke (National Highway Traffic Safety Administration), Current Research in Rollover and Occupant Retention, NHTSA Paper 96-S5-O-01, The International Enhanced Safety of Vehicles Conference, Australia, 1996.

22. John Winnicki (NHTSA), Estimating the Injury-Reducing Benefits of Ejection Mitigating Glazing, NHTSA Technical Report DOT HS 808 369, February, 1996.

23. Claes Tingvall and Anders Lie (Swedish National Road Administration, Chalmers University of Technology Department of Injury Prevention, and Karolinska Institute Division of Neurotrauma), The 1996 "Bertil Aldman Award" Lecture: The Implications of the Zero Vision on Biomechanical Research, Proceedings of the International IRCOBI Conference on the Biomechanics of Impact, Dublin, September, 1996. International Research Council on the Biomechanics of Impact Secretariat, 109 Avenue Salvador Allende, 69500 Bron, France.

24. Herbert Yudenfriend (Stott-Hecht Glass Works), The Issue and the Challenge, Glass Digest, page 44 and 45 and 2 more, January, 1961.

###



Figures 1 and 2. Reduced size views, taken between crossed polarizers, of two different door windows, showing the non-uniformity of bi-refringence - and hence of temper.

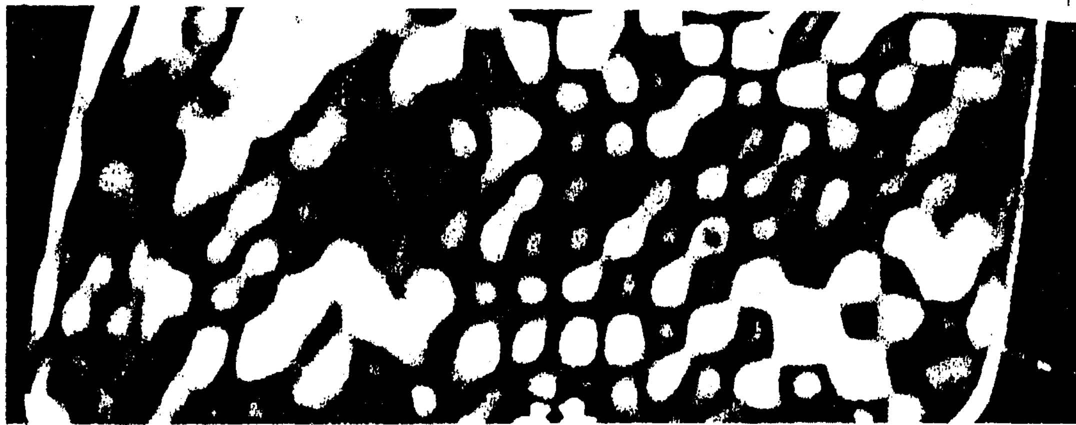


Table I: Variation of compression level of 27 motor vehicle door windows, identified by their National Automotive Glass Specification catalog numbers, with the years and make/models using this glazing. (Note: 138,000 kPa = 20,000 psi).

NAGS Catalog No.	No. Tested	Years	Make/Model	Total set, Highest	kPa Lowest	One window, Highest	kPa Lowest
FD3821	6	'89-92	Geo Prism	146,000	90,000	115,000	90,000
DQ5824	6	'80-96	Plymouth Voyager	162,000	77,000	108,000	77,000
		'88-96	Dodge Caravan				
FD4131	3	'90-91	Honda Accord	129,000	51,000	125,000	51,000
DD6967	3	'86-95	Ford Taurus	133,000	74,000	119,000	74,000
		'86-95	Mercury Sable				
DD7833	6	'91-96	Chevy Caprice	161,000	97,000	132,000	97,000
		'92-95	Buick Roadmaster				
		'93-95	Chev. Impala SS				
		'91-92	Olds Custom Cruiser				
DD7478	3	'92	Chev. Blazer	113,000	77,000	100,000	77,000
		'88-92	Chev. Pickup Truck				
		'92	Chev. Suburban				
		'88-92	GMC Pickup Truck				

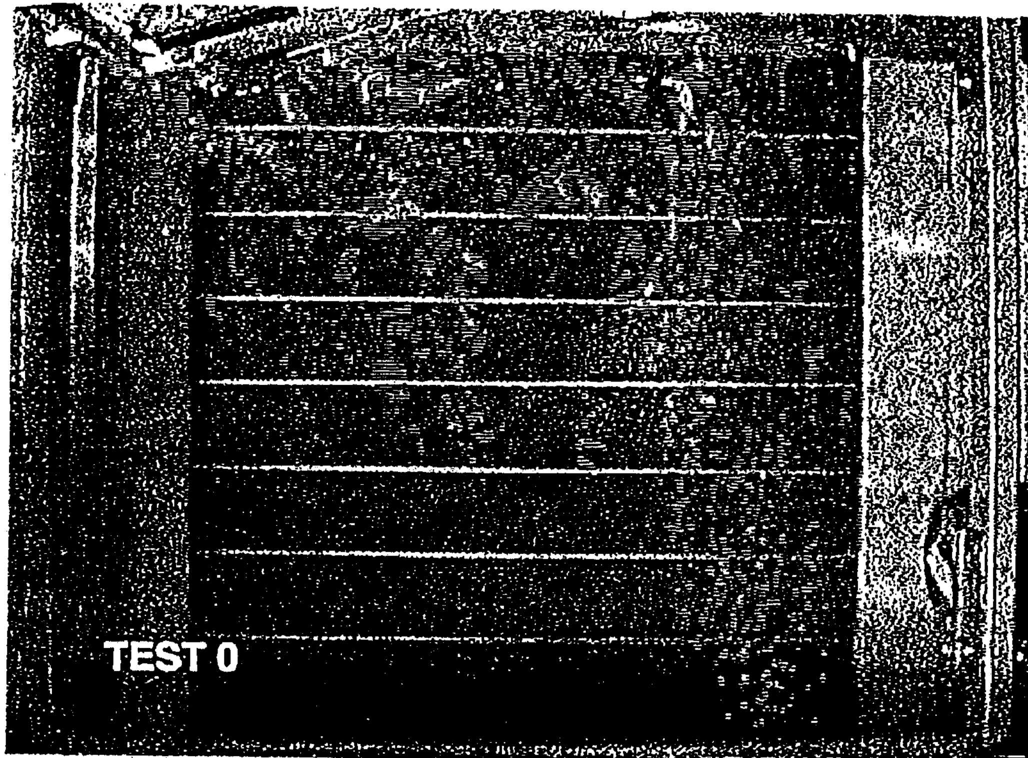
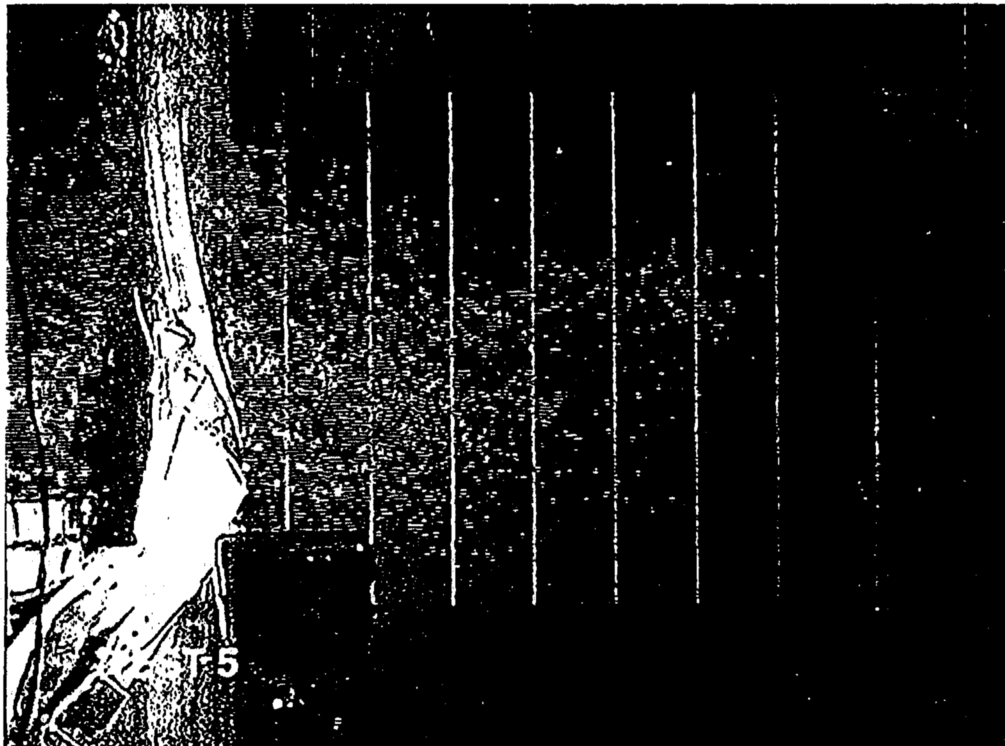


Figure 3: View looking down ("vertical view") of the tempered glass as it shatters and flies, 11 frames after the glass cracks due to door deformation from side loading at bumper level. The door and window are at the top of the frame. Test 0V, frame 11.

Figure 4: View looking from the side ("horizontal view") of the tempered glass as it shatters and flies, 7 frames after the glass cracks due to door deformation from side loading at bumper level. The door and window are at the left of the frame, with the bumper slowly deforming the door from the left, and the "occupant compartment" to the right. Test 5H, frame 7.



970122

Crashworthiness and Compatibility of Low Mass Vehicles in Collisions

P. Frei, R. Kaeser, R. Hafner, M. Schmid, A. Dragan, L. Wingeier, M. H. Muser, P. F. Niederer, and F. H. Walz
Working Group on Accident Mechanics at the Universities Zürich

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Reducing the weight of vehicles could be a strong means of reducing fuel consumption in urban traffic. Published accident and injury statistics however show an inverse correlation of vehicle mass against injury severity in car to car collisions, above all in head-on collisions. This inverse correlation is in part caused by current crash test standards, where compatibility in collisions between cars of different size and weight is not a requirement. Compatibility in frontal collisions demands for significantly different deceleration-time curves in rigid barrier impacts for cars with different weight. Low mass vehicles (LMV) must meet compatibility criteria to comply with current injury criteria in real car to car collisions. Cars designed according to compatibility criteria can change future accident and injury statistics in a way that injury severity in LMVs can be reduced significantly.

This study shows, that head-on collisions at closing speeds of up to 100 km/h with a mass ratio of 1:2 of the collision opponents can be survived with occupant loads well below the limits of actual injury criteria (FMVSS 208), if compatibility requirements are met with regard to car structure as well as restraint systems.

INTRODUCTION

FOCUS - The research work described in this paper is focusing on frontal collisions between vehicles of different mass and size and the currently observed lack of compatibility therein. This lack of compatibility is considered to be the main reason for the infavourable appearance of small vehicles in accident statistics [1].

GENERAL - The reasons for the higher injury risk for occupants of small vehicles can be found in the low mass which results in a higher total change of velocity in car to car collisions, and, in comparison to heavier cars, in a usually lower stiffness of the crush zone and the cabin structure. Furthermore, the design of the car structure and the restraint systems are - according to current FMVSS and European standards - still being optimised for a head-on collision of a car against one of its kind (impact on rigid barrier), and not against the statistically most probable counterpart. These compatibility issues have been addressed by various authors in the past 25 years ([2], [3]).

CONSEQUENCES FOR THE LMV - The success of low mass vehicles, i.e. small vehicles (2.5 - 3.0-m overall length) with an energy-saving drivetrain, depends to a high degree on the level of passive safety they will offer. The safety deficits of the small cars mentioned above become even more apparent in LMVs with a curb weight less than 600 kg as compared to 800 - 1000 kg of current subcompact cars.

The basic principles for the structural design of such vehicles, as well as first results of crash tests, have been published earlier by our working group [4]. A curb weight for the LMV of 600 - 650 kg appears to be feasible. In frontal collisions, such a LMV should be compatible with 70 - 80% of the currently circulating cars. In Europe, the mass of collision opponents to be considered for the LMV (i.e. the European 'compact' class) ranges between 1300 - 1400 kg, which leads to a design mass ratio of about 1:2. In a frontal car to car collision with 100% overlap and 100 km/h closing speed, the LMV is subjected to a Δv of 20 m/s and a mean car deceleration level of 40 - 50 g. A highly optimised restraint system is needed to reach the same degree of occupant protection as encountered in conventional cars.

The results of mathematical modeling and sled tests of such an optimised restraint system have been presented at the 1996 STAPP conference [5]. They clearly indicate the feasibility of the procedure, in that the envisaged protection criteria could be met with ample margins.

Since the interior of LMV's dimensions are given by ergonomical constraints, the demand for a small overall length leads to a shorter, more densely packed frontal deformation zone. This lack of adequate crush space and the need to avoid car deceleration peaks higher than 60 g (approx.) demand for the frontal structure of the LMV being designed such that it generates an

almost ideal rectangular curve for deceleration versus time in frontal collisions (cf. Figure 1).

RESEARCH WORK - In 1991 our group started to design and test prototypes and experimental vehicles with respect to compatibility criteria. Crash tests at speeds up to 71 km/h were performed against rigid walls, deformable barriers and heavier cars to analyse the behaviour of car structures and restraint systems. Based on these preliminary results, the experimental LMV CRATCH for urban areas has been designed and built. Mathematical modelling, static and dynamic component tests as well as sled tests were used for the development of the restraint systems and of the car structure. The experimental LMV CRATCH has been successfully tested in a head-on car vs. car collision.

COLLISION MECHANICS

GENERAL - A small, light car will always experience a higher deceleration level in a real car versus car collision compared to its heavier opponent, as the deceleration for each of the colliding cars can be calculated with a simplified model from the crush load at the interface of the two cars. Both cars experience the same crush load vs. time curve with this simplified model.

$$(\text{deceleration}) = (\text{crush load}) / (\text{car mass})$$

Furthermore, this type of frontal collision will subject the lighter car to a higher total change of velocity. The velocity change of both cars can be calculated with the law of conservation of momentum. In real collisions the Δv is 5 - 10% above the theoretical value due to residual elasticity of the deformed structures.

$$\Delta v_1 = m_2 * (v_1 + v_2) / (m_1 + m_2)$$

$$\Delta v_2 = m_1 * (v_1 + v_2) / (m_1 + m_2)$$

As this study shows, the main reason for the higher injury risk for the occupants of currently circulating small cars in frontal collisions is not intrinsic in these two physical constraints. Instead, it is given by the lack of compatibility in frontal collisions among cars of different size and weight.

The restraint system of any car will only be able to deploy its effectiveness if the passenger compartment does not collapse in a collision. Intrusion of structural components into the cabin and deformation of the passenger compartment should be small. A extremely important parameter for the development of the restraint system is the mean deceleration level of the car cabin with its restraint system interface points and surfaces.

Furthermore, the design of the car structure and the restraint systems, are - according to current FMVSS and European standards - still being optimised for a collision of a car against one of its kind, and not against the statistically most likely counterpart. (A collision of a car against a rigid wall simulates, in an idealised way, a collision against the 'mirror image' of that car). Thus, the frontal stiffness is chosen such that most of the crush zone is deformed under the test conditions, resulting in the lowest mean deceleration level of the car and therefore normally in the lowest dummy loadings. In rigid wall tests at speeds of 50 km/h, mean decelerations of current cars range between 15 g and 25 g. Larger cars normally exhibit mean decelerations closer to 15 g, subcompact cars with short crush zones on the other hand show mean decelerations up to 25 g in collisions. A car with a high mass will thus be lead to exhibit a higher stiffness of the frontal structure, since the required force to induce a deformation at a given deceleration level is proportional to the mass.

A low mass vehicle, designed according to current crash test standards, will be doubly disadvantaged in case of a collision with a heavier vehicle. Not only will it experience a higher Δv due to the mass ratio, but its frontal deformation space, which is relatively soft (according to the reasons mentioned above), will be crushed before significant deformation of the heavier car even starts.

In conclusion - the frontal structural stiffness of a low mass vehicle must be at least equal or slightly *higher* than the stiffness of its heavier counterpart [6].

COMPATIBILITY - Structural compatibility in collisions means that the different cars should deform at the same load level, yielding a length of the deformation zone proportional to the mass of the car. This even helps to design a short and light LMV, e.g. with a total length of 2.5 - 3.0 m. Analysis of advanced restraint systems indicate, that such an 'universal' deformation load level should be defined in a way that mean deceleration levels of LMV's do not exceed 40 - 50 g. The maximum load level in a frontal collision at a given speed is experienced with 100% overlap, the load decreases with smaller overlap ratios. The length of the crushable zone is defined by the amount of energy that has to be absorbed. Since frontal crash tests against a rigid barrier are performed at the same speed with 100% and 50% overlap, the maximum required deformation length is determined by the 50% overlap condition. Under all conceivable conditions, the LMV's crush zone must be capable of absorbing the LMV's own kinetic energy. In a collision against a heavier car, some of the LMV's kinetic energy might also be dissipated through deformation of the heavier car if the front stiffness of the LMV is designed according to the considerations mentioned above. Such a concept is reasonable, since the heavier car experiences a considerably smaller change in velocity in this type of collision than in a test against a rigid barrier. Earlier tests with an ultrastiff lightweight structure crashed against an Audi 100 have confirmed this concept [4].

Figure 1 shows, schematically, deceleration pulses during an impact on a rigid barrier with an initial velocity of 15.5 m/s for a LMV designed according to compatibility criteria in comparison to a typical compact car. Deformations in this example would be in the range of 300 mm for the LMV vs. 800 mm for the Compact Car.

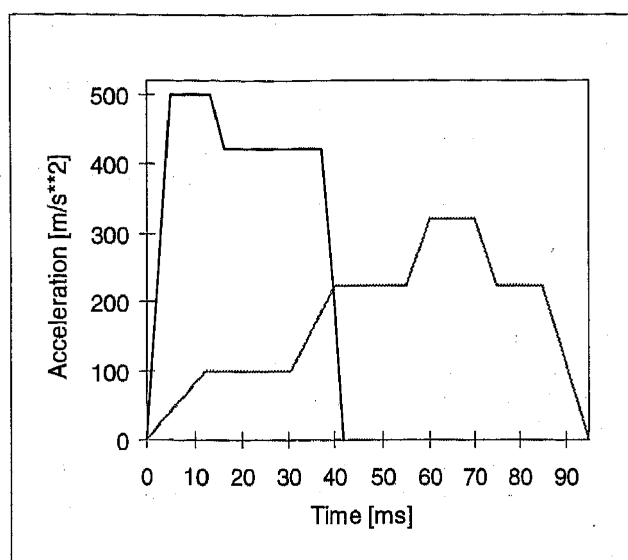


Figure 1: Schematic representation of the deceleration pulses of a typical Compact Car (gray) and a compatible LMV (black) against a rigid barrier at 15.5 m/s (56 km/h).

INTERIOR GEOMETRY

Low mass vehicles are intended primarily for use in urban areas and short ranges. A rather upright seating position is adept to this situation since it offers a better overview of the traffic situation for the driver and thus enhances active safety (cf. Figure 2). Such a seating position results in somewhat shorter ride-down distances for the upper body parts. On the other hand, given an adequate restraint system, the upright position will lead to less forward rotation of the upper body parts, thus reducing unnecessary decelerations especially of the head. Once the seating position is chosen, the position and angle of the steering wheel as well as the position of the feet and the knee bolster are determined by ergonomical considerations.

RESTRAINT SYSTEM CONCEPT

The restraint system of the experimental LMV CRATCH has been presented at the 1996 STAPP conference [5]. Here a short summary.

Restraint system simulations for the LMV with the pulse in Figure 1 and a Δv of 20 m/s clearly show that additional ride-down space in the car interior has to be provided, thereby compensating for the diminished crush zone. The effective use of the ride-down space must be ascertained, e.g. the restraint systems must become active as soon as possible and the components must deform ideally at a constant load level. Figure 2 illustrates the interior geometry and the restraint system components of the experimental low mass vehicle CRATCH.

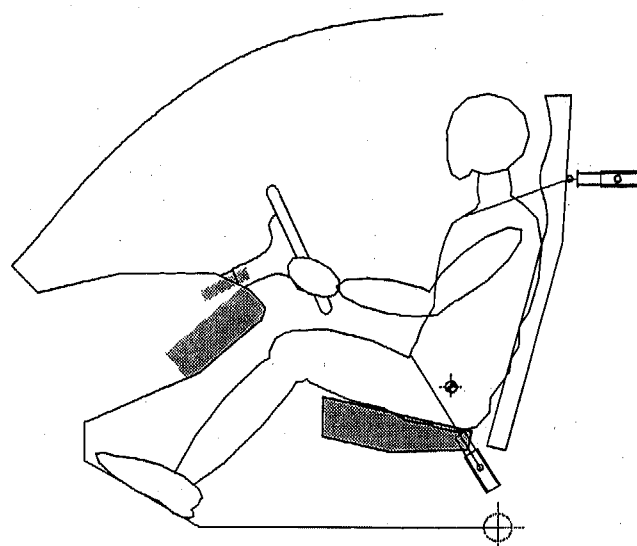


Figure 2: Interior geometry of the LMV. Additional energy absorbing components are marked in light gray.

Low mass vehicles are currently produced only in small quantities. Therefore, it is essential that generally available components are implemented in their restraint system. However, such components need adaptation to the specific LMV situation. The following components are integrated in the experimental LMV CRATCH.

AIRBAG - A full size airbag from a German midsize car is used, the exhaust orifices and the trigger time have been adapted.

STEERING COLUMN - A deformable retracting steering column has been developed, that must absorb energy and enlarge the ride down distance by 100 mm.

SEAT PAN - A hard, crushable foam material with a thin soft cushion on top is used. This provides better coupling of the pelvis and higher energy absorption.

KNEE BOLSTER - A bolster with an adequate thickness (~ 100 mm) and compressive strength is used as an energy absorbing component. The initial distance between knee and foam (50 - 100 mm) is a crucial parameter which should be minimised as far as comfort requirements permit.

BELT SYSTEM - The use of a belt system is essential if occupant protection criteria are to be met in a

collision as described above. A 3-point belt system with a stiff webbing (4 - 8 %) and belt load limiters with excessive belt extraction at the design load has been developed to provide the required ride down distance for the occupants and optimised energy absorption properties.

ABSORPTION OF COLLISION ENERGY

CONVENTIONAL CARS - A major part of the kinetic energy of conventional cars in frontal collisions is not absorbed through deformation of the front structure, but instead through direct contact of the wide, heavy engines and accessories and also the front wheels and suspension with the opponent or an obstacle. Deformation of engine and wheel suspensions as well as compression of these components contribute to energy absorption. This arrangement provides an important load path for various overlap conditions through engine and wheels into the cabin floor structure. The cabin floor structure usually absorbs energy through deformation, thereby allowing drivetrain and chassis to intrude into the cabin. Additionally, the very stiff and heavy elements of the drivetrain create a dynamic stiffness against lateral load components in frontal - side collisions.

LOW MASS VEHICLES - LMV's have small, light drive trains. The front structure must be capable to absorb most of the collision energy including lateral load components, since deformation of the cabin and intrusion of components into it should be avoided. An ideal design of a deformable front structure would allow to distribute energy absorption to all its structural elements even for a 50% overlap situation, thereby substituting the part of a heavier car's engine.

STRUCTURAL CONCEPT OF THE LMV CRATCH

GENERAL FUNCTION - A stiff transversal beam right behind the front bumper over the full width of the experimental LMV CRATCH can react various load conditions (point load or distributed load) and overlap situations in frontal and frontal-side collisions. The beam is located at the average bumper height of 440 mm above ground with a width of 120 mm. The beam transfers the load onto the deformation elements of the front structure. A deformation element inside each front wheel bay was selected chosen out of a selection of concepts (cf. figure 3). The two deformation elements are implemented as curved longitudinal beams. The cabin structure is stiff, it does not deform until the deformation elements are fully crushed. The structural interface between cabin and deformation elements is a second stiff transversal beam, connecting the two A-pillars in front of the floorpan.

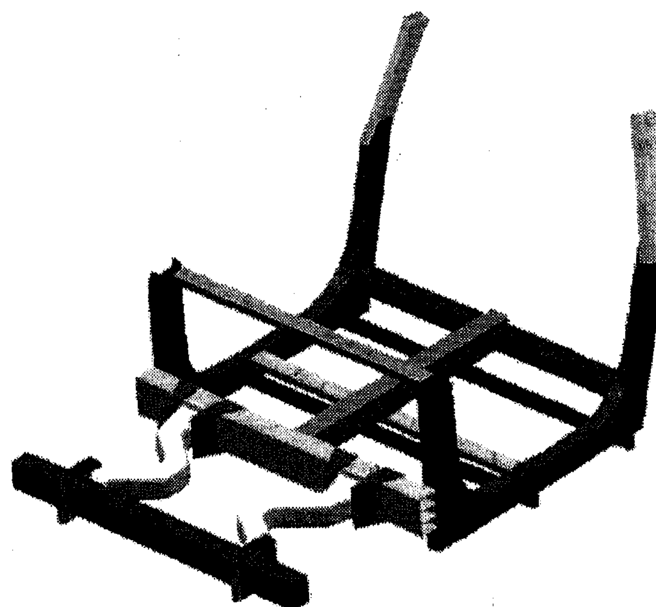


Figure 3: The CRATCH spaceframe

The deformation elements are thin-walled rectangular tubes. They are attached rigidly to the 'bumper' beam and the 'A-pillar' transversal beam, all together forming a stiff frame. This frame is designed to transfer longitudinal as well as transversal loads. In a 100% overlap situation, both deformation elements absorb energy through plastic bending and folding of the tube cross-section at three locations each (cf. figure 4), as opposed to a 50% overlap situation, where only one knee element bends at three locations and the other element bends at one location (cf. figure 5). This means that for 50% overlap, the near knee element accounts for 80 - 90 % and the far one for 10 - 20 % of the collision energy absorbed. The length of the deformation elements is determined by the average crush load and the amount of energy to be absorbed in the various collision situations.

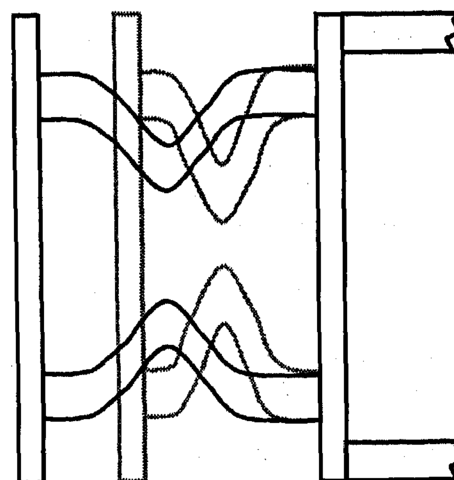


Figure 4: Deformation of front structure (100 % overlap)

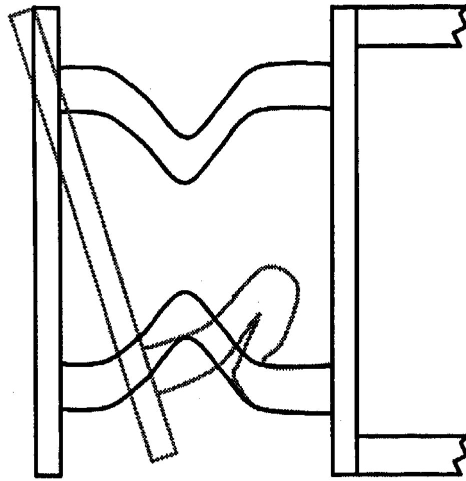


Figure 5: Deformation of front structure (50 % overlap)

TYPE OF STRUCTURE AND MATERIAL - For realization of the above described structural functions of an experimental LMV the design of a spaceframe was chosen. Most elements are tubes of rectangular cross section. Since buckling, rather than yield and rupture of the material is the failure criterion, aluminum was selected in place of steel. Using aluminum also helps to reduce the weight of this construction. The design is based upon single cell tubes without integrated stringers in the walls. The tubes are manufactured by welding U-profiles bent out of sheet metal. Most nodes of the spaceframe are welded, some special nodes had to be riveted.

DESIGN DATA OF THE STRUCTURE

The experimental LMV CRATCH is designed according to the following design requirements:

- cabin for 2 occupants plus luggage
- length 3.0 m
- width 1.65 m
- height 1.7 m
- space for a hybrid drivetrain
- space for 250 kg batteries
- max. curb weight 650 kg
- design frontal crash pulse for 50% and 100% overlap according to figure 6
- max. deceleration level for the cabin 60 g.
- collision angles maximum +20° and -20° to the longitudinal axis
- Stiff cabin structure, structural elements arranged such, that drivetrain or chassis components, bumpers and obstacles cannot intrude into the cabin in frontal and side collisions

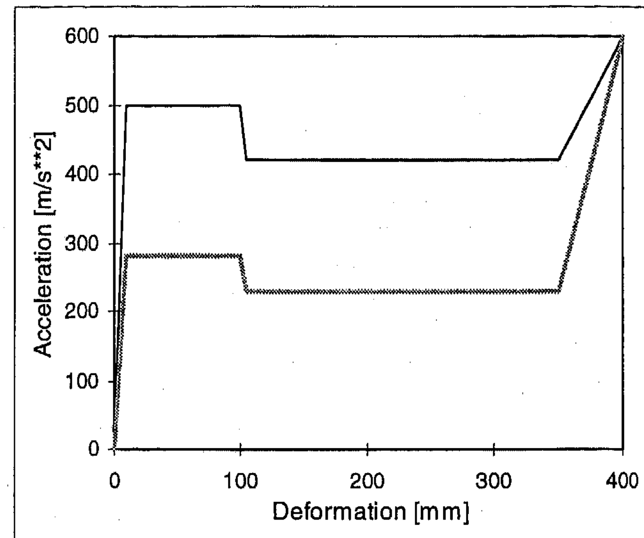


Figure 6: Frontal crash design pulses for the experimental LMV CRATCH, 100% (black) and 25% (gray) overlap against a rigid barrier. These 25% correspond to 50% overlap vs. a deformable barrier due to the nature of this structure.

ANALYSIS & OPTIMISATION OF THE STRUCTURE

ANALYSIS OF THE STRUCTURE - The static strength of the deformation frame was analysed using a simple FE model composed of beam elements and shear panels. The frame was analised for plastic deformations without folding of the cross sections. The load vs. deformation curve does not change for lateral load components up to a collision angle of 20°. As a result,

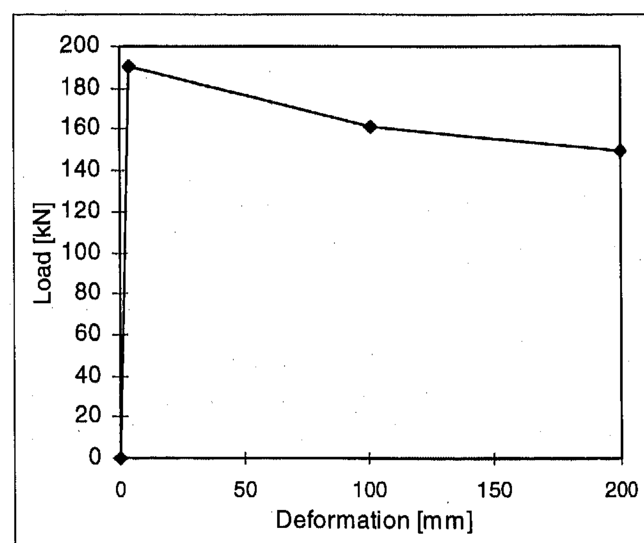


Figure 7: Theoretical crush load, 50% overlap, collision angle 0° - 20°

the crush zone should not be sensitive to lateral loads, e.g. as imposed in oblique impacts (cf. Figure 7).

STATIC TESTS - A test fixture to crush one deformation element according to the condition of a frontal collision with 100% overlap was set up. Four static tests were performed to develop a element with the specified load vs. deformation curve. The dynamic deformation curve was estimated to be 10% higher than the static curve (cf. figure 8).

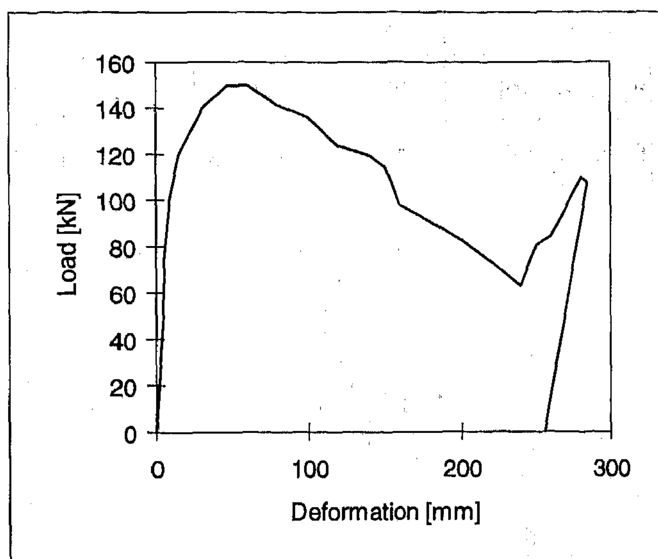


Figure 8: Static crush load of one deformation element, corresponding to a 100 % overlap impact. Load vector parallel to the longitudinal axis.

DYNAMIC TESTS - A preliminary dynamic test of the CRATCH spaceframe structure (without the costly hand-made body and without occupant surrogates and restraint systems) was deemed necessary in order to ascertain predictions and estimates of the dynamic behaviour of the two car structures. Such a test was performed in cooperation with the European car manufacturer Renault. The spaceframe of the experimental LMV CRATCH, equipped with hybrid drivetrain, suspensions and wheels, was tested against a production car (Renault LAGUNA). Ballast was added in order to match the design weight. Both vehicles collided with an initial velocity of 50 km/h, at 50 % overlap and a collision angle of 0°. Following the evaluation of this preliminary test, some modifications at the LMV were necessary. The crash pulse of the LMV was 15 % lower than the estimated pulse and the energy absorption was 25% lower than the kinetic energy of the CRATCH. As predicted, there was no intrusion of deformed components into the floorpan of neither car, and the total deformation of both occupant compartments was extremely small.

VERIFICATION OF THE THEORETICAL FINDINGS

Due to limited resources only one fully equipped experimental LMV CRATCH could be manufactured, i.e. only one frontal collision situation CRATCH vs. Renault LAGUNA could be tested. Therefore, it seemed adequate to use the estimated design pulse for a head-on collision with 100% overlap and a Δv of 20 m/s for the development of the restraint system and the subsequent sled tests, because this is the worst case for the restraint system. A frontal collision with 50 % overlap at 0° collision angle and a Δv of 20 m/s on the other hand is considered to be the critical load case for the crush zone, because only 50 - 60 % of the deformable structure is crushed and absorbs energy. This latter load case yields a lower and longer crash pulse but it shows the maximum deformations with a possible hardening of the structure at the end of the crushable zone, and an eventual intrusion of components into the foot space of the driver. Accordingly, this load case was selected to be tested in the final collision test CRATCH vs. LAGUNA.

FINAL COMPATIBILITY CRASH TEST

GENERAL - The French car manufacturer Renault was a highly welcome partner in the project, as this car manufacturer has a long tradition in offering a model range of largely different size and weight. Compatibility in collisions, at the very least among the currently circulating fleet of their own make, is a central issue for this manufacturer. Therefore, the LAGUNA already exhibits an adequate stiffness of the crush zone considering frontal collisions against small cars. Moreover, in terms of mass and exterior dimensions, the LAGUNA is representative for the current European 'compact' class cars.

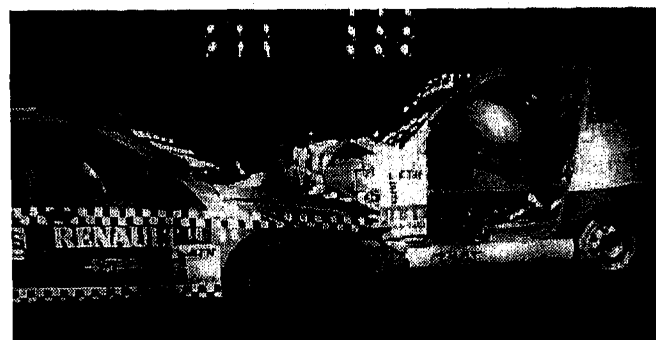


Figure 9: The collision opponents

The driver and passenger positions of the CRATCH as well as the driver position of the LAGUNA were occupied by instrumented 50th percentile Hybrid III dummies. A uninstrumented Hybrid II dummy was used for the passenger position of the LAGUNA, since occupant loads were expected to be well below actual protection criteria.

Table 1: Test parameters

		CRATCH	LAGUNA
Length width (without mirror) Height	[m]	3,05 1,65 1,7	4,508 1,752 1,433
Empty weight incl. measuring equipment, no fuel	[kg]	671,7	1356,6
Driver & passenger dummies	[kg]	160,3	153,4
Collision weight	[kg]	832	1510
Collision speed	[km/h]	50,5	50,5
Overlap (0° collision angle)	[%]	50	46,8
theoretical Δv (w/o rotation and elastic rebound)	[km/h]	65,12	35,88

The test was performed successfully. The high center of gravity of CRATCH caused a remarkable rotation about the lateral axis, i.e the rear bumper was lifted by 0,5 m approx. Both cars rotated about the vertical axis, the CRATCH by 140°, the LAGUNA by 80°.

As predicted, the CRATCH cabin showed no measurable deformation and no intrusion. The left deformation element was totally crushed and, after full deformation had occurred, the rear folding joint broke. Future improvements of the deformation elements should incorporate the elimination of such ruptures.

The LAGUNA cabin showed very small deformation and intrusion into the driver's floorpan (5 - 20 mm).

Table 2: Energy balance

		CRATCH	LAGUNA
measured Δv	[km/h]	71,6	32.76
kinetic energy	[kJ]	81,9	148,6
absorbed energy	[kJ]	65	140

The energy absorption capability for this type of collision of the CRATCH has been exhausted in the test. Further investigation is needed with respect to the energy absorption in other collision situations, e.g. against rigid barriers.

Table 3: Occupant loads

	CRATCH		LAGUNA	limit
	driver	pass.	driver	
HIC_{36}	632	546	257	1000
a_{3ms} head [g]	58,4	62,6	38	80
a_{3ms} chest [g]	39,7	33,3	35	60
F femur l/r [kN]	2,1 / 1.6	1,6 / 0.3	n.m	10.1

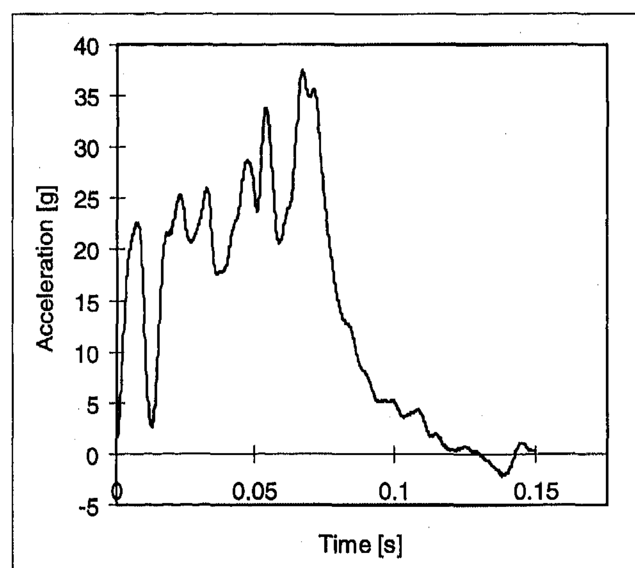


Figure 10: CRATCH - deceleration of car cabin

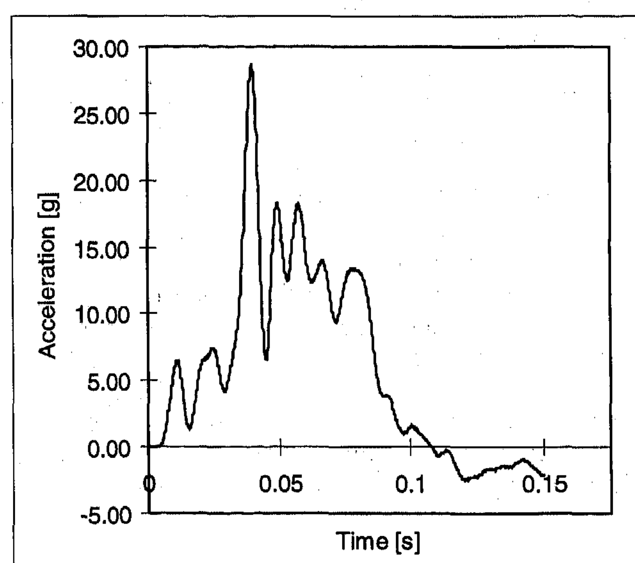


Figure 11: LAGUNA - deceleration of car cabin

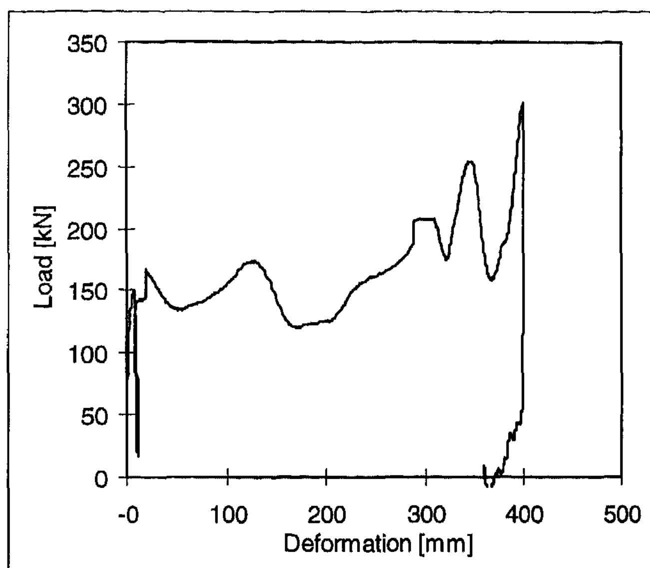


Figure 12: CRATCH - load at the interface

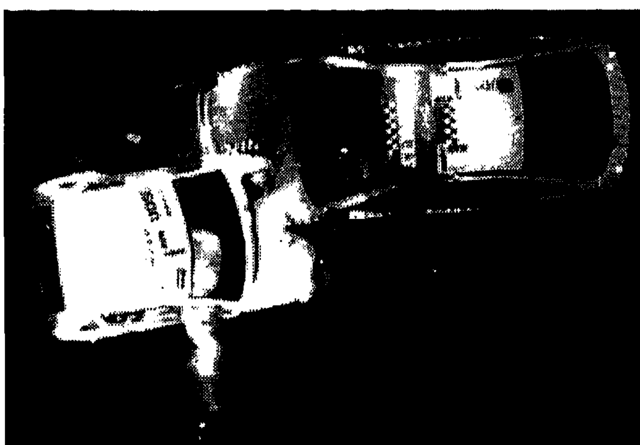


Figure 13: Maximum deformation of the two cars



Figure 14: Situation after the collision

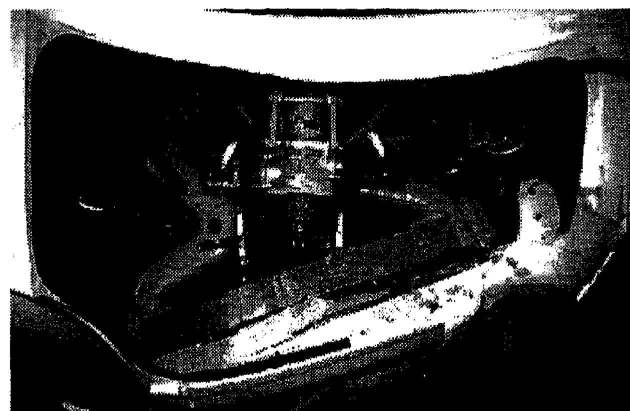


Figure 15: CRATCH - deformation of the front beams.

CONCLUSIONS

The safety deficits of small cars observed today in accident statistics can be alleviated if the structure and the restraint systems of these cars are designed and optimised for the situation they will most likely encounter in a real world situation, e.g. a collision against a heavier car.

For the extreme example of a low mass vehicle with a curb weight 25 % lower than current subcompact cars, it was shown that the development of a rigid lightweight structure with well defined properties of the crush zone is feasible and that a restraint system which can meet occupant protection criteria even at a Δv of 21.5 m/s (77.4 km/h) and car mean deceleration levels of 400 - 500 m/s^2 can also be conceived [5].

ACKNOWLEDGMENTS

The study presented here was in part supported by the Swiss Foundation for Traffic Safety and the Swiss Ministry of Energy.

We wish to thank ALUSUISSE (Neuhausen, Switzerland) for helping us to find an adequate technology and for building this experimental spaceframe, Renault (Paris, France) for the cooperation, for performing the preliminary crash test and for supplying two Laguna cars, PARS GmbH, Alzenau, Germany, for performing the sled tests, and MST GmbH, Aschaffenburg, Germany, for supplying the original and modified airbag modules and steering wheels.

REFERENCES

[1] Evans L., Frick M.
Driver fatality risk in two-car crashes
Dependence on masses of driven and striking car.
Proc. Conf. ESV 13, 1991

[2] Appel H.
Auslegung von Fahrzeugstrukturen im Hinblick auf Kollisionen zwischen kleinen und grossen Fahrzeugen
Verkehrsunfall (11) 1972, p221ff.

[3] Puleo G.
Small Vehicle Structural Design
Trans. SAE Vol. 82 (1973), SAE 730075

[4] Niederer P. et al
Occupant Safety of Low-Mass Vehicles
Proc. Conf. Stapp (37), 1993, p1ff., SAE 933107

[5] Muser M. et al.
Optimised restraint systems for Low Mass Vehicles
Proc. Conf. STAPP 1996, SAE 962435

[6] Kaeser R. et al.
Passive Safety Potential of Low Mass Vehicles
Proc. Conf. IRCOBI 1995, p337ff.

970123

The Relationship Between Car Size and Occupant Injury in Traffic Accidents in Japan

Koji Mizuno, Tetsuhiro Umeda, and Hideki Yonezawa
Traffic Safety and Nuisance Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The relationship between car size and occupant injury is examined using traffic accident data from all over Japan. For head-on collisions, the occupant injury rate is formulated based on the approximation of occupant injury by delta-V. The effects on occupant injury of the sizes of the striking and struck vehicles, as well as the effects of seat belt use and vehicle velocity, are examined in head-on, side-impact and single-car collisions.

As occupant injuries are also influenced by the other car or cars involved in a collision, the number and size distribution of vehicles is important. Sensitivity analysis shows that the effect of the number of lighter cars is greater in head-on collisions and that the effect of the number of heavier cars is greater in side-impact collisions, relative to the total number of fatalities.

INTRODUCTION

It is well known that the injury risk to the occupant is higher in lighter vehicles in head-on collisions. The effect of vehicle mass on occupant injury has been examined in detail by Evans.[1] He claims from the data of FARS that when 900-kg and 1800-kg cars collide head-on, the risk of fatality in the smaller car is 13 times higher than that in the larger car.

The higher injury risk due to the difference of mass between two vehicles is defined as the aggressiveness of the heavier vehicle. The aggressiveness of vehicle mass increases when combined with high vehicle rigidity, such as is found in light trucks and vans.[2] The compatibility between small car and large car is investigated and low mass vehicle is proposed with an optimized restraint system.[3]

In single-car accidents, the injury rate does not vary as much as in head-on collisions, though it has been shown by comparison to pedestrian fatalities that the risk of fatal injury increases as the mass of the vehicle decreases.[1] The analysis was carried out not only with respect to vehicle mass, but also the power/weight ratio of the car.[3] The risk of fatality increases in proportion to the power/weight ratio of the car. This is related to the fact that sports and specialty cars with high power/weight ratios and which are driven by younger drivers are frequently involved in high-speed single-car accidents.[5]

The automotive traffic environment varies from country to country according to the vehicle population, road conditions, and traffic laws. The average mass of a passenger car in Japan is 1131 kg (1994), which is smaller than that of cars in the United States. One reason the average mass is so small is the existence in Japan of mini-passenger cars, which average only 639 kg and the size of which is defined by Japanese law. Therefore, since occupant injury is affected by car size, the question must be examined in the context of the Japanese traffic environment to know what occurs in collisions with different car sizes.

This research focuses on the relationship between occupant injury and car size, especially vehicle mass, in traffic accidents in Japan. We mainly used the law enforcement data on traffic accident and analyzed statistically the size effect of the car.

METHODOLOGY

As law enforcement data on traffic accidents belongs to the National Police Agency and vehicle data belongs to the Ministry of Transport (MOT), in the past accident data was not correlated to vehicle design factors, such as mass, body dimensions, and engine displacement. Therefore the relationship between occupant injuries and vehicle designs was difficult to analyze. In 1993, the

Institute of Traffic Accident Research and Data Analysis (ITARDA) was established by the National Police Agency, the Ministry of Transport and the Ministry of Construction. ITARDA aims to reduce road accidents and injuries through comprehensive investigation and research on traffic accidents. Data on traffic accidents, drivers, roads and vehicles are integrated by ITARDA, so that it is now possible to correlate accident data with vehicle design factors such as body dimensions, mass, wheelbase or engine displacement.

The National Police Agency collects data only on accidents where at least one occupant suffered injury or death. Accordingly, in this paper the probability of injury means the fraction of injuries to occupants involved in such accidents. As the number of vehicle occupants killed in traffic accidents is about 4,500 per year and accident data for the four years from 1992 to 1995 was used, the total number of fatalities alone was insufficient for statistical analysis. Therefore, the broader classification of serious or fatal injury was examined. As no injured passengers are not included in the accident data, only driver injuries were examined in order to simplify the analysis.

Occupant injuries are examined in accident configurations such as head-on, side-impact, and single-car collisions, which are in turn classified according to location of impact. In this paper, a head-on collision is defined as one in which the impact location on both vehicles is on the front, including a frontal oblique collision. Side-impact collisions are defined as those in which the front of one vehicle collides with the side of another vehicle. Note that in the accident data, only eight locations of deformation are defined: front, right front, left front, right side, left side, rear, right rear and left rear.

Accidents between passenger cars and single-car accidents of passenger cars made up half (49.3%) of the total number of accidents involving 4-wheel vehicles in 1994. Since most passenger cars are bonnet types with similar mechanisms of injury, rather than cabover-types, this study deals with collisions between bonnet-type cars or single-vehicle accidents of bonnet-type cars.

Table 1 Categories of vehicles

	Dimensions			Engine
	Length (meters)	Width (meters)	Height (meters)	Displacement (liters)
Mini-size	3.3 or less	1.4 or less	2.0 or less	0.66 or less
Small-size	4.7 or less	1.7 or less	2.0 or less	2.0 or less
Ordinary-size	No limit	No limit	No limit	No limit

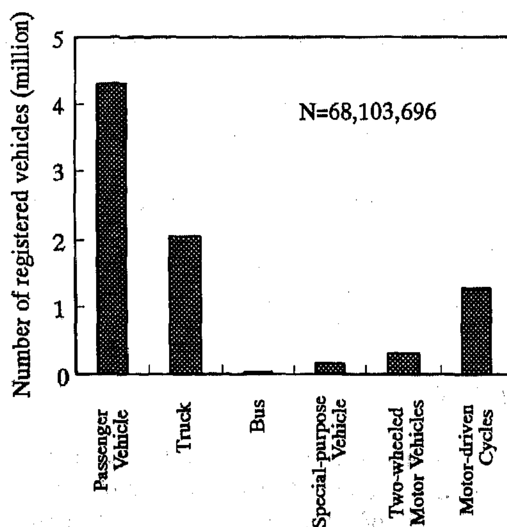


Fig. 1. Registered vehicles (March, 1995)

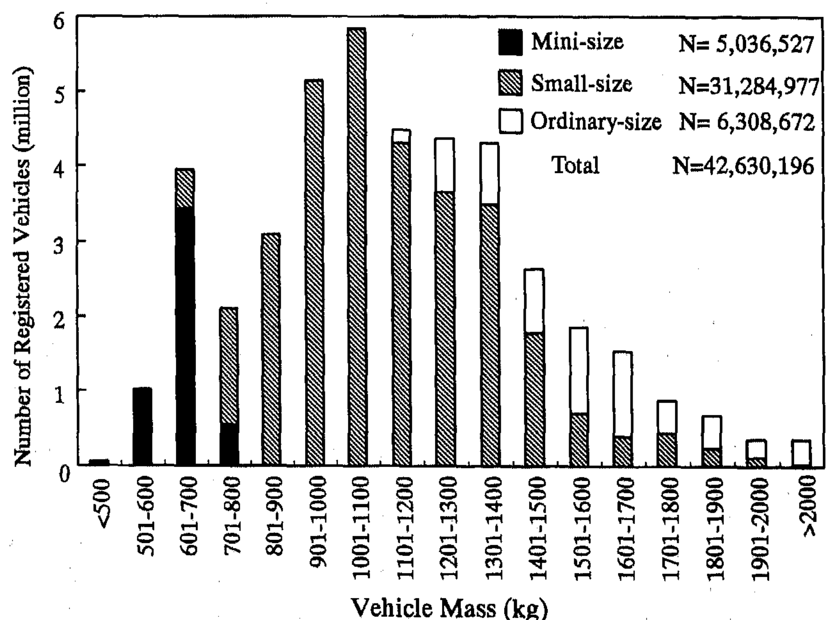


Fig. 2. Registered passenger vehicles versus vehicle mass in Japan (Dec. 1994, Data collected by ITARDA)

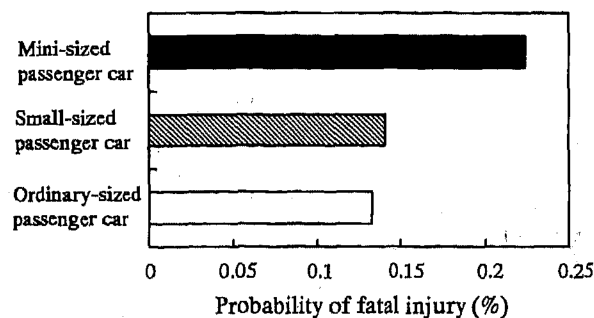


Fig. 3. Probability of fatal injury to driver involved in car-to-car collision (1994)[5]

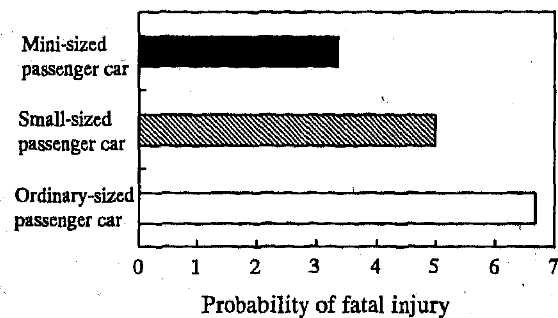


Fig. 4. Probability of fatal injury to driver involved in single car accident (1994)[5]

VEHICLE CATEGORY AND CLASSIFICATION

Vehicles are classified by their use, such as passenger vehicle, truck, bus and special-purpose motor vehicle. Fig. 1 shows the number of registered vehicles in Japan. The vehicles are categorized by their size and engine displacement, as mini-sized, small-sized and ordinary-sized vehicles (Table 1). The vehicle tax varies according to the vehicle's category and classification.

Fig. 2 shows the distribution of passenger vehicles according to mass. The mass of the vehicle is determined by its size and engine displacement, so the masses of mini-sized, small-sized and ordinary-sized cars cover different ranges. The mass of a mini-sized passenger vehicle is from 500 kg to 800 kg, that of a small-sized passenger vehicle is greater than 600 kg, and that of an ordinary-sized passenger vehicle is greater than 1100 kg. Many mini-sized cars are registered as trucks to take advantage of lower taxes, but they have almost the same structure and usage as a passenger car excluding the folding seats for goods loading facilities.

Bonnet-type passenger cars must be put through a front crash test against a fixed barrier, with the speed set at 50 km/h for small-sized and ordinary-sized cars, and at 40 km/h for mini-sized cars. In 1998, the legally defined minimum dimensions for a mini-size car will be changed to 3.4 meters long and 1.48 meters wide in order to meet the standards for a frontal crash test at 50 km/h.

As different drivers elect to use mini-sized, small-sized or ordinary-sized cars, the accident involvement rate and fatal accident rate are different for each category of cars. Fig. 3 shows the probability of injury to the drivers in car-to-car collisions including rear-end collisions, and Fig. 4 shows that of the drivers in single car accidents. In car-to-car collisions, it is shown that the probability of fatal injury to the driver in a smaller car is higher, especially in mini sized car. However, in single car accidents, the probability of fatal injury to the driver in a smaller car is lower. This is because the crash velocity of a smaller car is lower in a single car accident.

INJURY RATE

HEAD-ON COLLISION - The data on head-on collisions between bonnet-type cars for the years 1992 to 1995 include 950 driver fatalities, 11,634 serious injuries, 185,848 minor injuries and 167,034 cases of no injury to the driver. The percentage of fatally injured drivers wearing seat belts was 30.2%, which is far less than the 77.6% figure for all drivers involved in head-on collisions. According to the data, the percentage of drivers wearing seat belts does not vary much with vehicle mass.

Vehicle mass has a greater effect on the severity of occupant injuries in head-on collisions than in other accident configurations. Fig. 5 shows the probability of serious or fatal injury to the driver of car 1 in a head-on collision between car 1 and car 2. The probability of serious or fatal injury increases as the mass of car 1, m_1, decreases and that of car 2, m_2, increases. When m_1 is less than 700 kg and m_2 is greater than 1401 kg, the probability of serious or

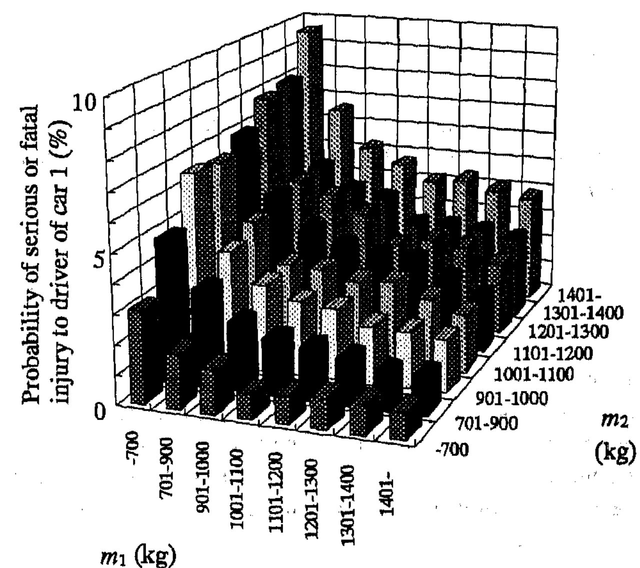


Fig. 5. Probability of serious or fatal injury to driver of car 1 in head-on collision

fatal injury to the driver of car 1 is 9.5%. By contrast, when m_1 is greater than 1401 kg and m_2 is less than 700 kg, the probability of serious or fatal injury to the driver of car 1 is 0.95%.

The probability of occupant injury depends not only on the mass of the car in which the occupant is traveling but also that of the other car. Thus, the probability of injury to the occupant in a car with a given mass is influenced by the numerical distribution of cars, and the travel and accident rates. Fig. 6 shows the probability of serious or fatal injury to the driver in a head-on collision relative to the mass of the driver's vehicle. Fig. 7 shows the same probability relative to the mass of the other vehicle. The probability of serious or fatal injury increases exponentially as the mass of the driver's car decreases, especially if the mass is less than 1000 kg. On the other hand, the probability of injury steadily increases as the other vehicle's mass increases.

The occupant injury rate versus car mass is estimated using delta-V. We define the probability of injury as

$R_i(m_i, m_j)$ = Probability of injury to occupant of car i, in collision with car j,

where m_i, m_j = masses of vehicles i and j. Delta-V is defined as the change in the car passenger compartment velocity during an impact and is used as an indicator of impact severity. As delta-V is the velocity change that the occupant experiences, it is potentially a factor for occupant injury. Thus the probability of occupant injury is approximated by delta-V to the power k, as[6]

$$R = r_0(\Delta v)^k , \quad (1)$$

where r_0 is the constant for injury severity.

Using Eq. (1) and Newton's third law, the relationship between delta-V and the probability of occupant injury can be expressed as

$$\frac{R_2}{R_1} = \left(\frac{\Delta v_2}{\Delta v_1}\right)^k = \left(\frac{m_1}{m_2}\right)^k , \quad (2)$$

where Δv_1, Δv_2 = delta-V of vehicles 1 and 2. The ratio of vehicle mass and probability of occupant injury is shown in Fig. 8 in the manner demonstrated by Evans.[6] The parameter k is obtained for serious or fatal injury, and for more than minor injury, according to the slope of the straight line.

$$\begin{cases} k = 2.67 \text{ (Serious or fatal injury)} \\ k = 1.07 \text{ (More than minor injury)} \end{cases} \quad (3)$$

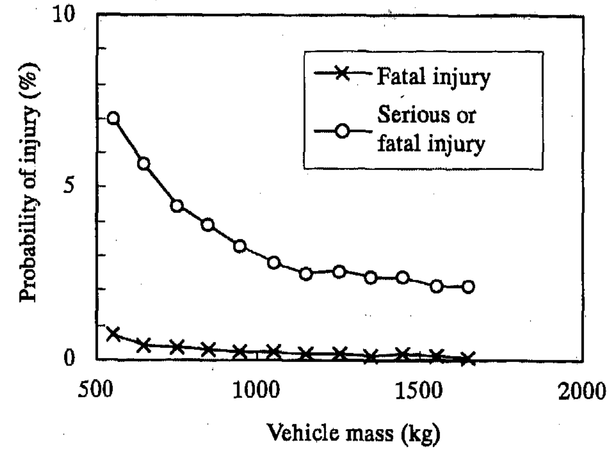


Fig. 6. Probability of injury to driver in head-on collision

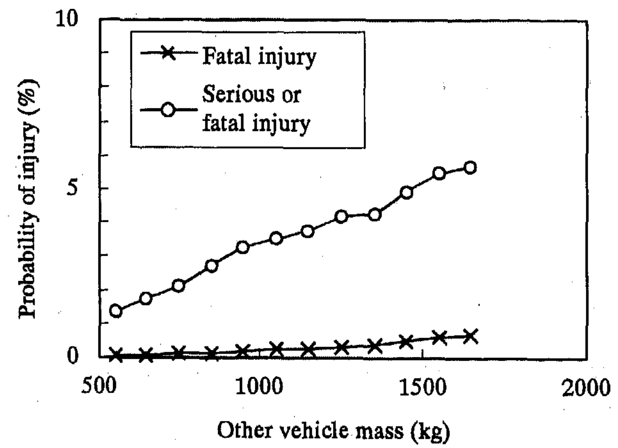


Fig. 7. Probability of injury to driver versus mass of other vehicle in head-on collision

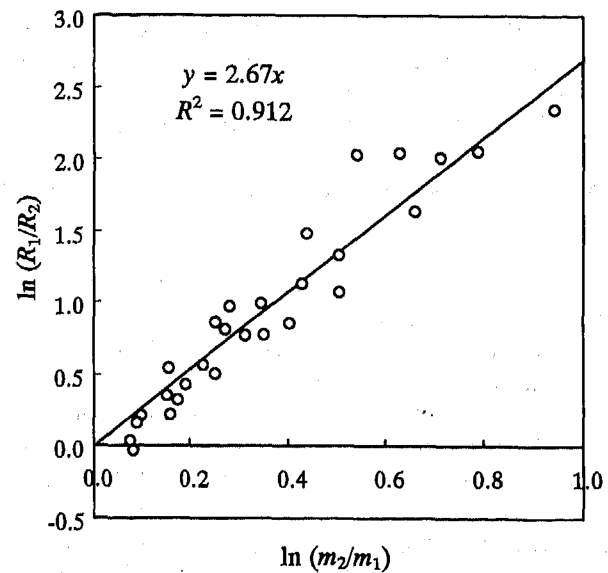


Fig. 8. Ratio of car mass and probability of serious or fatal injury in head-on collision

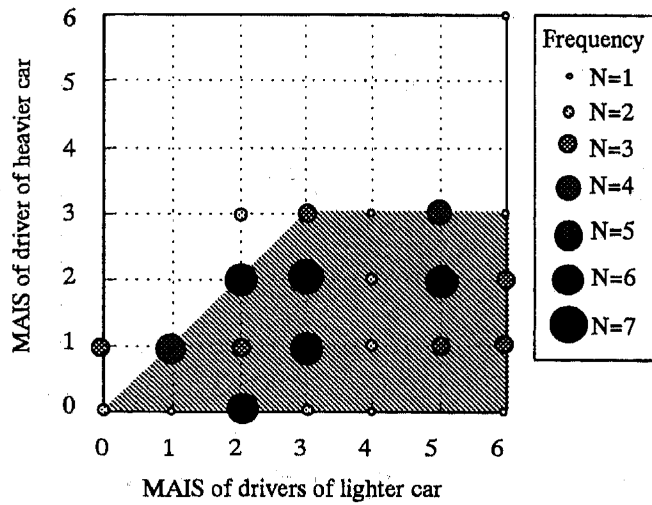


Fig. 9. Comparison of AIS of each driver in head-on collision (N=59)

It is well known that human head tolerance is proportional to $\alpha^{2.5}T$. Thus the probability of serious injury is expressed as

$$R \propto \alpha^{2.5}T \tag{4}$$

where α is acceleration and T is duration time. The ratio of probability of serious or fatal injury, R_2/R_1, is obtained as follows:

$$\frac{R_2}{R_1} = \left(\frac{\alpha_2}{\alpha_1}\right)^{2.5} = \left(\frac{m_1}{m_2}\right)^{2.5} \tag{5}$$

Comparing Eq. (2) with Eq. (5), the exponent value of 2.67 for serious injury or fatality derived from Eq. (2) is seen to be close to the posited value of 2.5, suggesting that the value 2.67 is appropriate for serious injury or fatality.

There is a tendency of higher risk to the driver of a lighter car. Fig. 9 shows the Abbreviated Injury Scale (AIS) for the driver of each car involved in head-on collisions investigated by the MOT. [7] In many cases the AIS for the driver of the lighter car is larger than that for the driver of the heavier car. Furthermore, in most of the head-on collisions investigated, the AIS for the heavier car is less than or equal to 3, whereas that for the driver of the lighter car is in the range 0 to 6. That is to say, the AIS for each driver is distributed in the region under the oblique line shown in Fig. 9.

For many head-on collisions, delta-V is approximated for a central collision. Assuming the restitution coefficient is zero, delta-V is obtained for a central collision as follows:

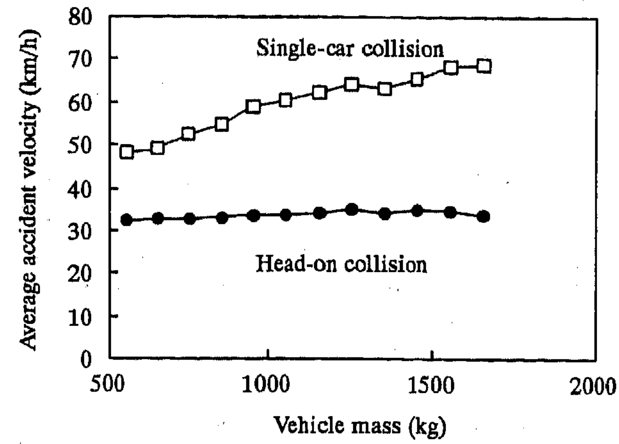


Fig. 10. Average accident velocity in head-on collision

$$\Delta v_1 = \frac{m_2}{m_1 + m_2} v_c \tag{6}$$

where v_c is closing speed. From Eq. (1) and Eq.(6), the probability of injury for the driver in car 1 is expressed as follows:

$$\begin{aligned} R_1 &= r_0 \left(\frac{m_2}{m_1 + m_2} v_c\right)^k \\ &= a \left(\frac{m_2}{m_1 + m_2}\right)^k \end{aligned} \tag{7}$$

where $a = r_0 v_c^k$.

Before discussing Eq. (7), the relationship between the coefficient a, closing speed and vehicle mass must be examined. Fig. 10 shows the average accident velocity versus vehicle mass in head-on and single-car collisions. The accident velocity is defined as the velocity at the moment the driver perceives the hazard of colliding with another vehicle, a pedestrian, or a parked car or other stationary object. Specifically, it indicates the velocity before the driver brakes or steers to avoid the accident, and is compiled mainly from drivers' testimony. For head-on collisions, the accident velocity does not change significantly with vehicle mass (32.5-35.2 km/h) as shown in Fig. 10. On the other hand, for single-car collision, the accident velocity increases from 48.4 to 68.6 km/h as the vehicle mass increases. It is possible that the characteristics of drivers who are involved in head-in collisions are different from those of drivers who are involved in single-car collision. The accident velocity is likely to be closely related to collision velocity. This fact suggests that the collision velocity should be considered in a single-car accident,

whereas it does not have to be considered in a head-on collision, when statistically evaluating the effect of vehicle mass on occupant injury. Thus, the coefficient a is thought to be independent of vehicle mass in head-on collisions in Eq. (7).

According to Eq. (7), there is a linear relationship between $\ln(R_1)$ which is natural logarithm of R_1 and $\ln\{m_2/(m_1+m_2)\}$. Fig. 11 demonstrates the strong relationship between $\ln\{m_2/(m_1+m_2)\}$ and $\ln(R_1)$ for serious and fatal injury. The coefficient a is obtained by an intercept of the straight line in Fig. 11 as follows:

$$\begin{cases} a = 20.6 & \text{(Serious or fatal injury)} \\ a = 112.2 & \text{(More than minor injury)} \end{cases} \quad (8)$$

The regression of the probability of injury due to vehicle mass in a head-on collision is obtained using Eq. (3), Eq. (7) and Eq. (8). According to Eq. (8), the probability of injury exceeds 100% when the mass of the other car, m_2, is infinite. This is because Eq. (8) applies only to bonnet-type cars under 2000 kg. Eq. (7) also suggests the probability of injury increases with the closing speed, which depends on v_c^k.

The effect of wearing a seat belt on the probability of serious or fatal injury versus vehicle mass is shown in Fig. 12. Wearing a seat belt reduces the risk of serious or fatal injury over the entire range of vehicle mass. The probability of fatal or serious injury for belted drivers is between 1.5 and 5.1%, whereas that for unbelted drivers is between 4.3 and 11.6%, which shows that the effect of vehicle mass is large for unbelted drivers. Thus, wearing a seat belt is more important for the occupant of a lighter car in order to mitigate the severity of injury.

The probability of serious or fatal injury versus mass ratio is compared for belted and unbelted drivers as shown in Fig. 13. The straight line for belted and unbelted drivers are parallel. When the slope of the line is fixed at 2.67 as obtained for all drivers in Eq.(8), a is estimated for serious or fatal injury as follows:

Table 2 The coefficient a and k

Injury severity	Seat belt	k	a
Serious or fatal injury	Belted	2.67	14.1
	Unbelted		37.9
More than minor injury	Belted	1.07	106.7
	Unbelted		133.7

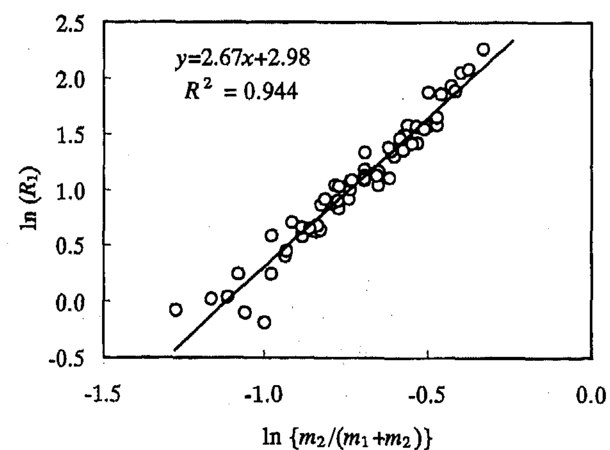


Fig. 11. Mass ratio and the probability of injury to driver in car 1 in head-on collision

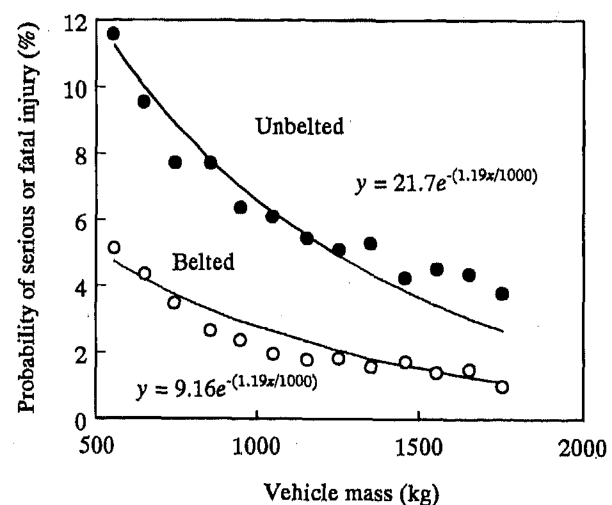


Fig. 12. Effect of seat belt on probability of serious or fatal injury to driver in head-on collision

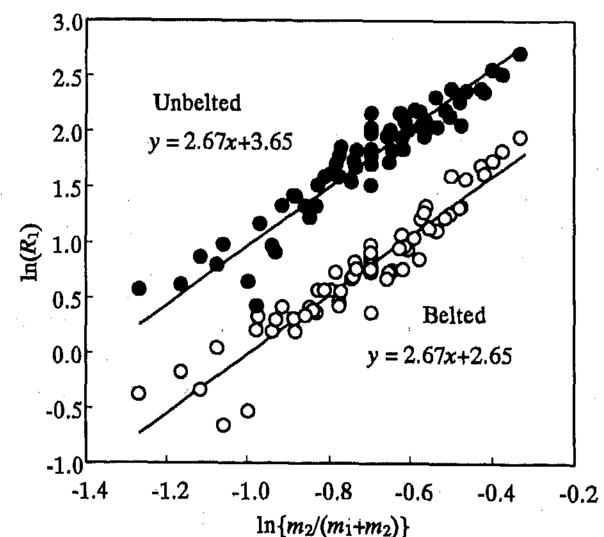


Fig. 13. The comparison of the probability of serious or fatal injury for belted and unbelted drivers in head-on collisions

$$\begin{cases} a_{Belted} = 14.1 & \text{(Belted driver)} \\ a_{Unbelted} = 37.9 & \text{(Unbelted driver)} \end{cases} \quad (9)$$

Table 2 summarizes a and k for restraint condition and injury severity.

When the combination of car mass (m_1, m_2) is equal for two head-on collisions, the probability of injury for belted and unbelted driver is obtained by Eq.(7) and Eq.(9) as:

$$\frac{R_{Unbelted}}{R_{Belted}} = \frac{a_{Unbelted}}{a_{Belted}} = 2.68 \text{ (Serious or fatal injury)} \quad (10)$$

Fig. 14 shows the average accident velocity classified by seat belt wearing. The accident velocity is higher for cars driven by unbelted drivers, which shows the risk involved in a severe accident is higher for unbelted drivers. In Eq.(7) a is dependent on crash velocity, however, in Eq.(10) the difference of the crash velocity between belted and unbelted drivers is not considered.

Thus, a double pair comparison method is used in order to compare the injury rate between belted and unbelted drivers. The ratio of seriously or fatal injury rate of the driver in car 1 to that of driver in car 2, R_1/R_2, is obtained by the ratio of the number of serious or fatal injuries of the drivers in car1 to that of the drivers in car 2. Using this method, injury rate is not affected by the difference of crash velocities between two cars driven by belted and unbelted drivers. Fig. 15 shows the ratio of injury rate classified by seat belt wearing of the driver in car 1 and car 2. When both drivers wear a seat belt or don't wear a seat belt in car 1 and car 2, R_1/R_2 is expressed using Eq.(2) and Eq.(3) as:

$$R_{1,Belted} / R_{2,Belted} = R_{1,Unbelted} / R_{2,Unbelted} = (m_2 / m_1)^{2.67} \quad (11)$$

where $R_{i,\,Belted}$ is the injury rate of the belted driver in car i and $R_{i,\,Unbelted}$ is that of the unbelted driver in car i. The data obtained by the double pair comparison method for belted/belted and unbelted/unbelted drivers is distributed along the approximate curve in Eq.(11). In the case where the driver in car 1 do not wear seat belt and the driver in car 2 wears a seat belt, the ratio of serious or fatal injury is approximated as:

$$R_{1,Unbelted} / R_{2,Belted} = 1.89(m_2 / m_1)^{2.67} = (1.27 m_2 / m_1)^{2.67} \quad (12)$$

This result shows when the combination of striking and struck car mass (m_1, m_2) is equal, the serious or fatal injury rate for unbelted drivers is 1.89 times higher than that for unbelted drivers. Eq. (12) also shows that in order to gain

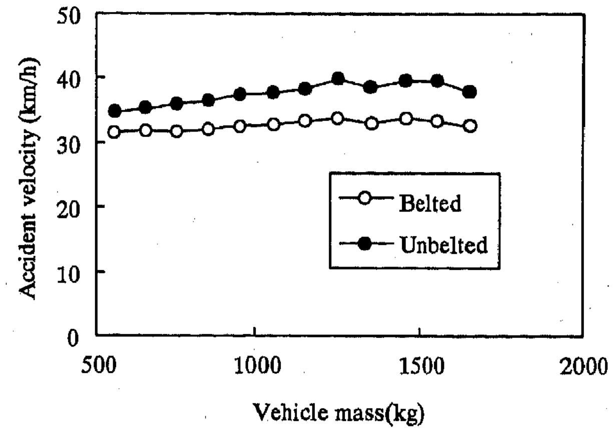


Fig. 14. Average accident velocity of cars with belted and unbelted drivers in head-on collisions

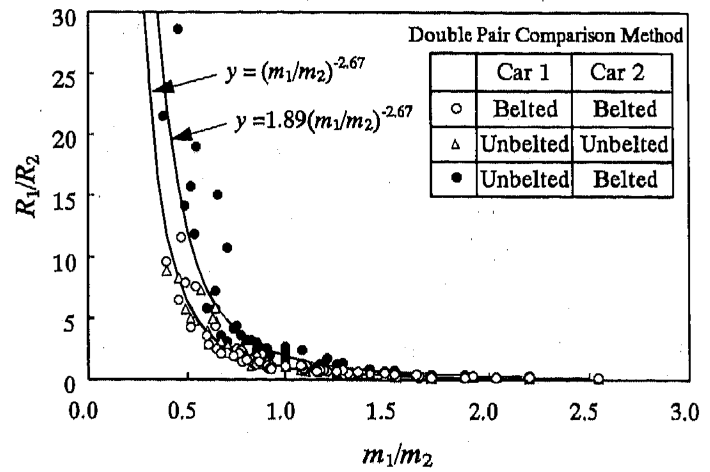


Fig. 15. Ratio of probability of serious or fatal injury for driver's restraint condition in car 1 and 2

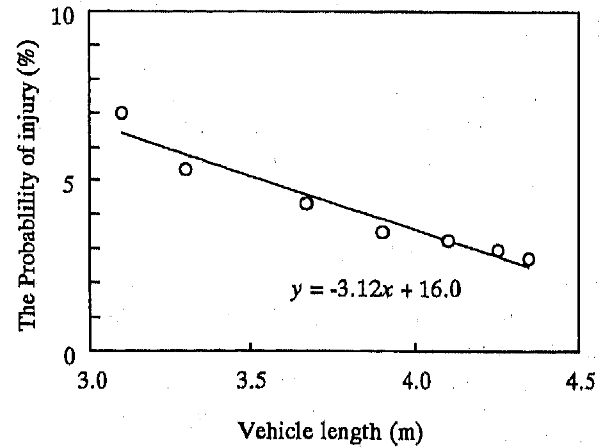


Fig. 16. Car length and probability of serious or fatal injury in head-on collision

the same probability of serious or fatal injury, the unbelted driver requires a vehicle mass 1.27 times that of vehicle mass driven by the belted driver.

The length as well as the mass of a car influences the severity of injury to the occupant. As the total length of the car increases, the hood is likely to be longer and to have a greater capacity to absorb impact. Fig. 16 shows the effect of car length on the probability of serious or fatal injury to the driver. The relation between length and the probability of serious or fatal injury can be described as straight line. However, since length is strongly linked to vehicle mass, it is difficult to identify the effect of length alone.

SIDE-IMPACT COLLISION - The data on side-impact collisions between bonnet-type cars for the years 1992 to 1995 includes 88,127 drivers struck from the right side (the driver's side in Japan) and 87,261 drivers struck from the left side. Among these, right-side collisions accounted for 296 deaths, 3,159 serious injuries, and 59,606 minor injuries, while left-side collisions resulted in 248 deaths, 2,199 serious injuries, and 49,859 minor injuries. Therefore, the number of deaths is larger in collisions on the driver's side, even though the number of side-impact collisions is almost the same for the right and left sides. In this section, the probability of injury to the driver when the other car strikes the right side is examined. Fig. 17 shows the probability of driver injury in a car struck from right side. The probability of injury to the driver increases as the mass of the struck vehicle decreases and that of the striking vehicle increases, although the effect of the struck vehicle mass on the probability of injury is somewhat smaller than it is in head-on collisions.

The effect of the mass of the struck and the striking vehicle on the probability of driver injury is shown in Fig. 18 and Fig. 19 respectively. As the mass of the struck vehicle decreases and the mass of the striking vehicle increases, the probability of injury increases. The probability of driver injury ranges from 3.0 to 6.3% relative to the struck vehicle mass, in contrast to the range from 2.1 to 7.0% for a head-on collision. Thus, the struck vehicle mass influences injury probability more significantly in a head-on collision than in a side-impact collision.

The relationship between the mass ratio of two vehicles in a side-impact collision and the probability of occupant injury is shown in Fig. 20. The correlation between mass ratio and probability of injury is small compared to a head-on collision (Fig. 11). Two reasons for this may be proposed. First, the impact direction diverges as the two cars travel in different directions, i.e., Eq. (6) does not hold for two-dimensional behavior in a side-impact collision. Second, the occupant injury rate is not fully correlated to delta-V. In a side impact, the interior of the struck side hits the occupant before the exchange of momentum between

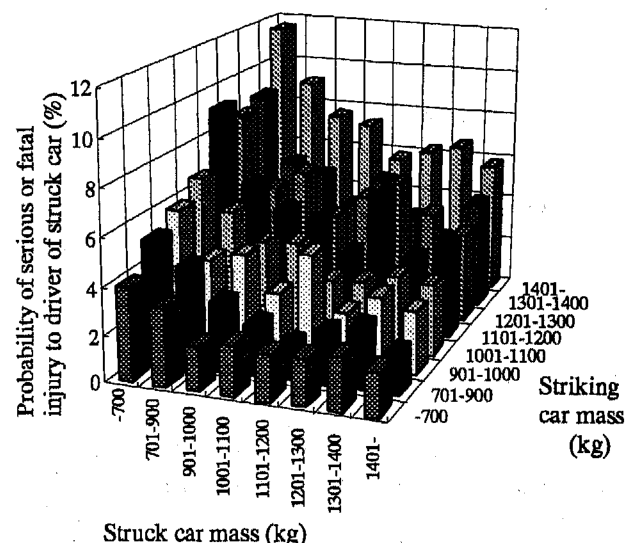


Fig. 17. Probability of injury to driver of car struck from right side

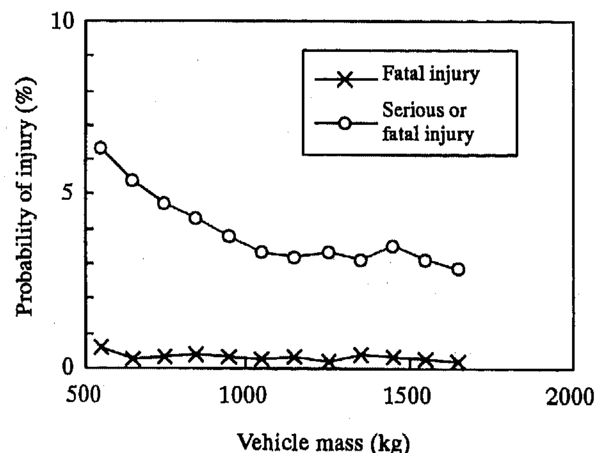


Fig. 18. Struck vehicle mass and probability of driver injury in side-impact collision. Location of impact is right side of car.

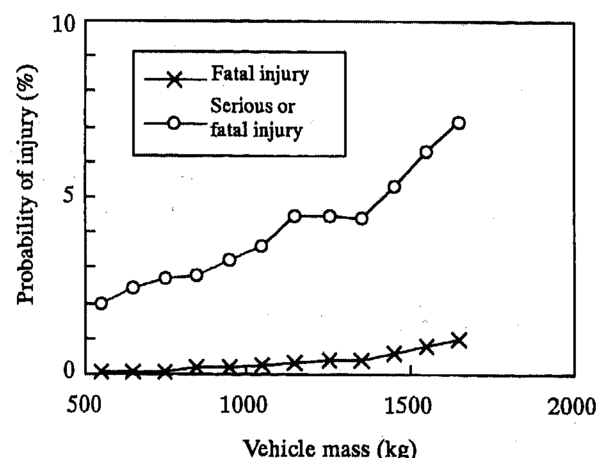


Fig. 19. Striking vehicle mass and probability of driver injury in side-impact collision. Location of impact is right side of car.

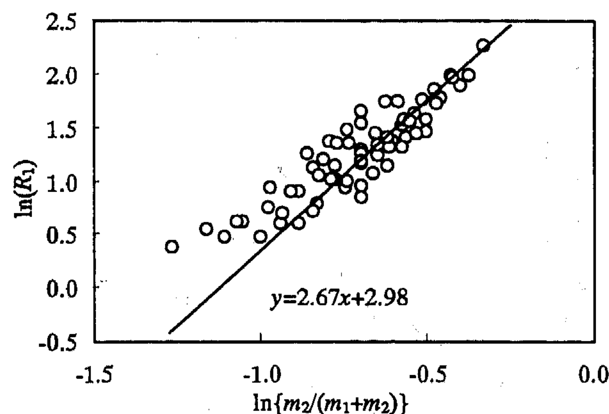


Fig. 20. Mass ratio and probability of serious or fatal injury to driver in struck vehicle in side-impact collision. Straight line is a regression line for head-on collision.

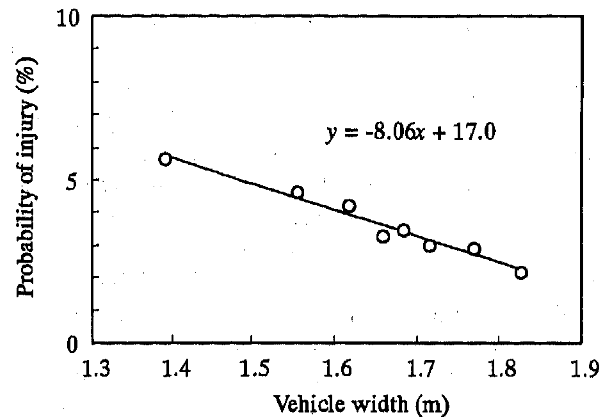


Fig. 21. Car width and probability of serious or fatal injury to driver in side-impact collision

two cars is over. The straight line in Fig. 20 is that for a head-on collision. The probability of injury in a side-impact collision can be seen to rise above this line when the mass ratio, $m_2/(m_1+m_2)$, is small, which demonstrates that the occupant on the struck side is likely to be seriously or fatally injured, even when the struck vehicle mass m_1 is larger or the striking vehicle mass m_2 is smaller in comparison to a head-on collision.

The width of the car affects the injury probability, as has sometimes been shown in side-impact tests and simulations.[8][9] As the width of the car increases, more space becomes available for reinforcing structures, door beams or door padding, and the distance between occupant and door increases. Fig. 21 shows the relationship between car width and the probability of injury to the struck side driver. The probability of serious or fatal injury decreases linearly with car width. Note that width directly affects the vehicle mass, as does length, so that we cannot pinpoint the width effect alone using the data in Fig. 21.

SINGLE-CAR COLLISION - In 1995, the number of fatal single-car accidents was 1,868, which amounts to 18.3% of all fatal accidents (total 10,227). Cars colliding with fixed objects accounted for 75.4% of single-car accidents. In many cases the fixed object with which the car collided was a telephone pole, signpost or guardrail. Our research examined accidents in which the front ends of bonnet-type cars collided against fixed objects. The data indicate that from 1992 to 1995, 59,101 drivers of bonnet-type cars were involved in single-car collisions, resulting in 2,552 deaths, 8,530 serious injuries, and 36,297 minor injuries, with 11,722 collisions resulting in no injuries to the driver.

Fig. 22 shows the probability of driver injury in a single-car collision with a fixed object. The probability of fatal injury increases with the vehicle mass, while that of serious or fatal injury reaches its minimum value when the vehicle mass is 1150 kg, though for both cases the effect of mass is not so large as it is in car-to-car collisions. However, the severity of the collisions in which drivers are involved are not factored into this result.

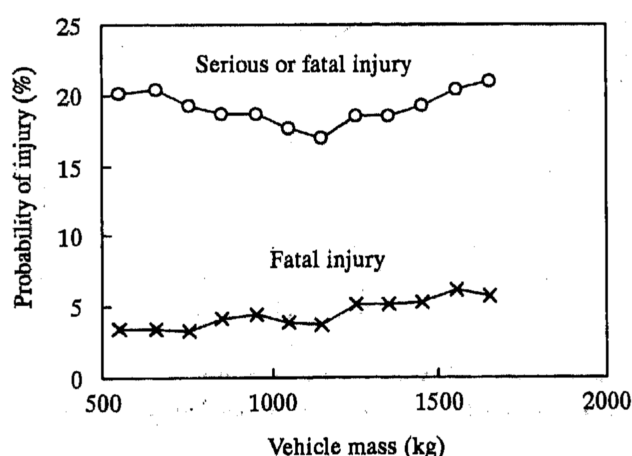


Fig. 22. Probability of driver injury in single-car collision

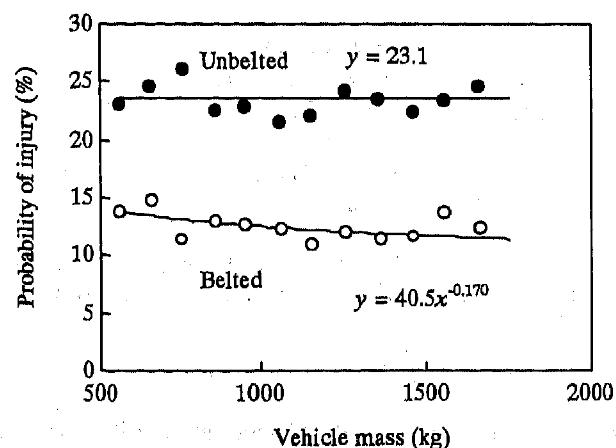


Fig. 23. Effect of seat belt on probability of serious or fatal injury to driver in single-car collision

The percentage of drivers wearing a seat belts in single-car collision is 44.2%, which is far lower than the 77.6% figure for head-on collisions. One reason for the low percentage is that younger drivers are disproportionately involved in single-car collisions. Note that there is a problem of reliability with the data on seat belt use, because it depends heavily on the testimony of the driver to the police. Nevertheless, the effect of wearing a seat belt is clear. For unbelted drivers, the probability of fatal injury is 6.6%, that of serious injury is 16.7% and that of minor injury is 61.3%, while for belted drivers, the probability of fatal injury is 1.4%, that of serious injury 11.3%, and that of minor injury 61.9%. Fig. 23 shows the effect of wearing a seat belt in a single-car collision. The probability of serious or fatal injury is higher for unbelted drivers, irrespective of vehicle mass. These results show the importance of wearing a seat belt in a single-car collision.

We examined the relationship between probability of occupants injury and vehicle mass, considering the occupant risk exposure in terms of accident velocity. Injury severity varies with accident velocity as shown in Fig. 24. As the accident velocity increases, the probability of fatal injury increases. A heavier vehicle is likely to be involved in a more severe collision, because the accident velocity tends to be higher for a heavier car than for a lighter car, as shown in Fig. 10. Fig. 25 shows that the accident velocity for a car weighing 550 kg is lower than that for a car weighing 1750 kg. The accident velocity distribution for a car weighing 1055 kg is similar to the average distribution. The probability of injury is calculated assuming the average accident velocity for all cars, regardless of mass.

The probability of injury to the occupant can be obtained from the velocity distribution and the probability of injury in a vehicle of mass m_i at a given velocity. Thus, the probability of occupant injury in a single-car collision, $R(m_i)$ is obtained as:

$$R(m_i) = \sum_j s(m_i, v_j) \cdot r(m_i, v_j), \qquad (13)$$

where i: vehicle category by mass,

j: velocity category by its range,

$s(m_i, v_j)$: probability of accident velocity classified by vehicle mass,

$r(m_i, v_j)$: probability of occupant injury for vehicle i with velocity v_j.

The probability of occupant injury in a single-car collision with the same distribution of accident velocity is given by

$$R_v(m_i) = \sum_j s(\bar{m}, v_j) \cdot r(m_i, v_j), \qquad (14)$$

where $s(\bar{m}, v_j)$: probability of accident velocity averaged for all vehicles.

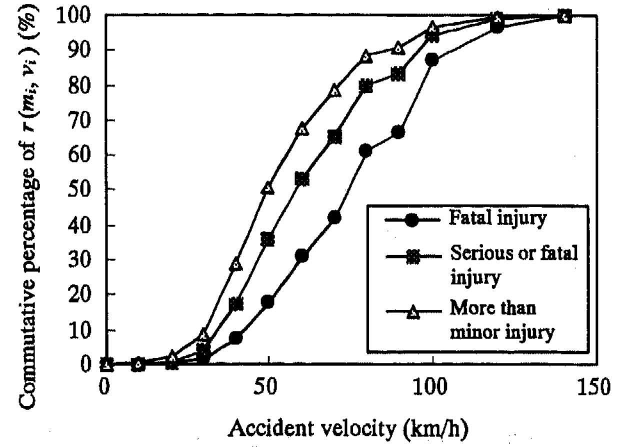


Fig. 24. Accident velocity distribution classified by injury severity in single car collision

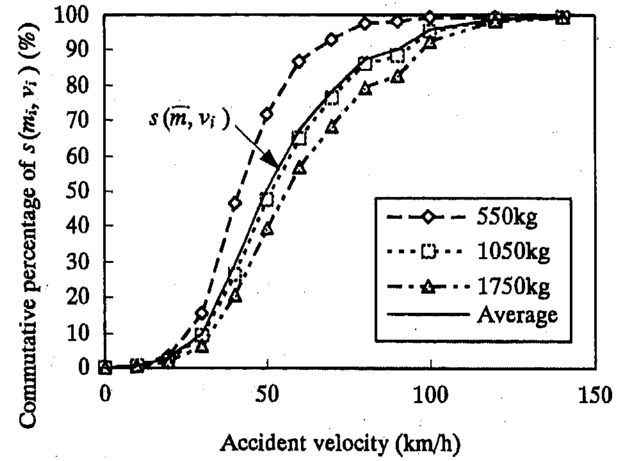


Fig. 25. Accident velocity distribution for vehicles weighing 550 kg, 1050 kg, 1750 kg in single-car collision

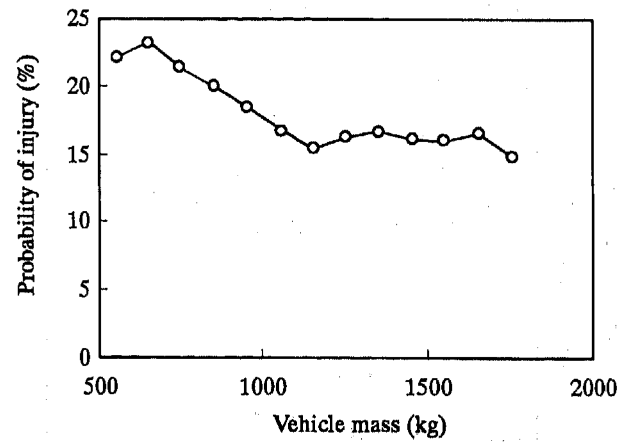


Fig. 26. Probability of serious or fatal injury corrected for accident velocity distribution in single-car collision

$R_v(m_i)$ is calculated for serious or fatal injury to the driver as shown in Fig. 26. As is evident from Fig. 26, the driver of a lighter car is at higher risk in a single-car collision when the accident velocity is the same as that of a heavier car. This can be explained by the fact that when the object with which the car collides is not as rigid as a fixed barrier, a heavier car is likely to destroy the object, making the occupant safer as a result.

BODY REGIONS OF INJURY

We examined the frequency of injury to different body regions in order to clarify the differences in injury patterns with different accident configurations and vehicle sizes. Fig. 27, Fig. 28 and Fig. 29 illustrate the frequency of injury to major regions of the driver's body, classified by severity, vehicle size, and occupant restraint in head-on, side-impact, and single-car collisions of passenger car.

Some characteristic differences can be found in the frequency with which injuries of a given level of severity occur in different body regions. The patterns of fatal injury in head-on, side-impact and single-car collisions are similar, irrespective of seat belt use. Among fatal injuries, head injuries account for 35-50%, chest injuries 18-26%, and the whole body, neck and abdomen about 10% each. However, for serious injuries, the patterns differ according to the accident configuration and occupant-restraint conditions. Serious leg injuries are proportionately more frequent in head-on and single-car collisions for both belted and unbelted drivers. In side-impact collisions, chest and pelvis injuries account for a higher proportion of serious injuries due to occupant contact with the side interior of the car. Neck injuries are the most frequent minor injuries in head-on and side-impact collisions. Wearing a seat belt prevents the occupant's head from striking the windshield or steering wheel in all accident configurations, thereby reducing the severity of injuries, but raising the proportion of minor neck injuries.

We can also find two remarkable differences in the pattern of body regions injured in different-sized cars. First, the frequency of head injury is higher for smaller cars in head-on and side-impact collisions, but is lower in single-car collisions. A smaller car is almost always a lighter car. In car-to-car collisions of the head-on and side-impact types, delta-V increases when the car is lighter or smaller. It is thought that the frequency of head injuries depends on delta-V. The second difference found is that the frequencies of chest and leg injuries are higher for smaller cars in head-on and single-car collisions with frontal impact. The frequency of leg injury is particularly high for mini-passenger cars. In single-car accidents, delta-V or deceleration is not necessarily higher for a smaller car because accident velocity tends to be lower for a smaller car. Thus, it is thought that the higher frequency of chest and leg injuries for a smaller car is due to its size rather than its lighter mass. In other words, it seems that the frequency of head injuries is due to delta-V and that of chest and leg injuries is due to the size of the car.

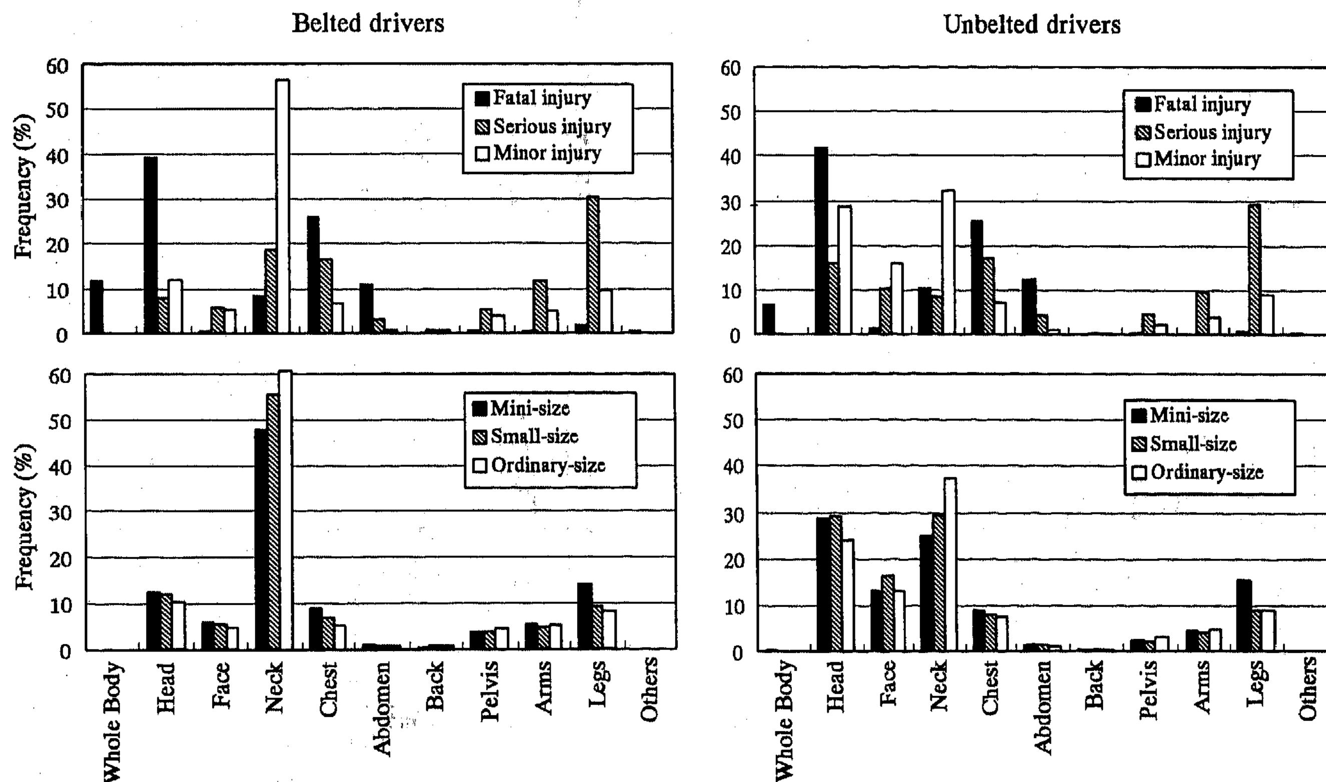


Fig. 27. Body regions of drivers injured in head-on collisions

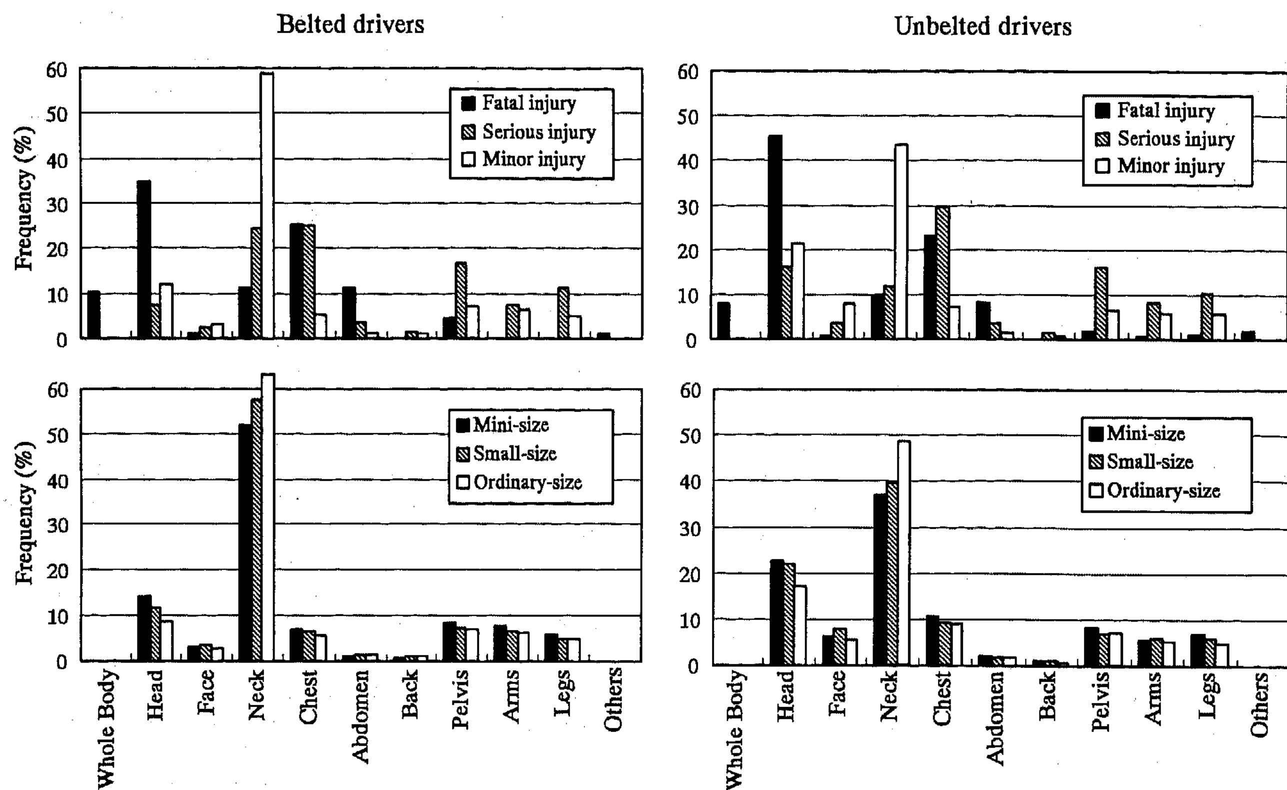


Fig. 28. Body regions of drivers injured in side-impact collisions

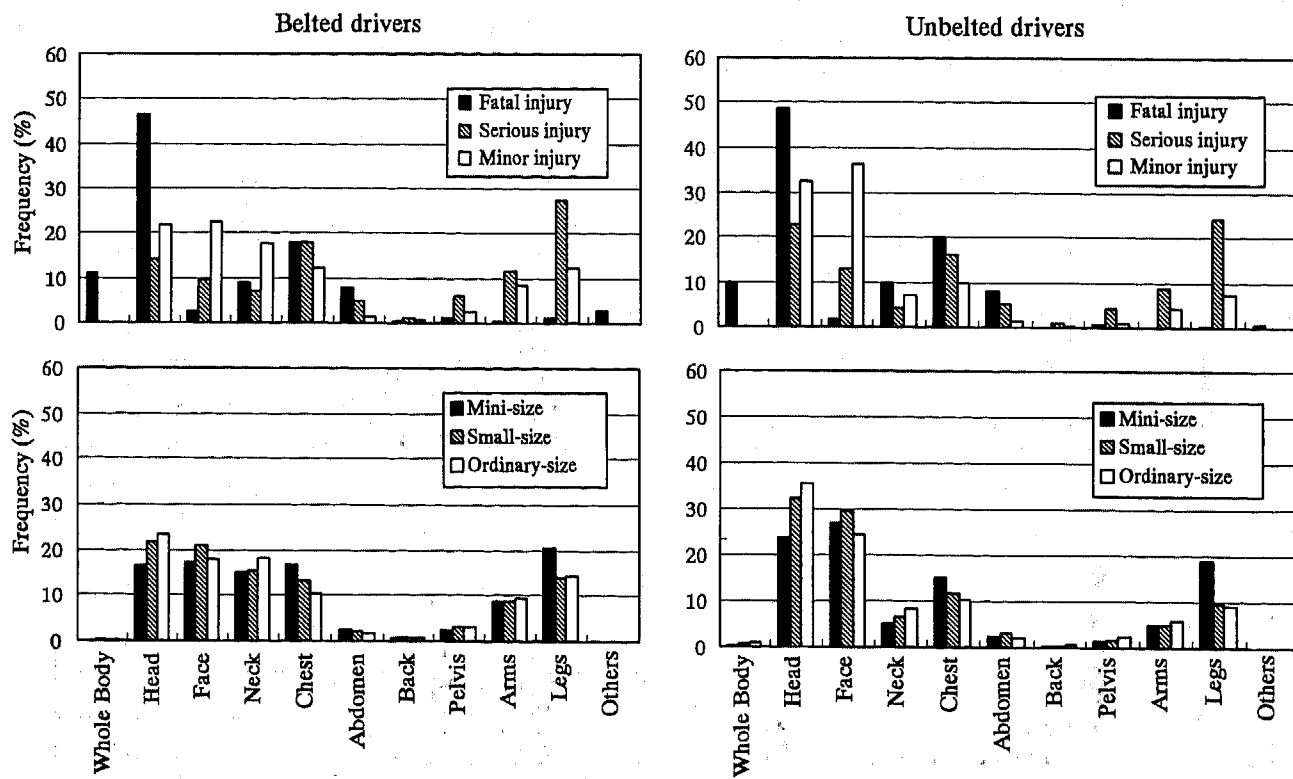


Fig. 29. Body regions of drivers injured in single-car collisions

NUMERICAL DISTRIBUTION OF VEHICLES

As the vehicle mass increases, the risk to occupants of the vehicle decreases, but the risk to occupants of the other vehicle increases. Thus, the number of injuries is influenced by the mass of both vehicles. The number of injuries in head-on collisions is expressed using Eq. (7) as follows:

$$N_I = (R_1 + R_2)N_A$$
$$= \frac{m_1^k + m_2^k}{(m_1 + m_2)^k} aN_A \qquad (15)$$

where N_I is the number of injuries and N_A is the number of accidents. N_I has the minimum value, $2^{1-k} a N_A$, when m_1 is equal to m_2 (k>1), which shows that N_I is small when the difference in the masses of the vehicles is small.

The total number of injuries in a collision will be influenced by the number of cars involved and the distribution of their respective masses. The total number of injuries, N, is calculated as

$$N = \sum_{i,j} p n_i n_j R_i(m_i, m_j), \qquad (16)$$

where i, j: vehicle category classified by mass, m_i, m_j,
n_i : number of vehicles in category i,
$R_i(m_i, m_j)$: probability of injury to occupant of vehicle i struck by vehicle j,
p : probability of accident.

Sensitivity analysis was conducted to clarify the effect of the number of vehicles of similar mass on the total number of seriously and fatally injured occupants. The sensitivity, γ_i, of the number of vehicles n_i to the total number of injuries N is calculated as:

$$\gamma_i = \lim_{\Delta n_i \to 0} \frac{\Delta N}{\Delta n_i} \qquad (17)$$

To simplify the analysis, we assumed that the probability of accident p is the same for all vehicles, though this is not true for the actual vehicle population. The probability of occupant injury in a bonnet-type car is used for $R_i(m_i, m_j)$ and the numerical distribution of passenger cars is used for n_i. The total number of cars is assumed to be constant ($\sum n_i = const.$). The probability of injury to the driver in a side-impact collision takes into account car's being struck from both sides. The results are indicated in Fig. 30 for head-on and side-impact collisions. When the sensitivity is positive, the number of serious and fatal injuries increases. In both head-on and side-impact collisions, the number of serious and fatal injuries decreases

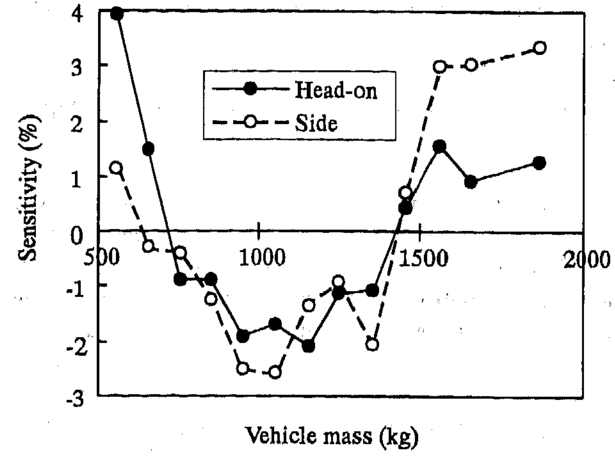


Fig. 30. Sensitivity analysis of the effect of the numerical distribution of vehicles of given mass on the number of seriously and fatally injured drivers

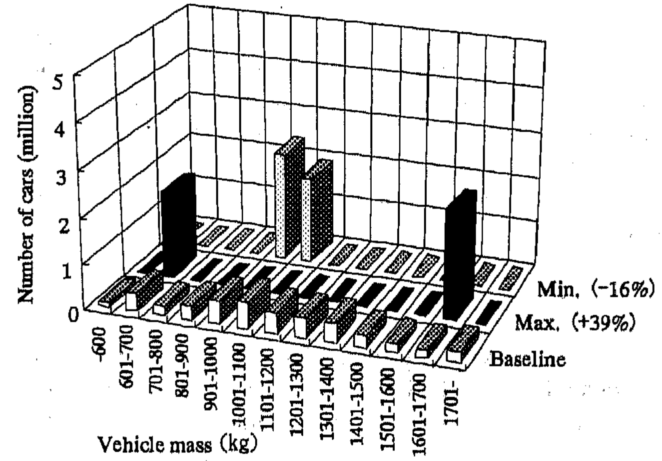


Fig. 31. The number of car distributions that gives the minimum or maximum number of serious and fatal injuries in head-on collisions

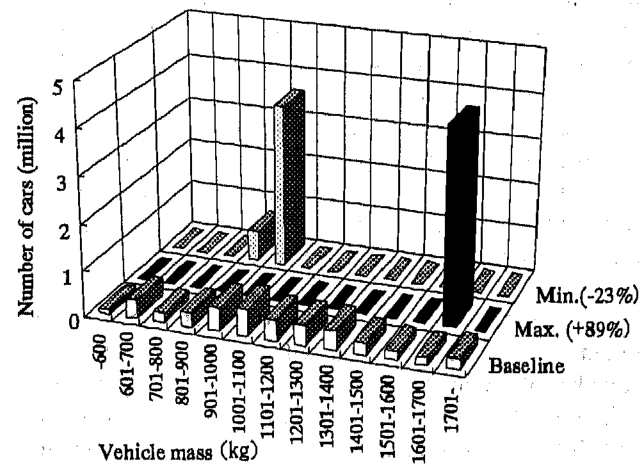


Fig. 32. The number of car distributions that gives the minimum or maximum number of serious and fatal injuries in side collisions

when the number of cars with mass from 700 kg to 1400 kg increases. Again it is found that the influence of the number of lighter cars is larger in head-on collisions, while that of the number of heavier cars is larger in side-impact collisions.

The total number of serious and fatal injuries per year is calculated by Eq.(16) as 2,645 in head-on collisions and 1,234 in side-impact collisions, using the actual numerical distribution of passenger cars as a baseline. The numerical distribution of cars which minimize or maximize the number of injuries in head-on and side collisions are given in Fig. 31 and Fig. 32 respectively. The number of serious injuries and fatalities takes the minimum value when the cars are concentrated in the mass region of 901-1100 kg and decreases injuries by 16% compared to the baseline distribution. On the other hand, the numerical distribution of cars in which the difference between car mass is large maximizes serious injuries and fatalities. In side collisions, the number of serious injuries and fatalities takes the minimum value when the cars are concentrated in the mass region of 801-1000 kg, however when all cars have an identical and large mass of 1601-1700kg, the number of serious injuries and fatalities takes the maximum value. The numerical distribution of cars which gives the minimum or maximum number of injuries is different in head-on and in side collisions because the effect of vehicle mass on the injury severity is different in each case.

DISCUSSION

The crash safety of a vehicle is estimated by a crash test with a fixed barrier in which the mass effect is small, so light and heavy cars are designed under the same safety concepts. The injury rate of the occupant is formulated by delta-V regardless of the vehicle mass. In head-on collisions delta-V depends on the car mass, the probability of injury in a collision can be expressed in terms of the vehicle mass as $a\{m_2/(m_1+m_2)\}^k$. As a result, the risk to the occupants in lighter cars is higher. The vehicle compatibility between light and heavy cars is discussed in many studies and some countermeasures are proposed —Heavier cars must be softer and have crush space for lighter cars, and lighter cars must be more stiff and have an optimum restraint system for the occupants. If the compatibility is accomplished for all vehicles, the injury rate will be expressed by delta-V in different ways for each car with different mass and the effect of car mass on injury rate will be smaller.

The effect of the numerical distribution of cars on the number of injuries is discussed. In head-on collisions, the number of serious and fatal injuries become larger when the difference between car mass is large. However, when the compatibility is accomplished and smaller cars become safer, the numerical distribution of cars does not affect the results so significantly as the analysis using the actual accident data of existing cars shows.

Mini cars occupies about one-eighth of the passenger car market in Japan. If all drivers of mini-passenger cars were to trade their current cars for small cars weighing 701-800 kg, it is estimated that the number of seriously and fatally injured drivers in head on collision accidents would fall by 108 from 2,645 per year, based on a calculation using the numerical distribution of passenger cars. Considering the economy of mini cars, it seems to be reasonable that mini cars should be made safer and compatible with large cars rather than replacing them with small cars.

CONCLUSION

This paper is concerned with the relationship between vehicle size, especially vehicle mass, and occupant injury, as determined from analysis of traffic accident data in Japan. The results are summarized as follows:

1. The fundamental formula of the probability of occupant injury in a head-on collision is expressed as $a\{m_2/(m_1+m_2)\}^k$ with k having a value of 2.67 for serious or fatal injury.

2. The struck vehicle mass influences occupant injury more significantly in head-on collisions than in side-impact collisions.

3. The occupants of smaller cars are likely to suffer chest and leg injuries when involved in frontal collisions.

4. As differences in mass in the vehicle population decreases, the number of serious and fatal injuries also decreases in head-on collisions.

REFERENCES

1. Evans, L.: Traffic Safety and the Driver,, Van Nostrand Reinhold, New York, 1991.
2. Hobbs, C., A., Williams, D., A. and Coleman D., J.: "Compatibility of Cars in Frontal and Side Impact," 15th ESV, 1996.
3. Muser, M., Krabbel, G., Utzinger, U., Prescher, V., Frei, P., Walz, F., Niederer, P., Kaeser, R.: "Optimized Restraint Systems for Low Mass Vehicle," SAE Paper 962435
4. Tarrière, C., Thomas, C., Trosseille, X,: "Frontal Impact Protection Requires a Whole Safety System Integration," 13th ESV, 1992.
5. Institute for Traffic Accident Research and Data Analysis, The Analysis Result of Relationship between Traffic Accident, Driver and Vehicle, 1996 (in Japanese).

6. Evans, L.: "Driver Injury and Fatality Risk in Two-Car Crashes versus Mass Ratio Inferred Using Newton Mechanics," Accident Analysis and Prevention, Vol.26, No.5, p.609-616, 1994.
7. Mizuno, K., Yonezawa, H.: "The Effect of Vehicle Mass in Car-to-Car Collisions," SAE Paper 960441
8. Watanabe,K., Yamaguchi, T.: "Analysis of Factors Affecting Dummy Readings in Side Impact Tests," 12th ESV, 1989.
9. Dieu, F., Riss, S.: "Computer Simulation Model for Side Impact Analysis," 14th ESV, 1994.
10. Evans, L. and Frick, M., C.: "Driver Fatality Risk in Two-Car Crashes – Dependence on Masses of Driven and Striking Car," 13th ESV, 1991.
11. Grime, G. and Hutchinson, P: "The Influence of Vehicle Weight on the Risk of Injury to Drivers," 9th ESV, 1980.
12. Wood, D., Mooney, S., Doody, M., Ríordáin, S.: "The Influence of Car Crush Behavior on Frontal Collision Safety and on the Car Size Effect," SAE Paper 930893.
13. Jones, I., S.: "The Effect of Restraint Use and Mass in "Downsized" Cars," SAE Paper 840199.
14. v Koch, M., Kullgren, A., Lie, A., Tingvall, C.: "Car Model Safety Rating - Further Development Using the Paired Comparison Method," 13th ESV, 1991.
15. Institute for Traffic Accident Research and Data Analysis, Statistics of Traffic (in Japanese).

970124

A Pneumatic Airbag Deployment System for Experimental Testing

Stefan M. Duma and Rodney W. Rudd
University of Virginia

Tyler A. Kress
University of Tennessee

David J. Porta
Bellarmine College

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper examines an originally designed airbag deployment system for use in static experimental testing. It consists of a pressure vessel and valve arrangement with pneumatic and electric controls. A piston functions like a valve when operated and is activated pneumatically to release the air in the tank. Once released, the air fills the attached airbag. The leading edge velocity can be controlled by the initial pressure in the tank, which can range up to 960 kPa. Three different test configurations were studied, which resulted in leading edge deployment speeds of approximately 20 m/s, 40 m/s, and 60 m/s. In experiments using this system, seven types of airbags were tested that differed in their material, coating, and presence of a tether. Data for each series of tests is provided. High speed video and film were used to record the deployments, and a pressure transducer measured the airbag's internal pressure. At 960 kPa initial tank pressure, the maximum leading edge velocity was 71 m/s. The maximum internal pressure of the airbag ranged from 88 kPa to 120 kPa. In addition to cost savings, the primary advantage of this system is its ability to quickly change the internal pressure while avoiding the hazards associated with the use of pyrotechnics. However, the leading edge velocity was typically below the average published values for common airbags. This discrepancy may be corrected in future designs by changing the orientation of the valve portion of this system.

INTRODUCTION

With increased interest in refining the airbag, the need arises for an inexpensive and safe means to test different airbag design and construction features. For example, if several airbag design characteristics are to be examined at several different pressures, the test matrix will require a large number of inflators to properly test each iteration. Given the significant cost of pyrotechnic inflators, alternate methods of testing should be considered. With the reusable pneumatic system described in this paper, the only two primary costs will be the bags themselves and the initial machining. Moreover, the increased risk of storing the pyrotechnic devices and the adverse effects of handling the combustion products after airbag deployment can be avoided [1].

The concept of a reusable compressed gas inflator for experimental testing is not new. Although a different design was used, previous laboratory studies note the advantages this type of system offers [2,3]. For example, the ability to readily change deployment parameters. In those studies, fast acting valves were used to control the compressed air as compared to the piston-valve arrangement presented here. Additionally, Melvin et al. (1993) compensated for the flow characteristics of high temperature pyrotechnic inflation gases by combining gases of various densities. This system uses compressed air, but could be adapted to use the combined inert gas approach, with the only drawback being increased cost.

REVIEW OF THE DEPLOYMENT SYSTEM

The airbag deployment system is shown in Figure 1 with a functional schematic detailed in Figure 2. An air compressor is used to provide the initial pressure for the 37 L pressure vessel. The horizontal structure comprises a large piston which acts as the valve. An aluminum alloy was the primary construction material.

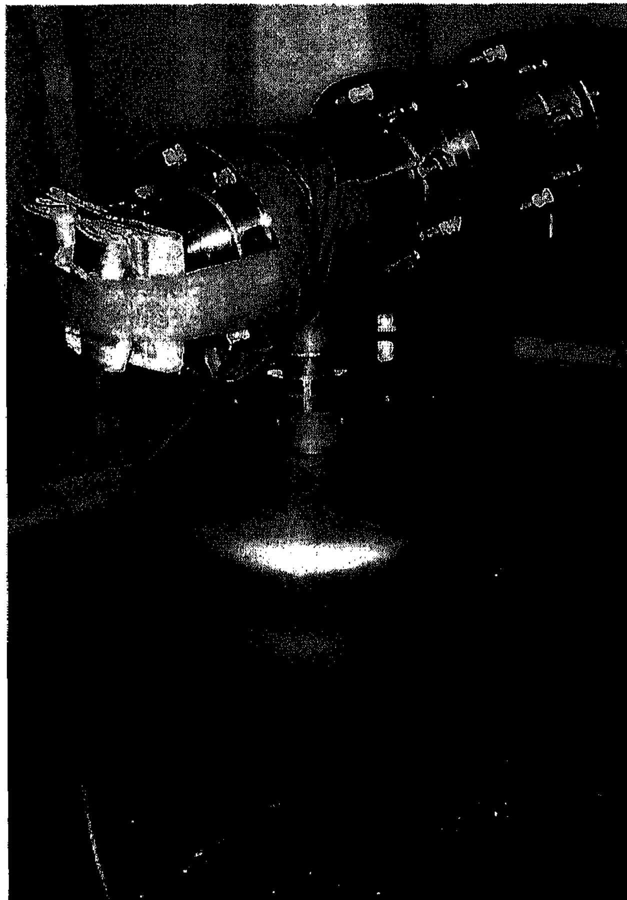


Figure 1: Airbag Deployment System

AIRBAG DEPLOYMENT - The airbags used with this system are modified so that they may be attached directly to the front of the system via the manifold as shown in Figure 3. On the airbag side of the manifold, a pressure transducer was installed to measure the bag's internal pressure throughout each deployment. A strap is used to hold the bag in place and model the performance of the airbag's cover as seen in an automobile. Before the tank is pressurized, the piston is placed in the forward position, thereby creating a seal between the tank and the airbag. Once the bag is attached and the tank is pressurized, the system is ready for deployment. In order to deploy the bag, a solenoid is activated which in turn opens the airway and supplies air to the middle portion of the cylinder as shown in Figure 2. This pressure forces the piston to the right allowing the air in the tank to fill the airbag. The piston assembly remains in the right position until reset for the next test.

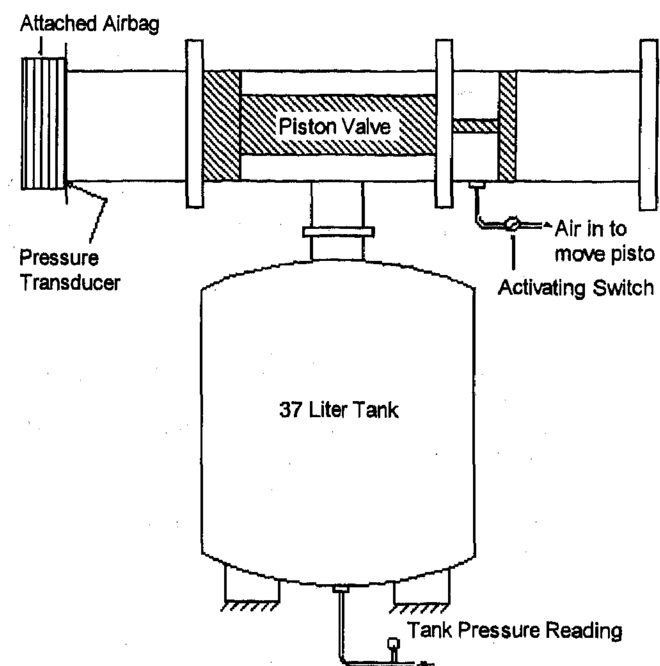


Figure 2: Functional Schematic of the Airbag Deployment System

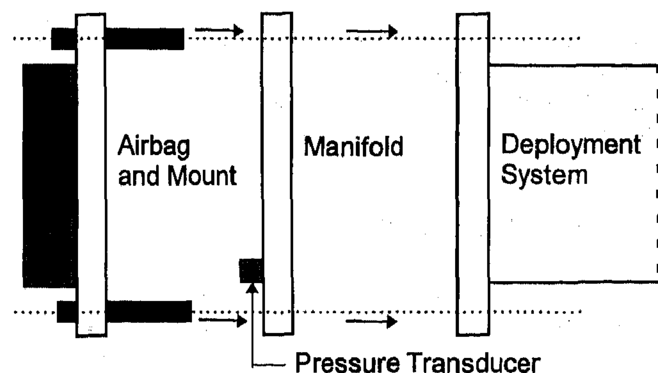


Figure 3: Assembly Diagram Depicting the Airbag Mount Attaching Through the Manifold to the Deployment System

DEPLOYMENT VARIANCE - The leading edge velocity can be varied by adjusting two factors: the internal tank pressure and the strap. First, a pressure gauge is attached to the line that feeds the deployment tank. From this gauge and data from previous tests, a reasonable estimate can be made of the future deployment speeds. The second variable is the presence of a cover. For the higher deployment speeds, a breakable strap is installed to allow a buildup of pressure behind the bag just before release. Three configurations have been established as detailed in Table 1.

Table 1: Deployment Configurations for Three Leading Edge Velocity Approximations

Test Config.	Initial Tank Internal Pressure	Cover/ Strap Present	Approximate Leading Edge Velocity
1	405 kPa	No	20 m/s
2	960 kPa	No	40 m/s
3	960 kPa	Yes	60 m/s

To demonstrate the ability of changing the leading edge velocity, all three configurations were tested using only one type of airbag. These tests are plotted at each of the three configurations in Figures 4, 5, and 6. Here it is shown that the leading edge velocity goals of 20 m/s, 40 m/s, and 60 m/s were met. Configuration 3 (60 m/s) was selected for the experimental tests since it was most similar to the production models.

SAFETY PRECAUTIONS - To avoid possible injury and lost time from an inadvertent deployment, the system has several important safety features. A pin, which is inserted between each test, is positioned perpendicular to the piston through two stays. With the pin installed, the piston cannot move and therefore, the bag cannot deploy. The secondary safety features are combined with the controls. The activation switch must be reset between each deployment and runs from a separate power supply to further decrease the risk of accidentally activating the system. Finally, an alarm is automatically sounded once power is supplied to the trigger system to warn of the imminent deployment.

EXPERIMENTAL RESULTS

Table 2 describes the seven types of airbags that were used in this test series. Material A was a high denier fabric while B, C, D all having subsequently lower deniers. All airbags were standard 60 L driver-side airbags, modified accordion fold with two windshield facing vents, that varied only in their coating, material, and presence of a tether as noted. Each had been modified to facilitate the mounting procedure by installing a new mounting flange; however, the modification was minimal and had no effect on the deployment characteristics.

Table 2: Airbag Construction Specifications

Airbag Type	Material Type	Coating Type	Tether/ Non-tether
1	A	A	N
2	B	B	N
3	B	B	T
4	B	none	N
5	B	none	T
6	C	none	N
7	D	none	N

LEADING EDGE VELOCITY COMPARISONS - The leading edge velocities for airbags deployed with this system were determined using high speed video and film. For each of the seven types of airbags, three deployments were conducted at configuration 3, which provided for the maximum deployment capability of the system. The leading edge velocity profiles of airbags with similar construction features, presence of tether and folding pattern, as those used in this test were gathered from published data [4,5]. For example, airbag type 1 has a tether and uses the modified accordion fold. Ideally all variables would be held constant, but the analysis is limited on the availability of published data. From the literature, it was determined that the 1990 Oldsmobile Toronado, 1990 Chevrolet Camaro, and the 1993 Chevrolet Caprice also have a tether and modified accordion fold. Therefore, in order to compare similar airbag types using the deployment system, three test profiles of airbag type 1 are plotted with the three production models as shown in Figure 7. This same technique was used in Figure 8 through Figure 13 for the other six airbags.

The average maximum leading edge velocity for 1990 airbags was 64.3 m/s and is given on these plots as well [5]. It should be noted that the comparisons in Figures 7-13 only consider the bag characteristics and not the inflator properties.

DISCUSSION

The leading edge velocity profiles in Figure 7 show that the deployment system mimics the three General Motor's models with similar bag characteristics reasonably well. Two of the peak velocities for the three experimental tests exceed the 1993 Caprice, which has the maximum velocity of the three comparison airbags. The leading edge velocity profiles match well for each airbag type. Figure 8 depicts the leading edge velocities of airbag 2 showing consistent behavior with the same models used in Figure 7. The identical three models are used again due to the fact that airbags 1 and 2 differ in material and coating, but both are non-tethered and have the same folding pattern as the comparison models.

Airbags 3 and 5, in Figure 9 and 11, are similar to their respective counterparts, but have a slower rise time and do not reach the peak velocity of the 1993 Ford Taurus. Although both the 1993 Saturn and 1993 Ford have a tethered airbag with a modified accordion fold, the differences in the inflators and material are two reasons that account for the discrepancy between these comparisons.

The experimental tests exhibited a decrease in velocity from 75 mm to 110 mm deployment distance in Figures 7, 10, 12, and 13. This observation comprised airbag types 1, 4, 6, and 7 which show no significant similarities, thus implying that this behavior is a random aspect of the deployment system rather than an attribute of any one particular airbag characteristic.

The sample industry airbags chosen, with the exception of the 1990 Ford Taurus, do not exceed the average maximum leading edge velocity of 64.3 m/s for

airbags in 1990 vehicles. Therefore, even though the deployment system approaches similar velocity curves, its maximum velocity is still below the average maximum speed.

CONCLUSIONS

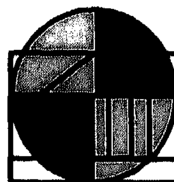
The pneumatic airbag deployment system proved to be a good tool for static airbag deployments with various leading edge speeds. The maximum velocity is easily adjusted by changing the initial tank pressure, but the velocity is limited in this design to approximately the 1993 industry average of 64.3 m/s. The velocity could be increased by designing the deployment system with a straight path from the tank to the airbag as opposed to the right angle in the current design.

The low cost of this system and its increased safety versus pyrotechnic airbags are two additional advantages. Two disadvantages of this system are the use of a strap to simulate the cover, and the inability to correlate the pressure curves with manufacturer's airbag tank tests. Future designs should incorporate a cover that reflects the exact characteristic of an actual airbag cover. This would insure a more realistic duplication of a deploying airbag.

REFERENCES

1. Chan T, et. al: Exposure Characterization of Aerosols and Carbon Monoxide from Supplemental Inflatable Restraint (Automotive Airbag) Systems. J. Aerosol Sci., Volume 20, Number 6, pp. 657-665, 1989
2. Melvin JW, Horsch JD, McCleary JD, et al: Assessment of Airbag Deployment Loads with the Small Female Hybrid III Dummy. SAE Paper 933119, 37th Stapp Car Crash Conference Proceedings, San Antonio, Texas, pp. 121-132, 1993
3. Horsh J, Lau I, Andrzejak D, et al: Assessment of Airbag Deployment Loads. SAE Paper 902324, 34th Stapp Car Crash Conference Proceedings, Orlando, Florida, pp. 267-288, 1990
4. Powell MR, Lund AK: Leading Edge Deployment Speeds of Production Airbags. SAE Paper 950870, International Congress and Exposition, Detroit, Michigan, 1995
5. Sullivan LK, Kossar JM: Airbag Deployment Characteristics. National Highway Traffic Safety Administration, DOT HS 807 869, February, 1992

CONTACT FOR ADDITIONAL INFORMATION



University of Virginia
Automobile Safety Laboratory
1011 Linden Avenue
Charlottesville, VA 22902
Phone: (804) 296-7288
Fax: (804) 296-3453

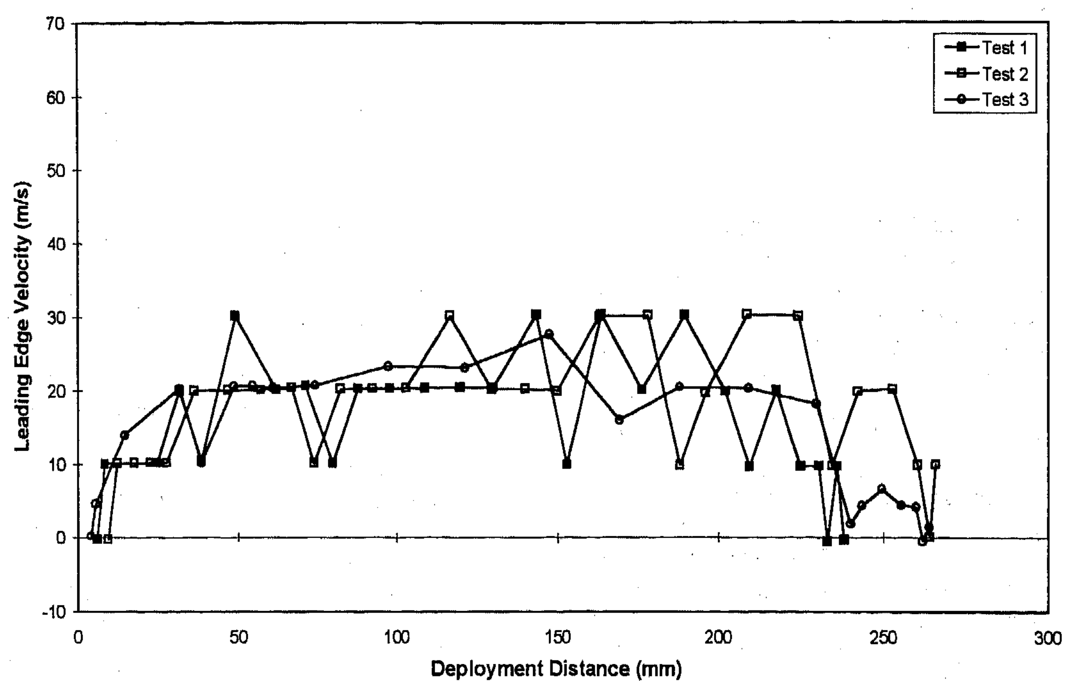


Figure 4: Air Bag Type 6 at Configuration 1 with 405 kPa Internal Tank Pressure, No Cover (Goal of 20 m/s)



Figure 5: Air Bag Type 6 at Configuration 2 with 960 kPa Internal Tank Pressure, No Cover (Goal of 40m/s)

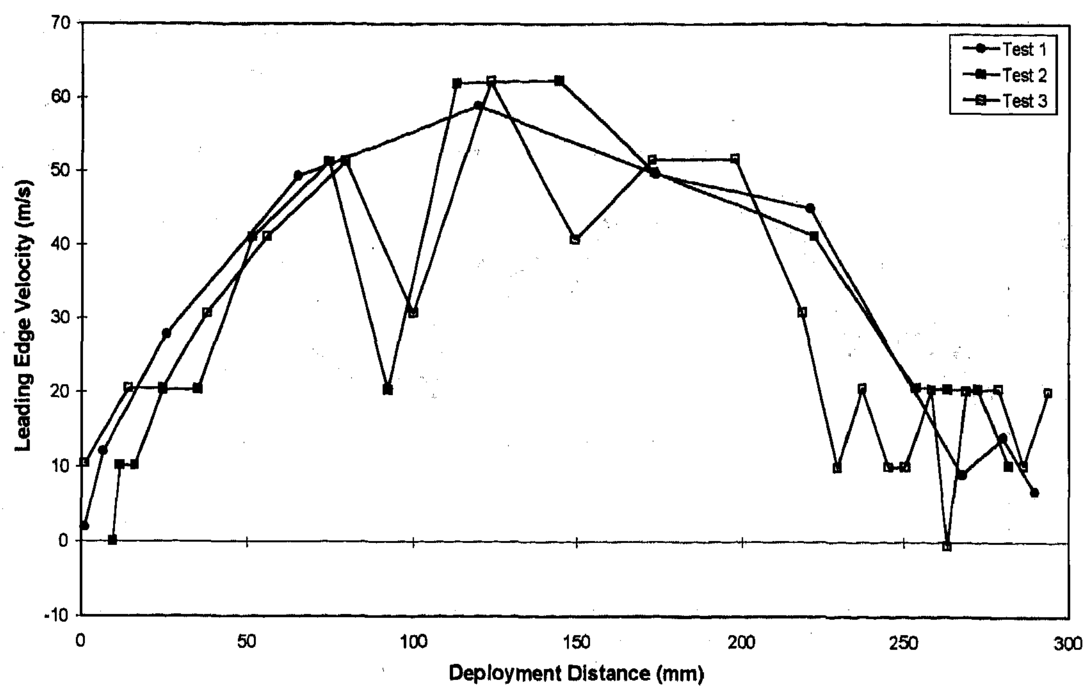


Figure 6: Air Bag Type 6 at Configuration 3 with 960 kPa Internal Tank Pressure, Cover Present (Goal of 60m/s)

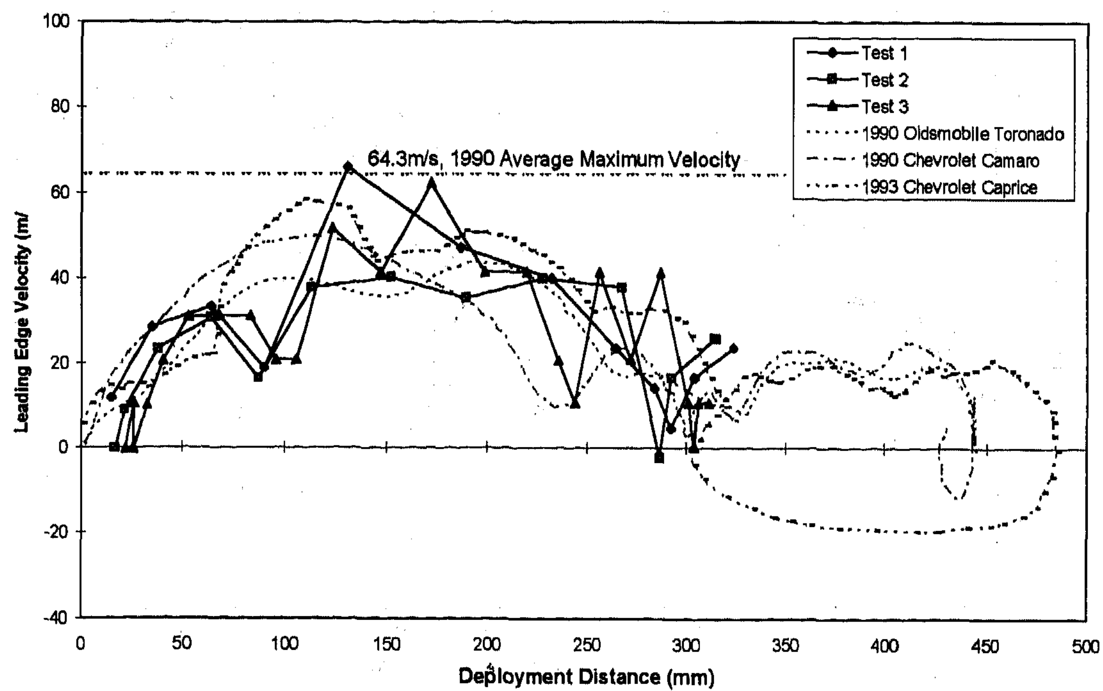


Figure 7: Leading Edge Velocity Profiles for Air Bag Type 1 Configuration 3 compared to similar production models

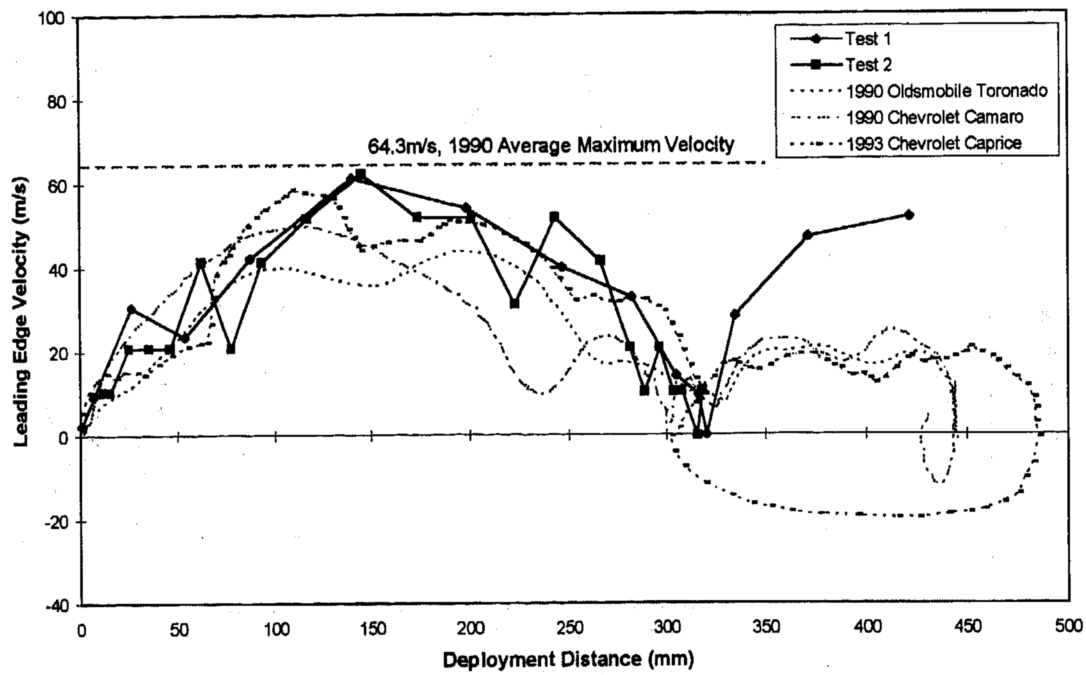


Figure 8: Leading Edge Velocity Profiles for Air Bag Type 2 Configuration 3 compared to similar production models

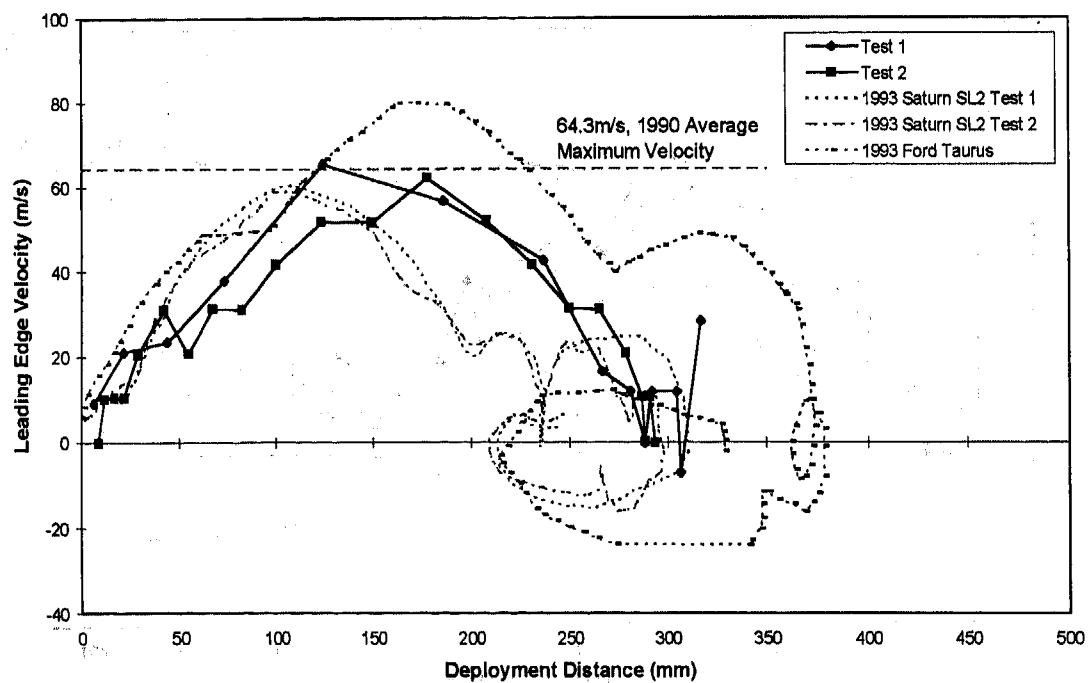


Figure 9: Leading Edge Velocity Profiles for Air Bag Type 3 Configuration 3 compared to similar production models

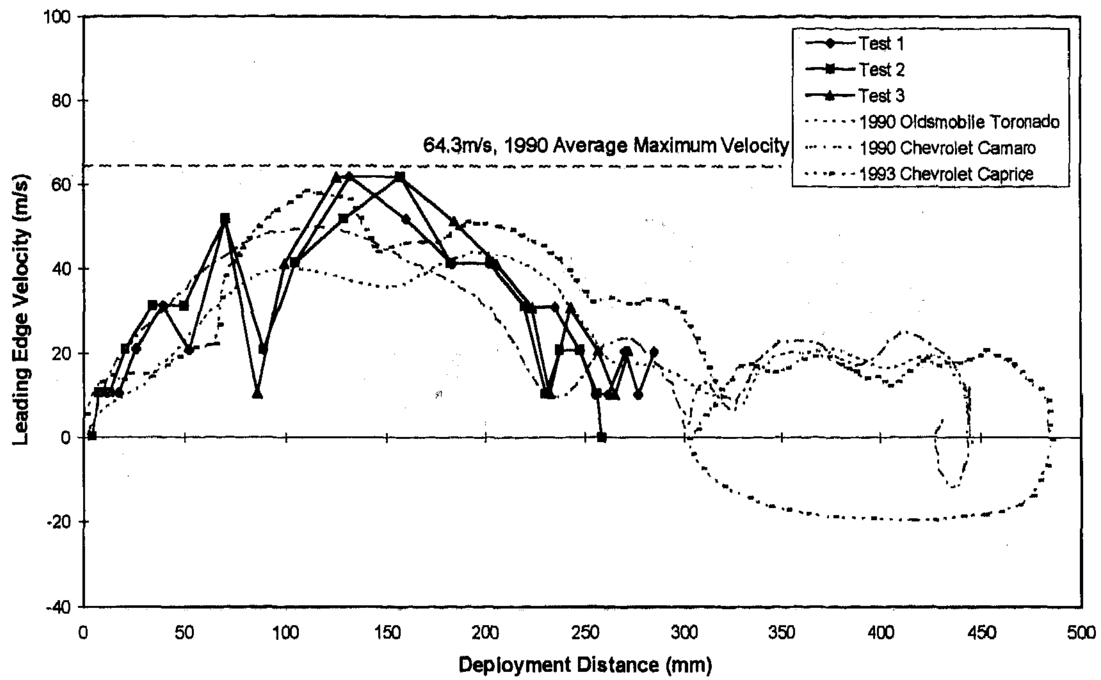


Figure 10: Leading Edge Velocity Profiles for Air Bag Type 4 Configuration 3 compared to similar production models

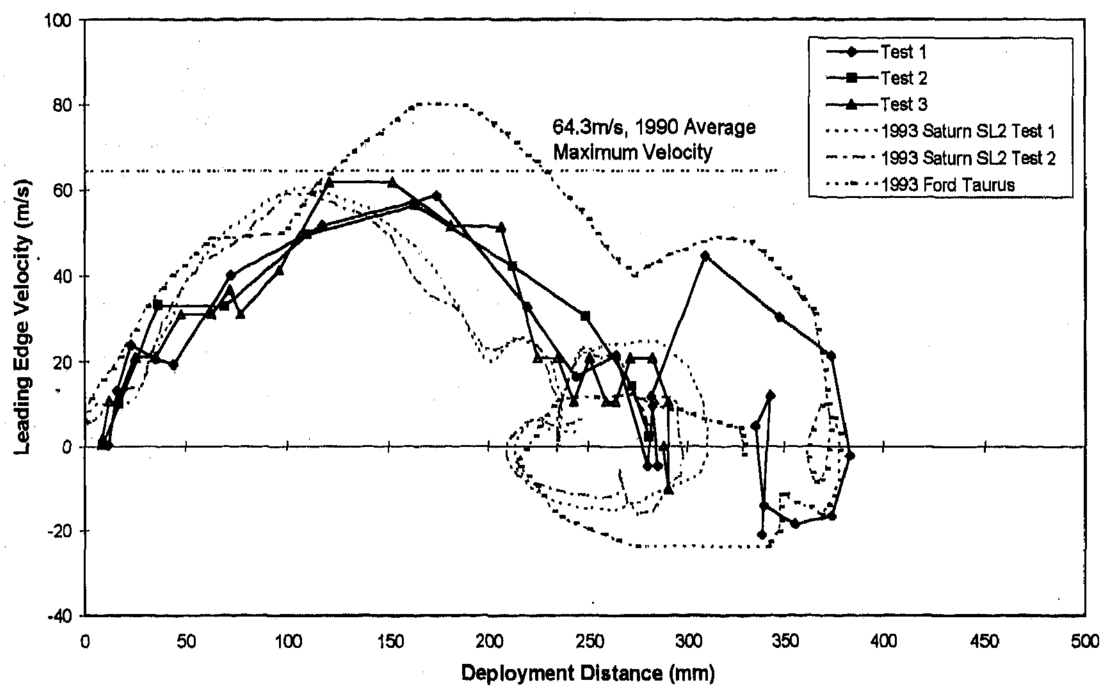


Figure 11: Leading Edge Velocity Profiles for Air Bag Type 5 Configuration 3 compared to similar production models

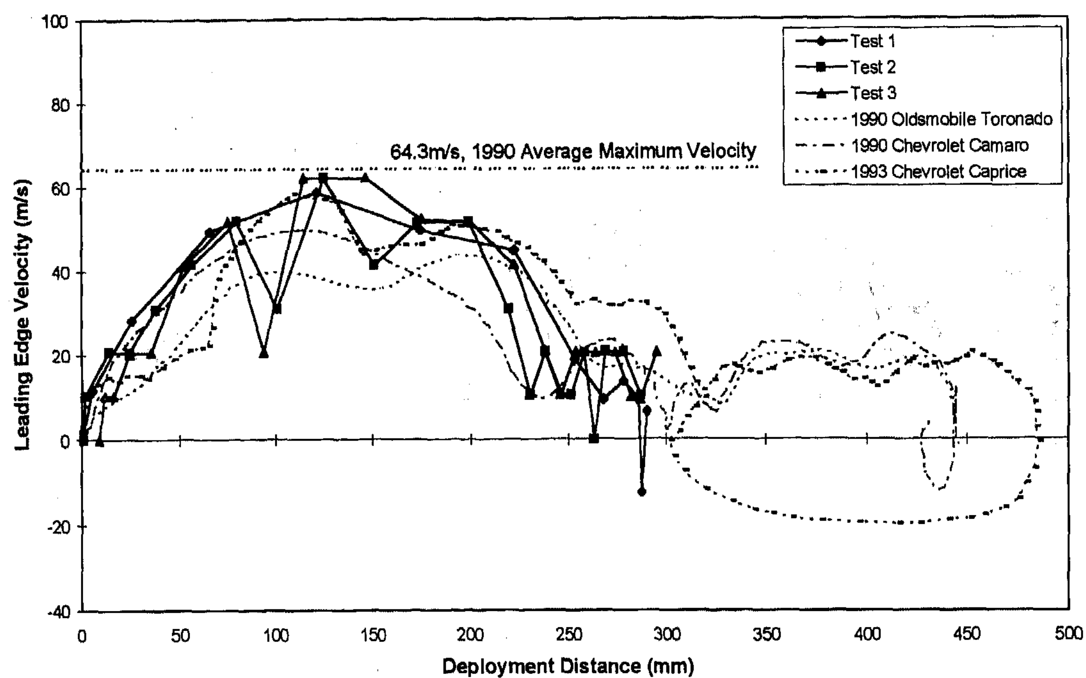


Figure 12: Leading Edge Velocity Profiles for Air Bag Type 6
Configuration 3 compared to similar production models

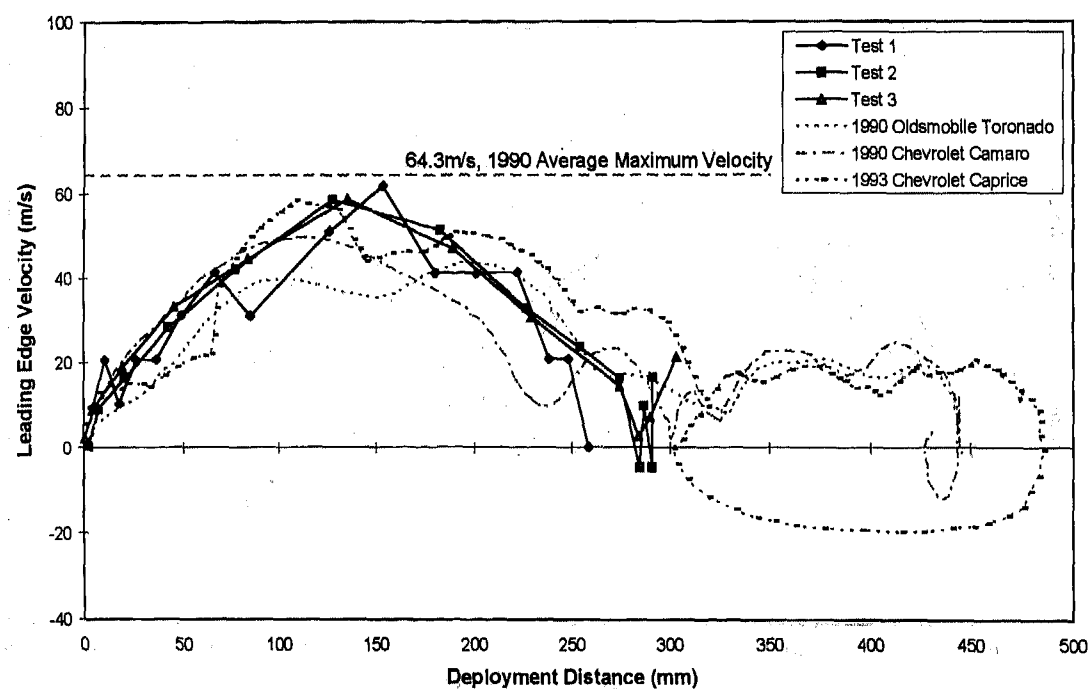


Figure 13: Leading Edge Velocity Profiles for Air Bag Type 7
Configuration 3 compared to similar production models

970127

Simulation of Seat Integrated Side Airbag Deployment

Anand S. Tanavde and Krish Gudipaty
Engineering Systems International Corp.

Srini Vaidyaraman
TRW Vehicle Safety Systems Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Modeling and simulation are becoming increasingly popular to help develop restraint systems and enhance the value of the design. Various types of tests are being developed and used extensively in evaluating and optimizing the performance of restraint systems by both OEM's and system suppliers. The use of analytical tools along with laboratory testing can help improve the effectiveness and the speed of developing and implementing a restraint system to maximize the occupant protection.

The side airbags are used in conjunction with energy absorbing foams to manage the energy transfer to the occupant from the structural intrusions in a side impact. The requirement of deployment times are shorter in side impact than in a frontal impact due to the limited space between the occupant and door structure. In the design and development of seat integrated side airbag modules, it is important to predict deployment time and pressure which are affected by the characteristics of the seat-fabric material and tear seam. The PAM-SAFE numerical simulation tool allows us to simulate the tearing of seat fabric and subsequent deployment of the side airbag and its complex interaction with the module and the seat foam.

INTRODUCTION

Modeling and simulation are increasingly used to help develop restraint systems and enhance the value of the design. Parametric studies with a simulation model are regularly used by both OEM's and system suppliers and are effective in evaluating and optimizing the performance of restraint systems.

Several automobile manufacturers are implementing side airbags in addition to energy absorbing padding and structures to provide occupant head and thorax protection and reduced pelvic acceleration for the occupant during side-impact crashes. The development of the side airbag system is a considerably more difficult problem than the frontal airbag system due to the limited space and faster airbag deployment time requirements. Hence it is essential to learn as much about the physics as early as possible in the development of side airbags. The seat mounted side airbag which deploys through a seat foam and fabric and gets in-position between the occupant and the door and B-pillar of the vehicle, is one design to help protect the occupant's torso and head.

Proper deployment of the airbag is crucial in providing protection to the occupant. One of the factors controlling airbag deployment is the characteristics of the tear seam of seat fabric and the airbag cover structure. Hence, the measurement of pressure distribution on the fabric cover with tear seam or a door structure is important for the design and development of Side Impact Supplemental Inflatable Restraint systems. In the past, a methodology was developed at TRW-VSSI to experimentally measure deploying pressure distribution on the inner surface of a passenger-side airbag door [1].

The objective of this study was to develop a methodology to numerically simulate a full deployment of the folded side impact airbag including the breakage of the seat-fabric at the tear seam and the module deformation. Explicit dynamic finite element code PAM-SAFE [2] was used to simulate TRW dynamic test of a folded side airbag

interacting with the airbag module and seat-fabric. The main purpose was to analyze the effects of the bag pressure and bag unfolding interaction with a seat-fabric resulting into a tearing failure along the tear-seam. This type of simulation allows a detailed study of the separation process of the seat-fabric, module deformation, loading on the seat-frame, and the side airbag deployment kinematics. The model developed in this study could be used to estimate dynamic loading on the seat-fabric for different materials.

AIRBAG MODELING

The airbag mesh was generated from CAD drawings of actual design. The airbag dimensions were measured on the airbag in the flat configuration with the side tucked in once from each side. The mesh for the flat configuration of the bag was oriented to match with the position of the tethers. The mesh density was chosen such that there are at least 3 elements between two folds. An accordion folding pattern was incorporated. PAM FOLDER was used to fold the flattened airbag in a folded state. The distance between the folded layers is 0.0006m. The final airbag mesh consisted of approximately 1800 membrane elements.

The jet effect [2, 3] was considered to simulate the dynamic effect of the mass flow during the first 10 ms of the inflation. The leakage was defined by a pressure dependent outflow rate curve. The airbag fabric properties are given in the following table.

TABLE 1

AIRBAG FABRIC PROPERTIES	
Mass density	774.3 Kg/m^3
Poisson's ratio	0.20
Young modulus	1.e6 Pa for parent sheet 2.01E+8 Pa & 2.24E+8 Pa for fiber layers
Thickness	0.0003 m

As pointed out earlier, the mesh was based on an actual bag in a folded state. It is generally time consuming to design a mesh which can be folded easily to obtain actual folded bag configuration and which can be packaged in the limited space available. In order to save time in meshing and to obtain a properly folded state without penetration of the airbag fabric layers, an initial metric method [3] may be used. However, in the author's opinion the initial metric method does not capture correctly the airbag kinematics and pressure, which determine the time of breakage of the tear seam.

The airbag module consists of a rectangular shaped module which houses the inflator. Part of the folded bag is attached to the module. The following table depicts module material properties:

TABLE 2

Module Properties (steel)	
Thickness	0.001m
Mass density	7850 Kg/m^3
Poisson's ratio	0.30
Young modulus	207.E+09 Pa
Yield stress	500.E+06 Pa

TRW DYNAMIC SIDE AIRBAG DEPLOYMENT TEST

A special, dynamic side airbag test was devised to study the strength of various fabric materials. This test was simulated in PAM-SAFE to calibrate tear seam properties. The test setup as is shown in figure 1, consists of calipers across which the fabric sample is constrained such that it covers the top of the folded airbag. The reaction forces on the calipers due to the deployment forces exerted on the fabric are measured with a load cell.

The FE mesh for the seat-fabric contains approximately 500 shell elements. It was restrained on both sides along the edges as per the test set-up. The fabric material model uses only the elastic part of the force Vs deflection curve obtained from the actual seat-fabric material test. The following table shows the seat-fabric material properties:

TABLE 3

Airbag Seat-fabric Material Properties	
Mass density	400 Kg/m^3
Poisson's ratio	0.20
Young modulus	7.0E+06 Pa
Thickness	0.002 m

CALIBRATION OF THE TEAR SEAM

In the test, release occurs along a pre-defined tear-seam in the seat-fabric. In order to model this tearing effect, a discontinuity or a gap of 1 mm is

introduced at the location of the tear seam. The seat-fabric material strength is distributed along this tear-seam using nonlinear bar elements, one per each node of the element. The strength of each nonlinear bar element is calibrated according to the force deflection curve (Fig. 2) obtained from the test data. The fabric sample used for the tension strength test is 0.008 m thick, 0.1016 m wide and 0.1524 mm in length. The actual seat-fabric used in the calibration test, is 0.344 m along the tearing line with a same thickness. Thus the experimental curve has to be scaled by an appropriate scale factor. As the end nodes share only half the nodal forces, the scale factor for these nodes should be halved.

The airbag was attached to the housing through the nodal constraints along the bottom peripheral nodes of the module . Then, the whole assembly was fixed at bottom. In order to get a correct kinematics of the airbag interacting with the module and the fabric cover, proper definition and working of contact interfaces are important.

KINEMATICS OF DEPLOYMENT FOR THE SIDE AIRBAG

The kinematics of the unfolding sequence from the front side, during first 5 ms is shown in Figures 3. Figures 4 and 5 show the side and top views of the same sequence. The comparison of these results with the test data show very good correlation. The simulation is able to capture the bulging of the fabric cover during 0 to 3.5 ms. The breakage of the fabric cover occurs at approximately 3.5 ms. The airbag completely breaks out of the cover at about 6 ms. The airbag gets fully inflated at about 8 ms.

The side airbag deployment was simulated with and without the jet effect. The reaction forces from the simulations are compared with the test data in figure 6. When the jet effects are considered, tear-seam release occurs at 3.5 ms at the reaction force level of 1800 N., which compares very well with corresponding test values of 3.4 ms and 1860 N. However, without the jets a tear-seam release is delayed to 4 ms at the force level of 2486 N. This behavior is quite reasonable and can be explained as follows. The effect of the jet is to provide additional concentrated pressure forces on the parts of the airbag where it directly impinges. This causes early local release of the seat-fabric at a lower force level. When the jet effect is absent, the airbag pressure forces are more evenly distributed on the seat-fabric, allowing it to stretch more, and causing delayed release at a higher force level.

The module starts deforming due to the airbag pressure forces; this phenomenon is called fish mouthing and is captured very well as depicted in Fig. 7. The airbag max. module deformation occurs at 4 ms.

AIRBAG MODULE INTEGRATED WITH A SEAT STRUCTURE

The side bag validated with a dynamic test was packaged inside the seat with the seat foam surrounding the module (see Fig. 9) The seat fabric covering the seat foam was modeled using the same fabric material as described in Table 3, above. A similar methodology (as described earlier) was used to model the tear-seam in the seat.

Developing a suitable seat model was an extremely complex task; especially the modeling of seat-foam geometry. Only the parts of the foam immediately surrounding the airbag module are modeled as deformable. The rest of the parts of the seat are assumed to be rigid. The stress-strain curve for the seat-foam is presented in figure 8. As the punching forces due to the deploying airbag are locally concentrated, only a part of the seat was modeled and constrained at the boundaries as shown in the figure 9.

Figure 10, depicts the simulation of the kinematic sequence of side airbag deployment out of the seating structure.

It was observed from the simulation results that at about 3.5 to 4 ms the fabric tears off and the airbag comes out of the seat fabric. At about 8 ms the airbag is fully inflated. Various contacts; self-contact for the bag, contact between the bag and the module, the bag and the foam, the foam and the fabric work reasonably well. The contact between the bag and the module may be tuned further to enhance the contact behavior at the edges of the module. In this simulation, no attempt is made to correlate with the test data which in any case was not available for this complex case. In the future, the results of this simulation could be compared with the test data.

CONCLUDING REMARKS

TRW's dynamic test of the side impact airbag interacting with the seat-fabric and deformable module was successfully simulated with the finite element code PAM-SAFE. The kinematics of the

unfolding airbag, the plastic deformation of the module and the release along the tear-seam is captured very well in this simulation. This simulation was used to calibrate tear seam characteristics which were further implemented in the model of a folded airbag module integrated inside of the seat.

A deployment of the side airbag out of the seat is also successfully simulated using PAM-SAFE. In this simulation, a complex interaction amongst the bag, the module, the seat foam, seat fabric is captured fairly well.

A simple thermodynamic approach was adopted to simulate the dynamic effect of the gas pressure onto the airbag seat-fabric and the module. This approach illustrates the fact that the complex effects due to the gas flow dynamics in the folded airbag and its interaction with the foam and the seat-fabric during deployment, module deformation, and subsequent tearing failure of the seat-fabric material can be captured by a simple uniform pressure gas model coupled with a jet technique using PAM-SAFE.

As the jets exert additional pressure loading concentrated locally on the parts of fabric where the jets strike, the fabric can tear at an earlier time and at a lower pressure force magnitude than in the case without the jets effects. Hence it is important to consider the effect of the jets in the simulation where the fabric tearing phenomenon needs to be simulated.

The methodology developed in this study could be used to predict the deploying forces or the loading condition on the airbag fabric cover and the kinematics of bag deployment.

The seat integrated airbag model developed above can be used to study the performance of various tear-seam strengths and patterns.. The model can also be used in a full side impact, sled or barrier simulation to evaluate the seat integrated side airbag system and its effectiveness in reducing occupant injuries.

ACKNOWLEDGEMENT

Our thanks are due to Dr. Ilango Shanmugavelu of Johnson Controls Inc. for providing foam data.

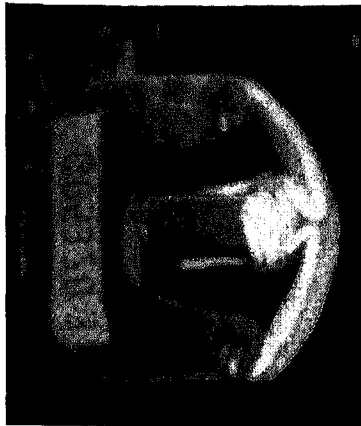
REFERENCES

1. Bayley, G. S. et. al., “Experimental Measurements of Deploying Pressure Distribution on a Passenger-Side Airbag Door”, SAE Paper 950339

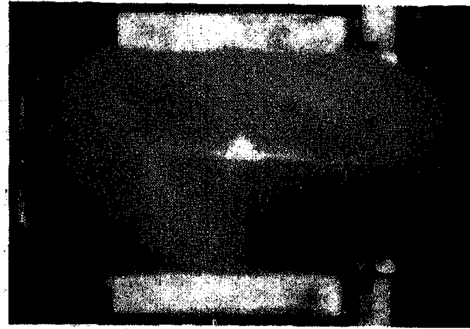
2. PAM-CRASH/PAM-SAFE User's Manual, Version 96, Engineering Systems International, SA, Rungis, France, 1996.

3. Groenenboom, D. et al., "A Diffusive Gas Jet Model in PAM-SAFE for Airbag Inflation", SAE Paper 930238, March 1993

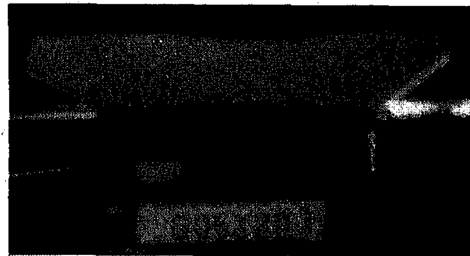
4. Tanavde, A. S. et. al., “Airbag Modeling using Initial Metric Methodology”, SAE Paper 950875, March 1995.



TOP VIEW



FRONT VIEW



SIDE VIEW

FIGURE 1. DYNAMIC SEAT FABRIC TEST SET-UP

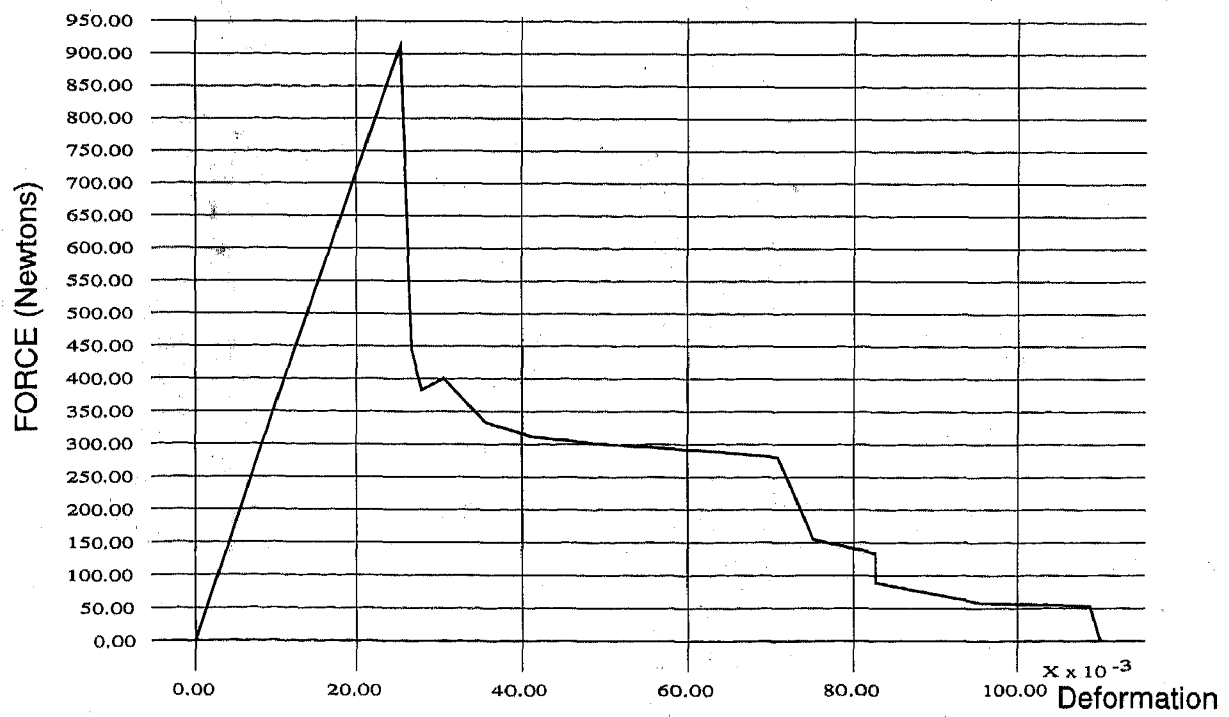


FIGURE 2. TENSION TEST ON A FABRIC COVER SAMPLE

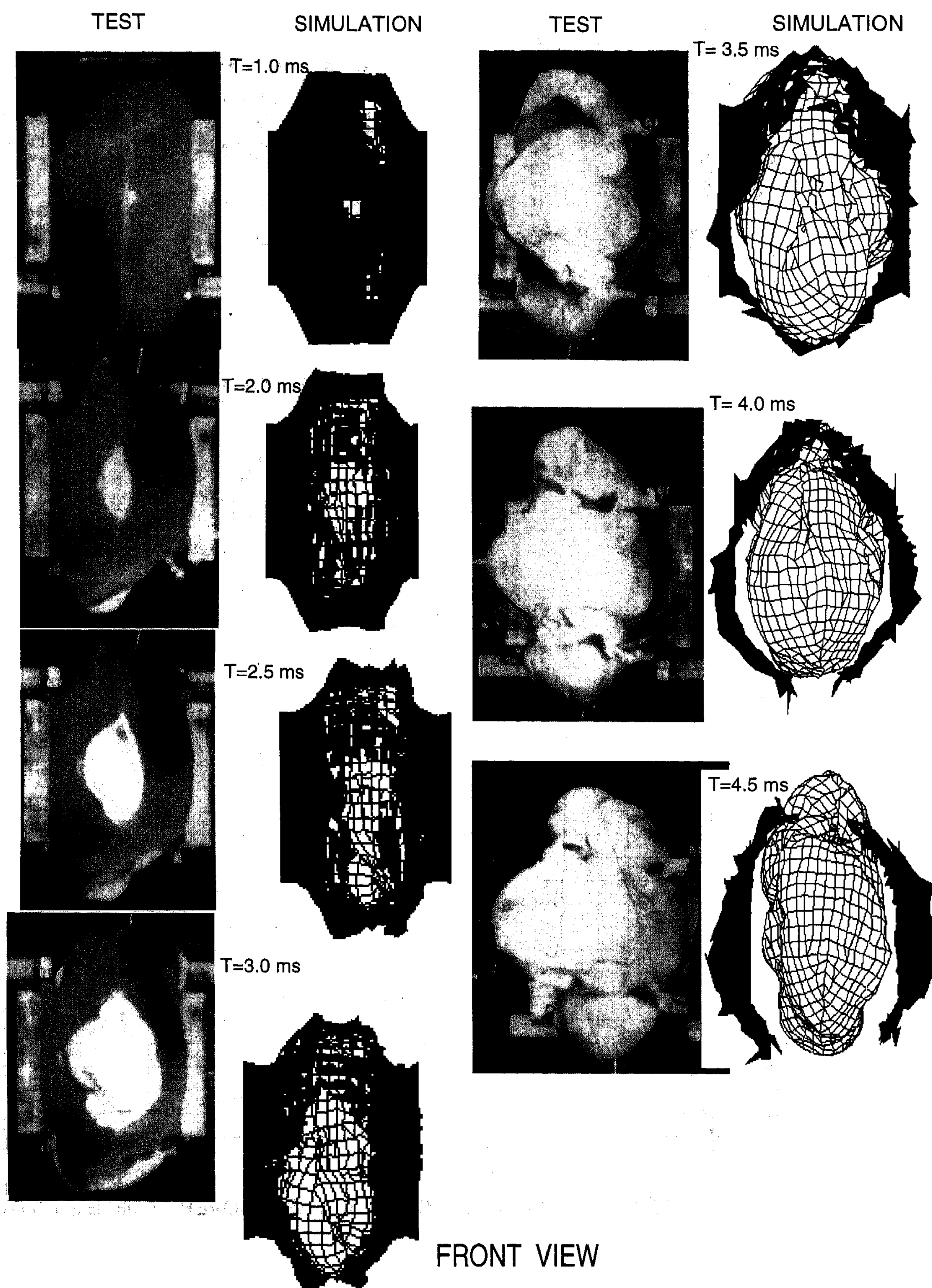


FIGURE 3. COMPARISON OF TEST AND SIMULATION FOR THE SIDE AIRBAG DEPLOYMENT

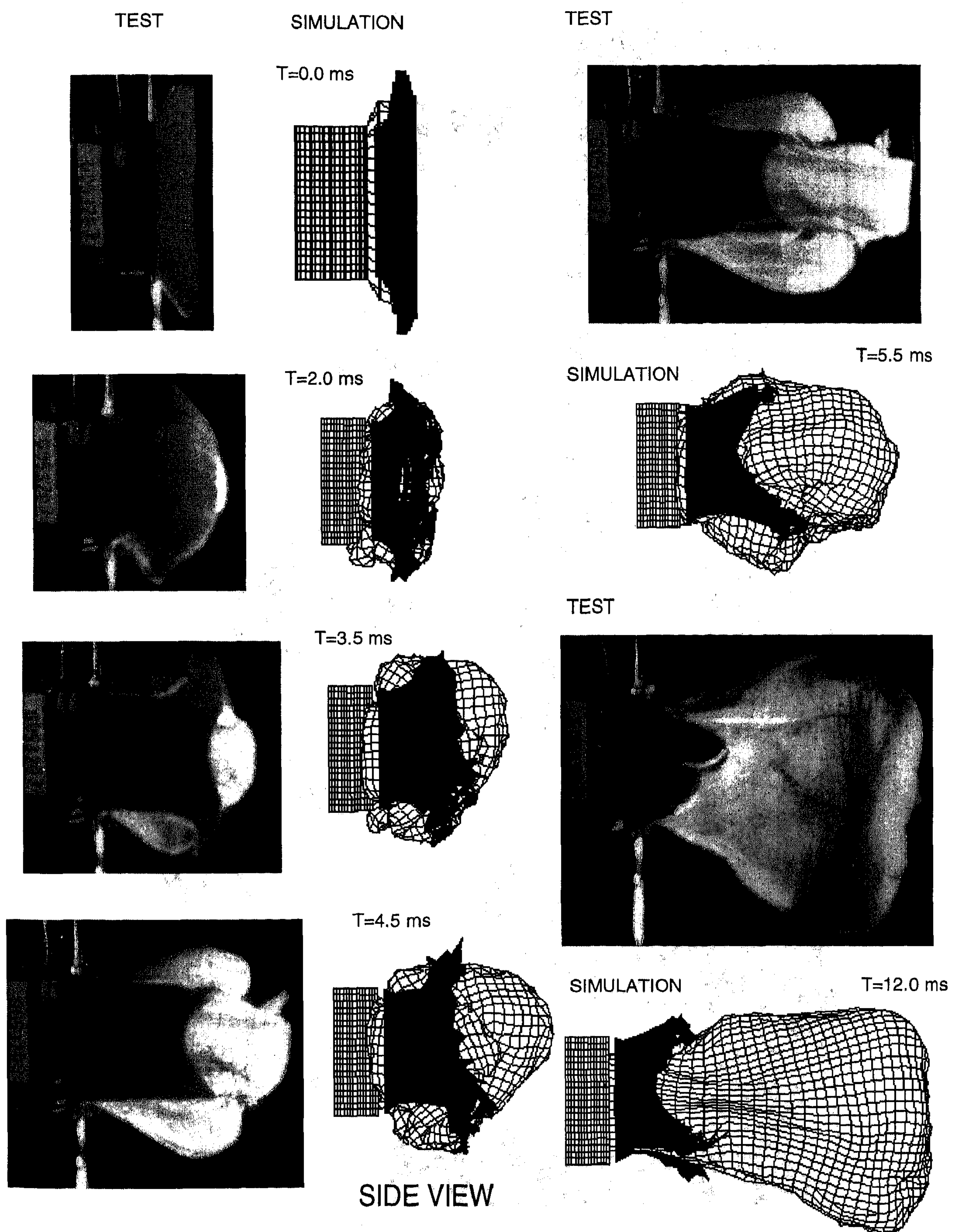


FIGURE 4. COMPARISON OF TEST AND SIMULATION FOR SIDE AIRBAG DEPLOYMENT

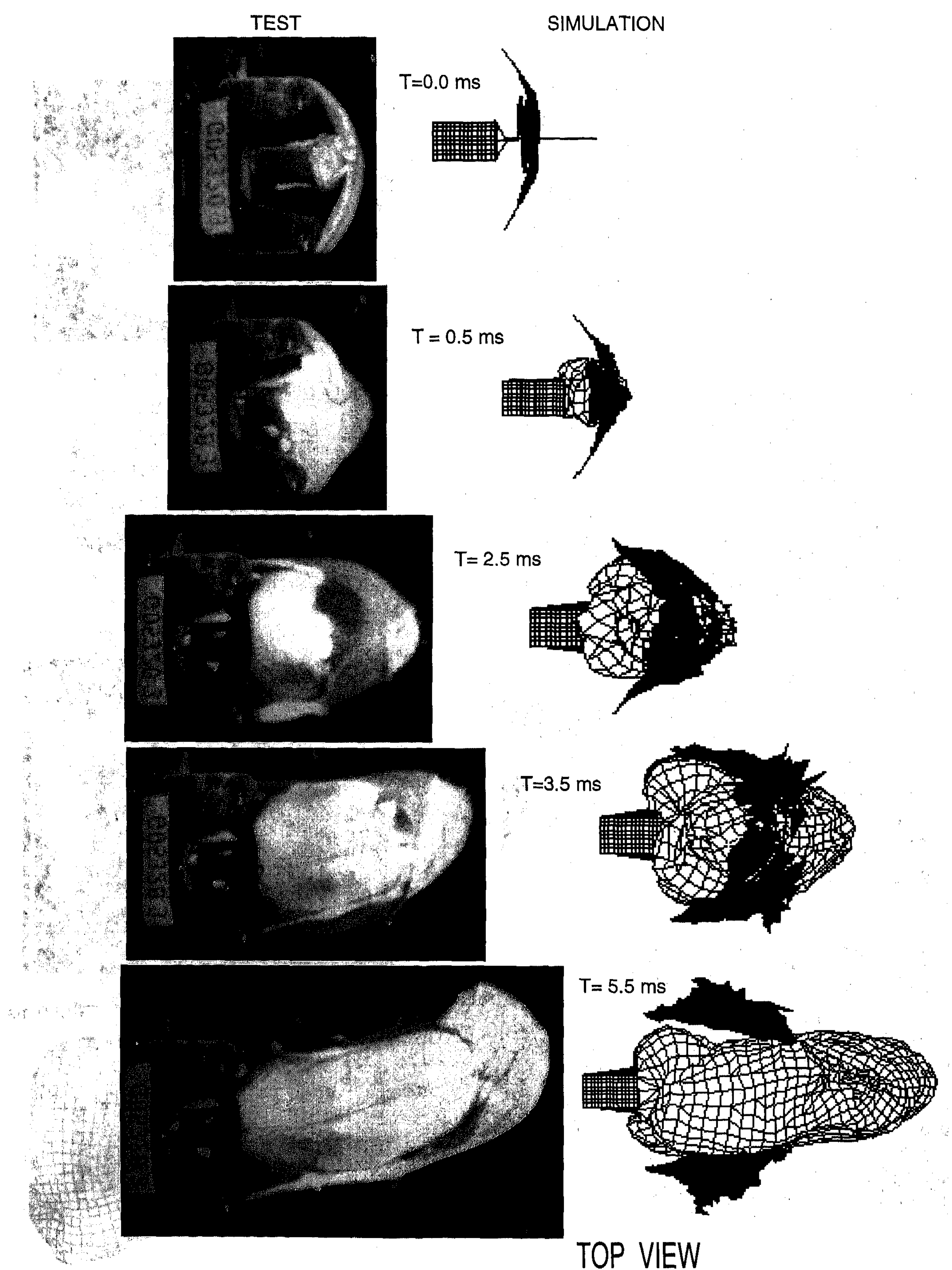


FIGURE 5. COMPARISON OF TEST AND SIMULATION FOR THE SIDE AIRBAG DEPLOYMENT

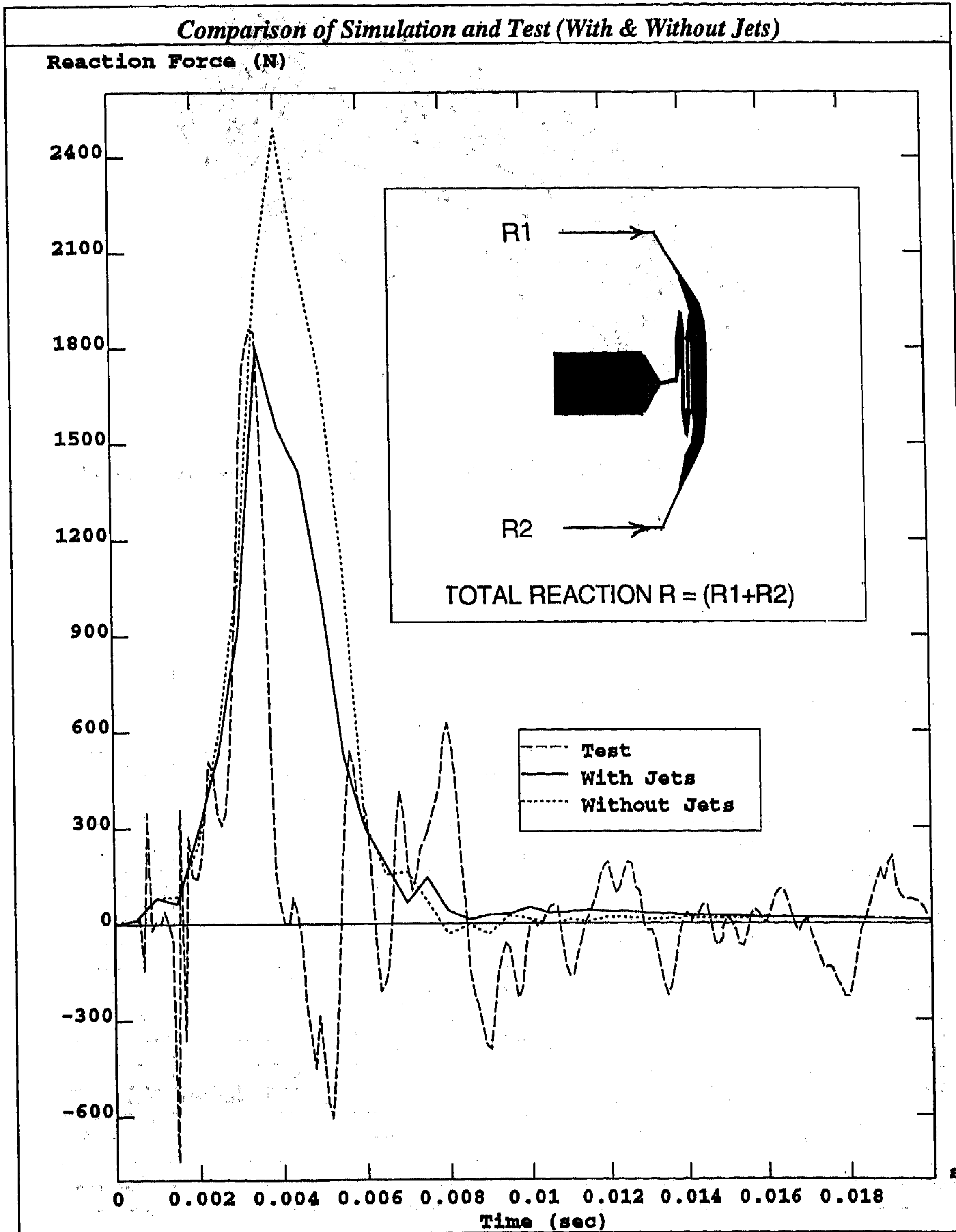


FIGURE 6. COMPARISON OF DYNAMIC TEST WITH A SIMULATION WITH AND WITHOUT THE JET EFFECTS

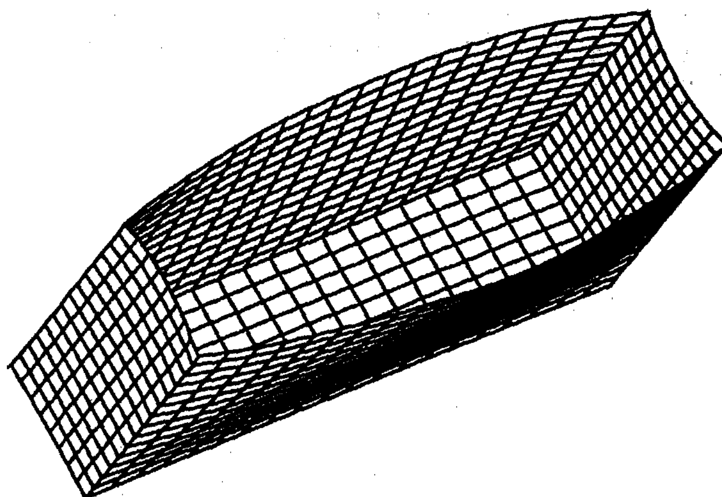


FIGURE 7. FISH MOUTHING DEFORMATION OF THE AIRBAG MODULE

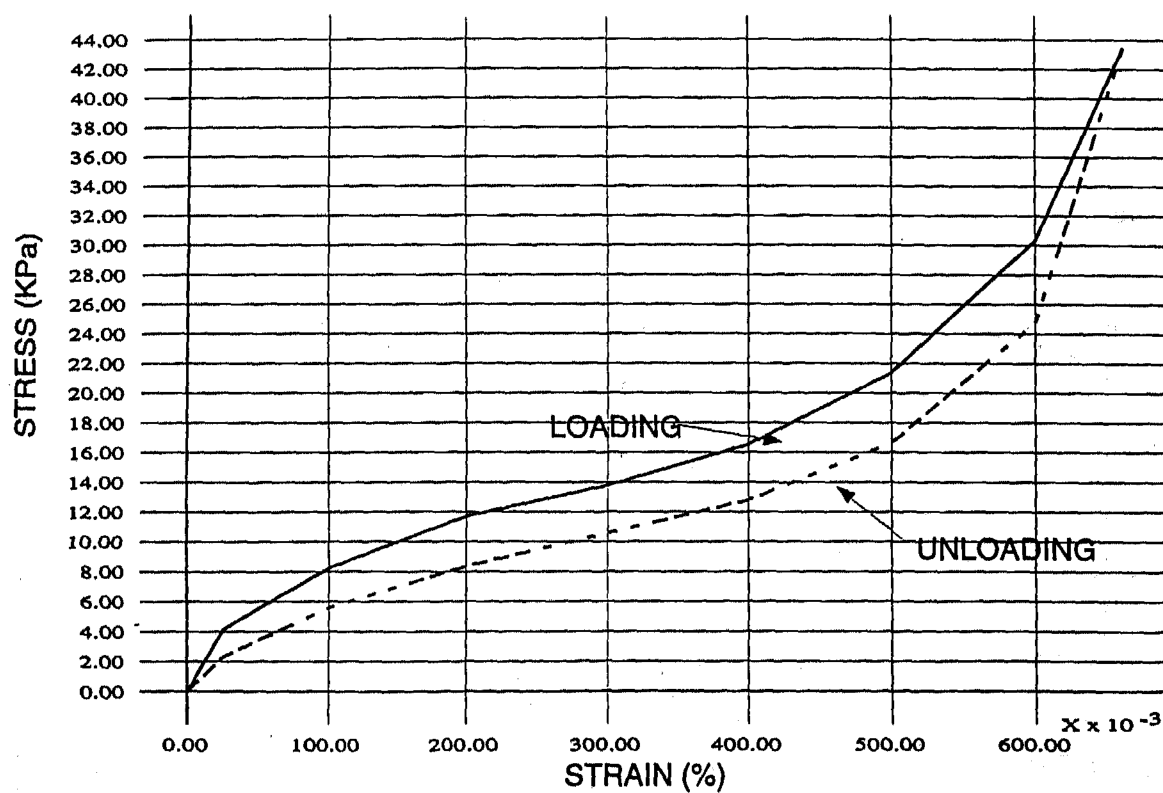


FIGURE 8. SEAT FOAM COMPRESSIVE STRESS-STRAIN CURVE

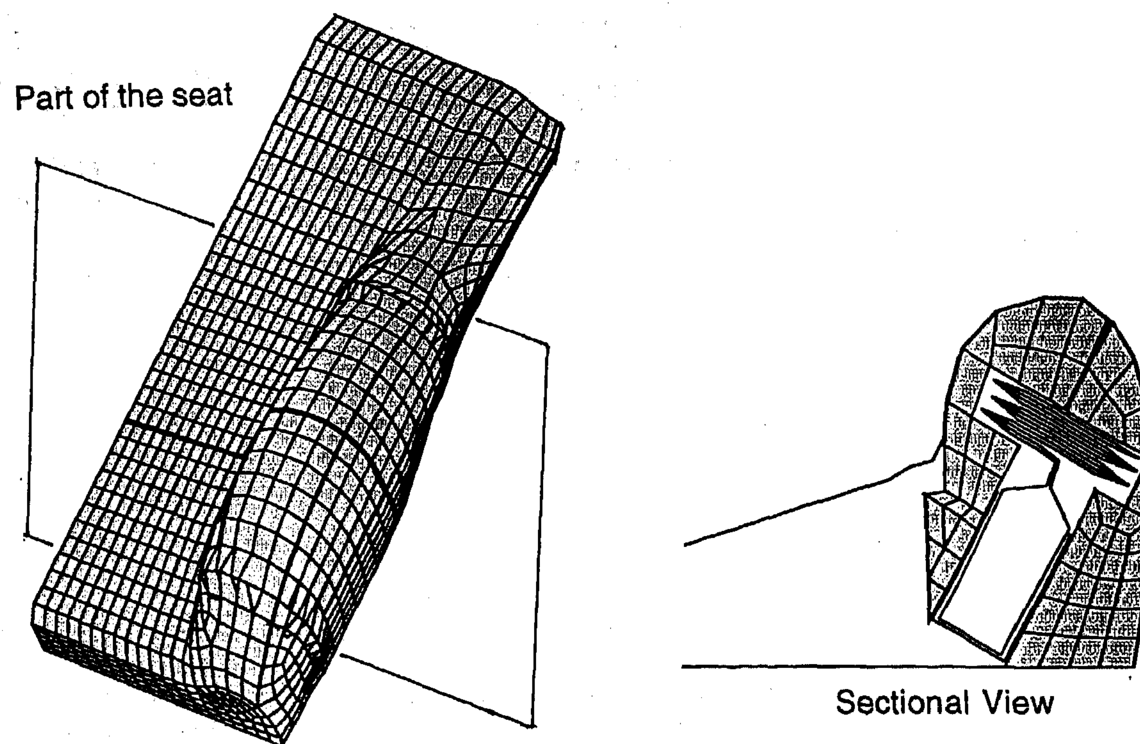


FIGURE 9. SEAT INTEGRATED SIDE AIRBAG (Only a part of the seat is shown)

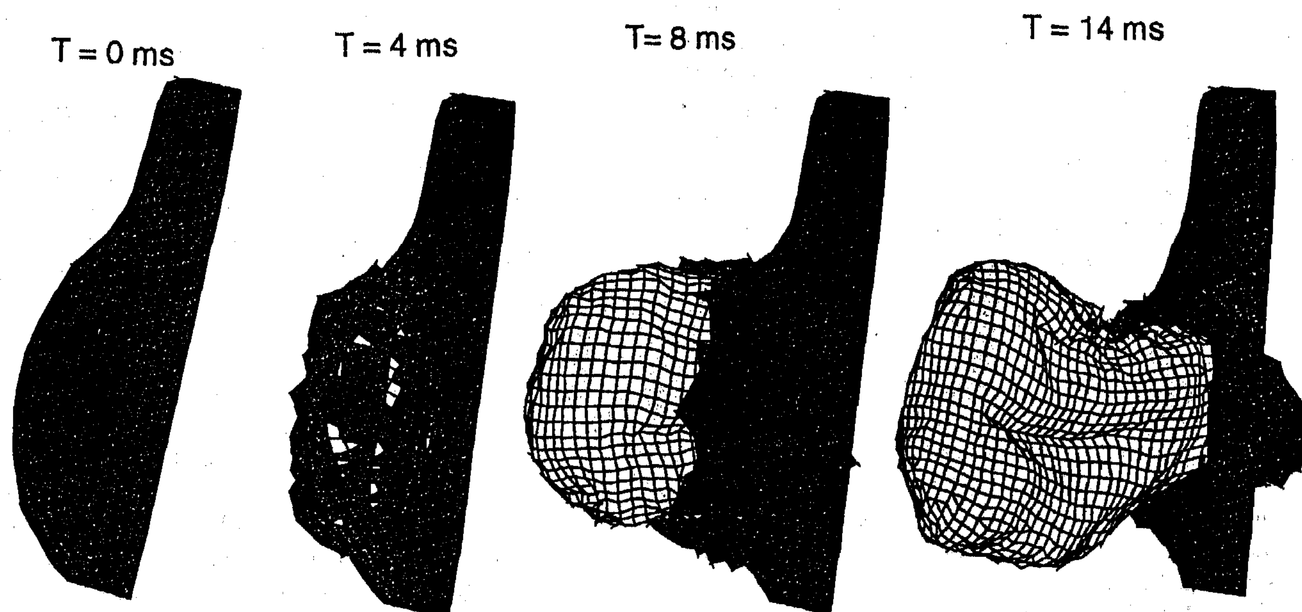


FIGURE: NOT TO SCALE

FIGURE 10. SIMULATION OF SEAT INTEGRATED SIDE AIRBAG DEPLOYMENT

970128

Critical Comparisons of US and European Dynamic Side Impacts

Nripen Saha, Stephen Calso, Djamal Midoun, and Priya Prasad
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Global engineering is increasingly becoming a practice within the automotive industry. Due to added engineering and manufacturing benefits, more and more new vehicles are being developed with common structure to meet the consumer needs in many local regions. While vehicle development and manufacturing process is becoming global, automotive safety regulations in various parts of the world have not been as uniform. A good example is the differing requirements for dynamic side impact protection of new vehicles. United States National Highway Traffic Safety Administration (NHTSA) and European Union (EU) have each produced their own distinct test procedures such as, different barrier faces, impact configurations, and anthropomorphic test devices (dummies). Although both test procedures have the same final objective - estimate occupant responses in side impacts, they differ greatly in execution and emphasis on occupant response requirements.

This paper discusses the differences in both side impact test procedures and vehicle & dummy performance in each case. Both dynamic side impact test data and analytical models are employed for this purpose. A full vehicle dummy-structure finite element is developed to investigate the two test conditions. Comparison of a typical vehicle performance such as, door and bodyside intrusion profiles, contact velocities with dummies, dummy kinematics in both test conditions are examined closely. It is determined that although each test induces different loading and responses, there may be some commonality in impacted vehicle performance. As the side impact regulations are still evolving both in USA and Europe, it appears that a unified approach for developing side impact countermeasures to meet the more demanding test condition may result in test mode simplification and robust side impact designs in new vehicles. One of the ideas examined in this paper is designing global new vehicles with a unified NHTSA/European procedure that simulates NHTSA side impact conditions with two EUROSID1 dummies, one in the front and one in the rear seat of the impacted vehicle.

INTRODUCTION

At present, two different side impact test procedures exist in North America and Europe. These are the US NHTSA FMVSS 214 and European Directive 96/27/EC dynamic side impact procedures. A NHTSA or European dynamic side impact test is composed of four basic elements. These are:

Test vehicle in stationary position with appropriate weight and instrumentation,

Deformable barrier to simulate striking vehicle stiffness at a specified velocity,

Anthropomorphic Test Devices (ATD) to help measure injury risk during impact, and

Test requirements based on various dummy responses.

Although both side impact test procedures have the similar objectives to provide understanding of vehicle and occupant responses, considerable differences exist in all four major entities described above. The procedures have unique dummies and barrier faces in each case.

Comparison of Test Procedures

The NHTSA dynamic side impact procedure was developed for passenger cars sold in the United States. It is being extended to truck vehicles ($\leq$ 6000 lbs GVWR) on September 1, 1998. The test uses U.S. DOT SID dummies located in the front and rear seats on the impacted side. The barrier is of honeycomb construction with different strength of honeycomb in barrier bumper (245 psi) and barrier body (45 psi). It represents an average U.S. vehicle size and stiffness. The barrier impacts with a crabbed angle of 27° to the target vehicle at 33.5 mph.

The European Directive dynamic side impact procedure has been developed for M1 and N1 vehicles sold in

Europe and it is still evolving, mostly in the area of barrier types, vehicle/barrier height and dummy seating positions. The European test procedure differs greatly from the NHTSA side impact procedure in that the barrier face, dummy, dummy measurements, impact speed and direction are all different. The test uses a single EUROSID1 dummy located in the front seat on the impacted side. The barrier specification in the European side impact procedure includes corridors for load-deformation behavior. It can be constructed of dense foam or honeycomb materials. Several European barrier types are available currently for side impact evaluation. These include KENMONT, UTAC, AFL, CELLBOND, FRITZMEIER, SHOWA, YOHAMA and PLASCORE constructions. The European barrier face is divided into six blocks with varying stiffness and represents an average European vehicle size and stiffness. The barrier impacts perpendicular to the target vehicle at 31 mph.

Comparison of Dummies

The NHTSA side impact procedure includes two US DOT SID dummies, and the European procedure requires one EUROSID1 dummy in the driver's seat on the impacted side. The dummies for the two types of tests are totally different in their construction. The Code of Federal Regulations (FMVSS 214) specifies the US DOT SID Thoracic Trauma Index (TTI) and the peak lateral Pelvis acceleration as the two criteria that have to be complied with for a vehicle to pass the dynamic performance requirements.

On the other hand, European Directive (96/27/EC) requires that the EUROSID1 dummy Head Performance (HPC), Thorax Rib Deflection, Thorax Soft Tissue (V*C), Abdominal Peak Force and Pubic Symphosis Peak Force criteria be used for approvals.

There are several requirements, as described below, regarding the seating positions for each of these two types of dummies.

SID Dummy Positioning

- Seats should be placed in mechanical mid position and full down (for the first two years the dummy's H-point must be within the door aperture). Seat backs are in nominal or 25 degrees, head rests full up, lumbar supports in lowest position and arm rests are lowered. Dummy is placed in the front seat and rear seats on the impacted side.

EUROSID1 Positioning

- Seats can be placed anywhere along the full seat track travel and full down. Seat backs are in nominal or 25 degrees, head rests full up, lumbar supports in lowest position and arm rests are lowered. Dummy is placed in the front seat on the impacted side.
- Arm is set at 40 degrees with respect to the dummy.

Comparison of Barriers

Listed below, in Table 1, is a brief summary of NHTSA and European barriers and test conditions.

Table 1: Comparison of Side Impact Barriers

Test	Barrier Wt	Barrier Speed	Impact Angle	Barrier Height
NHTSA	1360 kg	15.0 m/s	27°	280 mm
European	950 kg	13.8 m/s	0°	300 mm

Comparing the two deformable barrier data in Table 1, it is seen that the NHTSA barrier is heavier than the European barrier by 410 kg. Combined with a higher initial velocity the NHTSA barrier has approximately 40% more kinetic energy at the pre-impact state than the European barrier. Based on barrier dimensions and stiffness it is also seen that NHTSA barrier is bigger and stiffer than its European counterpart as the front end of an average US vehicle is bigger and stiffer than that of an average European passenger car or light truck. Figure 1 illustrates the difference in barrier stiffness by comparing dynamic force deflection characteristics of the two deformable faces. Based only on these statistics it may look as though the NHTSA test is more formidable than the European side impact test.

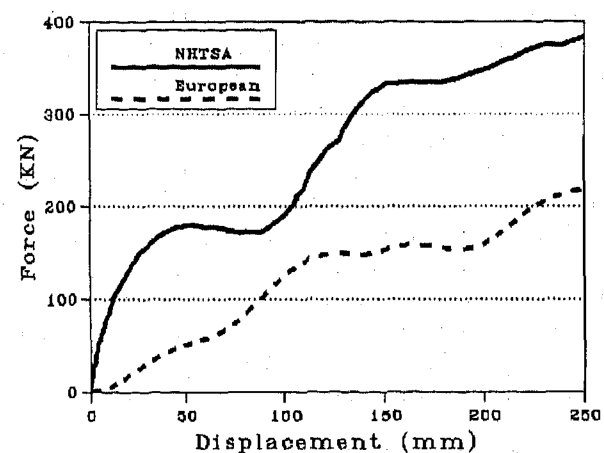


Figure1: Comparison of Barrier Face Force Deflection Characteristics

However, the barrier footprint on the body side structure must also be considered. Figure 2 illustrates how the two barrier faces are positioned against a mid size vehicle in NHTSA and European test conditions. The European barrier face being much smaller may contact only the B-pillar whereas the NHTSA barrier face may make solid contact with both hinge pillar and B-pillar. The effect of missing the hinge pillar in European test condition will be discussed further based on analysis results.

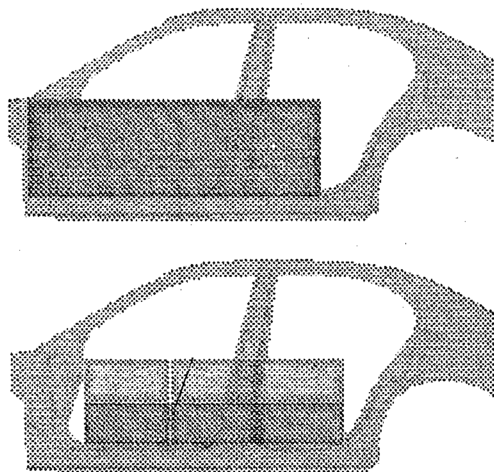


Figure2: Comparison of Barrier Overlap with Bodyside (NHTSA top / European bottom)

ANALYTICAL APPROACH

In this study, comparisons of the US and European Directive side impact performances are made using both side impact test data and full vehicle finite element models. Of the two vehicles modeled, one is a four door mid size vehicle and the other is a four door sub compact. The finite element system models of two vehicles contained the same vehicle structure and interior plus corresponding barrier and side impact dummies in each test set-up. The nonlinear simulations are performed using the commercially available code RADIOSS. Comparisons are made for barrier and vehicle energy absorption, impacted door and body side intrusion and velocities. Attempts are made to identify the vehicle components that play key a role in each test condition.

Since the European procedure has only a front seat dummy, in this study the front dummy responses are compared. As different types of responses are sought in US DOT SID and EUROSID1 dummies, for comparative purpose the individual dummy responses are examined in terms of margin relative to corresponding regulatory requirements. In addition to model responses, several sets of US and European test data from three similar vehicle tests are compared for this purpose.

SIDE IMPACT SYSTEM MODEL

Finite element models simulating dynamic side impact test are in use in the industry for the past several years. The FEA dummy-structure system model contains the vehicle structure, trim/bolster, side impact dummy and a barrier. Details of a typical side impact dummy-structure system model have been published by several investigators [1 3,4,7,10]. The primary component of the system model is of course, the vehicle structure model, built in detail with finite elements to include all sheet metal components, door closures, interior trim/padding, as well as suspension and powertrain components. The barrier model and the SID and EUROSID1 dummies are also created using finite elements based on component crash characteristics derived from various sled test configurations [2,5,6]. The dummy models are capable of predicting the physical responses accurately and sensitive enough to localized impacts as observed in side impact vehicle tests. The side impact system models provide insight into both structural performance such as, intrusions, accelerations, velocity histories and the occupant responses at pelvic, torso and head level.

ENERGY ANALYSIS

A side impact test involves contact between a deformable barrier face and the vehicle structure. Initially, the impact energy is mostly contained in the deformable barrier. As the impact proceeds the vehicle structure behaves as the primary energy absorbing component. A breakdown of the strain energy distribution between the vehicle and barrier at 25 and 50 milliseconds is shown in Figure 3 for impact against a mid size vehicle. The European barrier face absorbed about 48 percent of total strain energy and NHTSA barrier absorbed about 19 percent at 25 ms. At 50 ms these amounts are 39 percent and 24 percents, respectively. Both at 25 and 50 milliseconds the share of internal energy absorption by European barrier was higher than the NHTSA deformable barrier.

Figure 4 indicates how the vehicle internal energy is distributed among the body structural components of the vehicle at 25 and 60 milliseconds. It is seen that the major structural subassemblies contributing to the vehicle energy absorption are i) body side, ii) front door and iii) rear doors. The body side, front door and rear door percentages decrease at 60 milliseconds. This is primarily due to progressive inward crush of backup structures, i.e. front floor pan, seat cross members and roof panel. The rear door in the NHTSA test condition has a larger percentage of energy absorption (18.2%) than in EEC condition (11.1%). This is a direct result of the crabbed angle of the NHTSA barrier. As indicated above it appears that in both NHTSA and European procedures the body side played a large role in absorbing impact energy, followed by front and rear door systems.

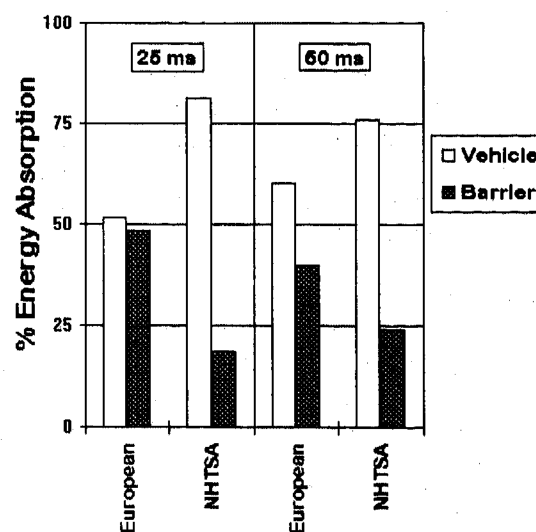
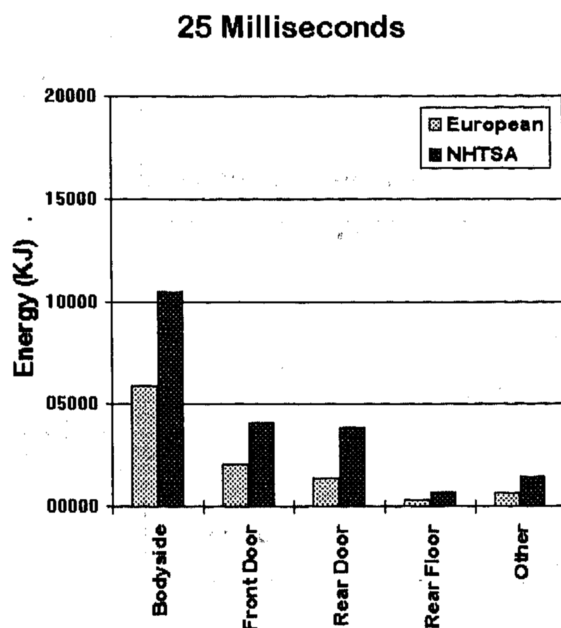


Figure3: Comparison of Vehicle and Barrier Energy Absorption in European and NHTSA Side Impact

25 Milliseconds



60 Milliseconds

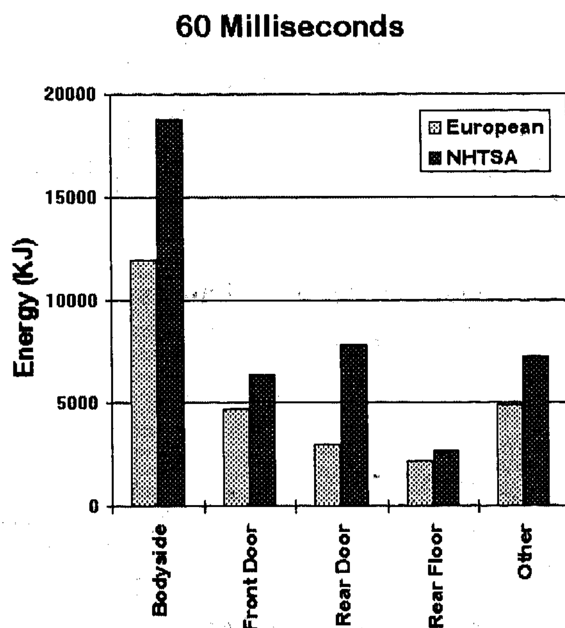
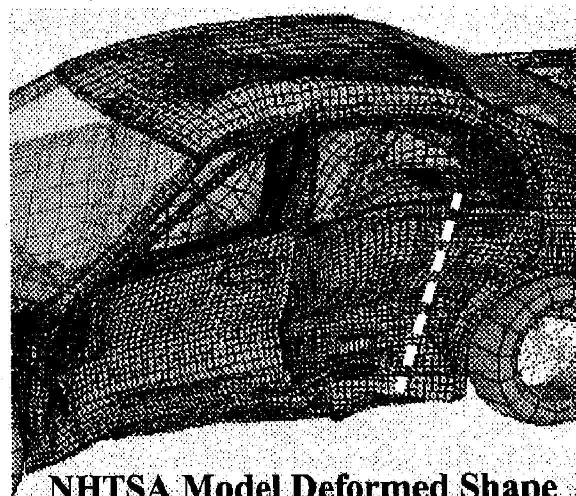


Figure4: Comparison of Internal Energy Absorbed by Vehicle Structure Sub-assemblies

INTRUSION COMPARISON

Shown in Figures 5 are the deformed bodyside geometries of the NHTSA and European side impact models. Figure 6 shows vehicles which have undergone similar testing. It is seen that the NHTSA model deformation pattern captures the barrier bumper imprint along the bottom of both doors and the deformation around the lower hinge pillar area. The model also captures the rear door vertical bending about mid belt line. The European model deformation pattern also matches well with what was observed in the test vehicle. Both test and model for the European Directive shows more deformation at the B-pillar and roof side rail joint when compared to that in the US procedure. The hinge pillar area is relatively untouched.



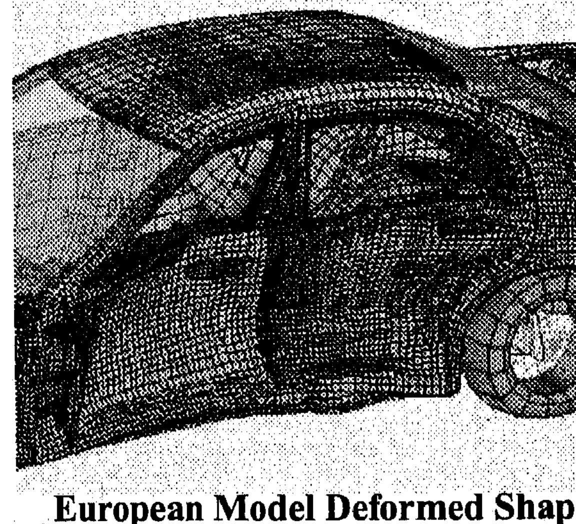


Figure5: Comparison of Deformed Geometry at 60 milliseconds

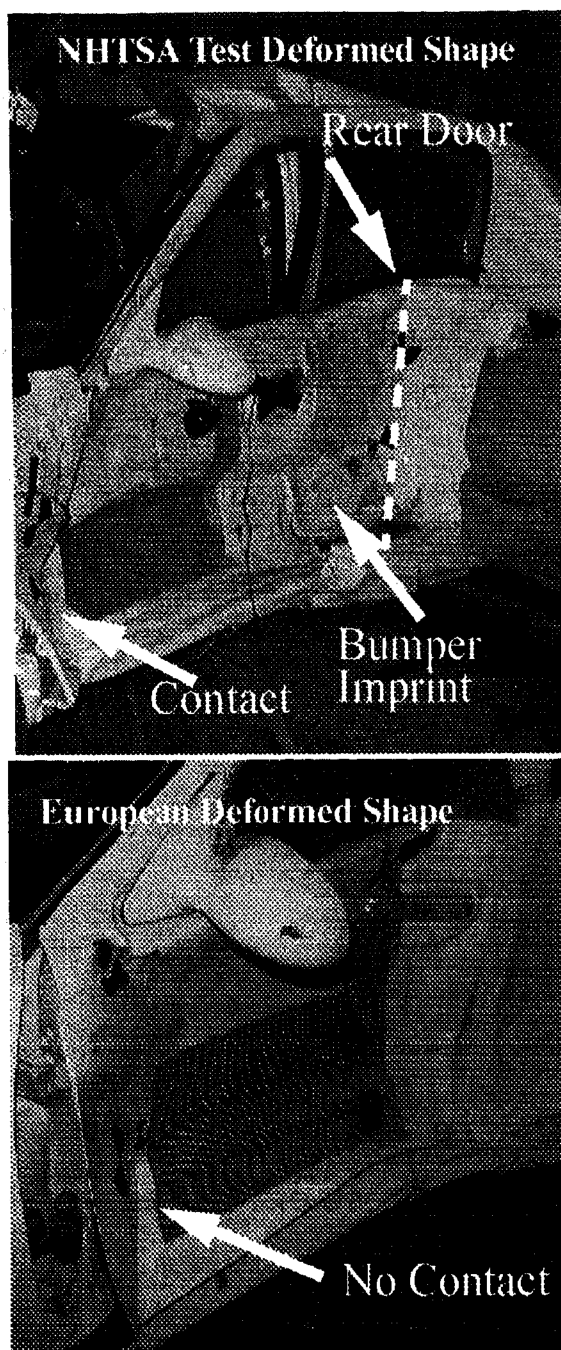


Figure6: Body Side and Door Deformation of a Vehicle in NHTSA and European Side Impacts

The deformation profiles of the front and rear door at the belt line at 60 ms are shown in Figure 7. The Figure includes door profile with a typical bodyside design to meet or exceed FMVSS 214 requirements only and the door profile with stiffened bodyside structure to meet or exceed both FMVSS 214 and European Directive requirements. For the design "as is", the door beltline intrusion in European condition is larger than NHTSA. However, with structural enhancements intrusions in European condition are close to or fall below NHTSA intrusions. The European belt line profile is somewhat circular primarily due to the normal impact and soft barrier edges whereas, NHTSA deformation pattern is flat and stretched rearward due to the crabbed velocity vector.

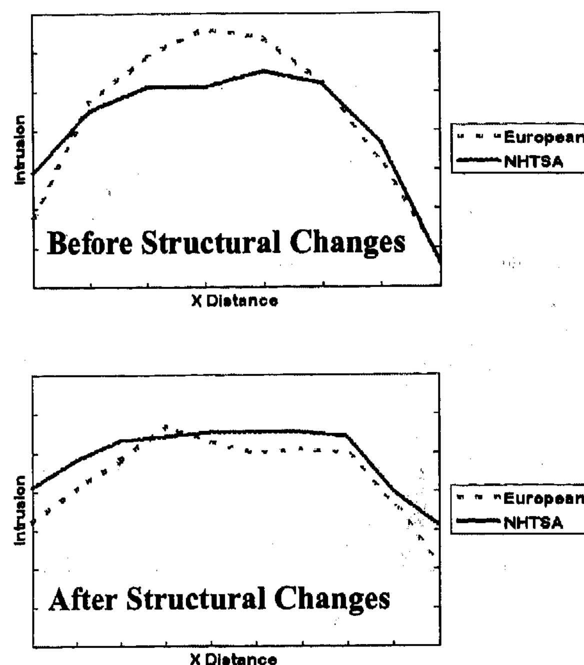


Figure7: Deformation Profile at Front and Rear Door at Beltline at 60 milliseconds

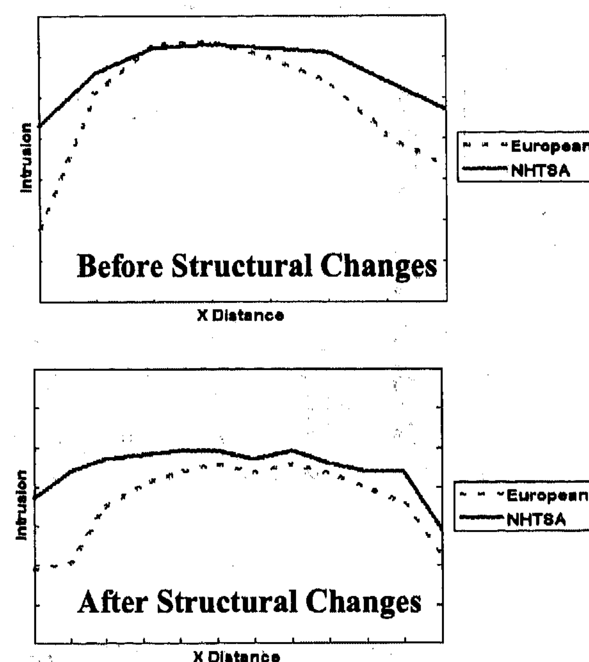


Figure8: Deformation Profile at Front and Rear Door H-point at 60 milliseconds

Figure 8 shows similar comparisons of the door profile at the H-point. The more protruding NHTSA bumper profile and stiffer barrier face accounts for its higher H-point intrusion. This is further illustrated in Figure 9 which shows vertical profile through the dummy's H-point at 15 ms. The

flatter European barrie face completely crushes the door outer panel and the NHTSA barrier face deforms only at the bottom portion of the door leaving more door volume at the top of the door at rib level.

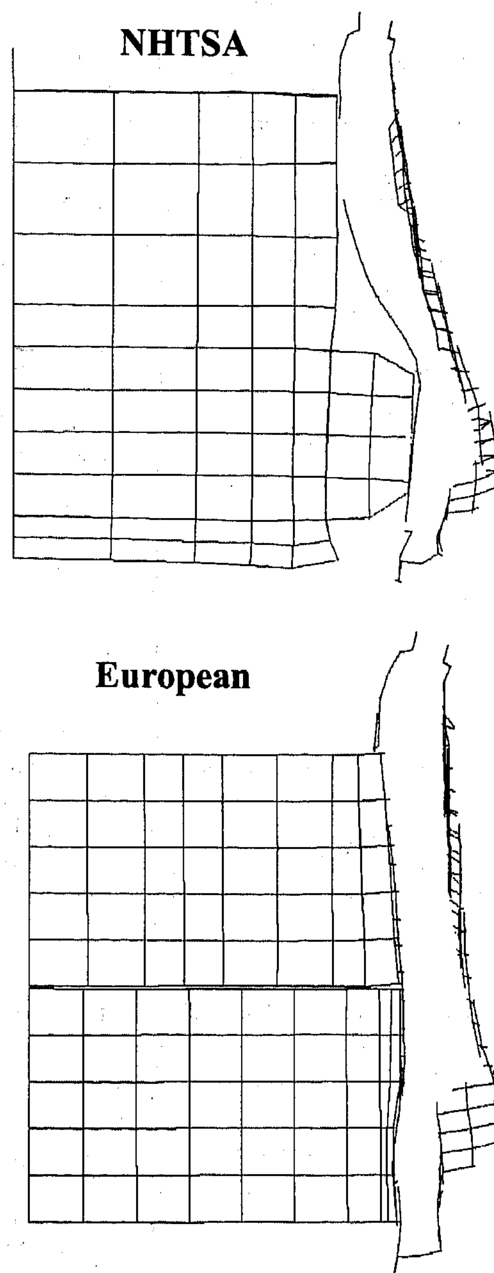


Figure9: Deformed Door Sections of NHTSA and European Barriers at 15 milliseconds

FRONT DUMMY RESPONSES

As mentioned earlier, the requirement for dummy response measurements in US DOT SID and EUROSID1 dummies are considerably different. Generally, the SID dummy responses are measured in terms of acceleration at ribs, lumbers spine and pelvic locations. In addition to pelvic G, the other FMVSS 214 requirement, Thoracic Trauma Index (TTI), is a combination of rib and lower spine acceleration.

On the other hand, the EUROSID1 measurements include many more variables, such as HPC, rib displacements, rib compression in terms of V*C, and lateral load at abdomen and pubic symphosis locations. The comparison of severity becomes harder in this case since measurements in two dummies are so different.

However, it is generally seen that meeting the US DOT SID dummy TTI requirement is more demanding than meeting the pelvic G requirement. Similarly meeting the EUROSID1 dummy rib deflection requirement is considerably harder than others.

In Tables 2 and 3, the model predicted front dummy responses of the mid size vehicle are shown against the corresponding requirements in NHTSA and European Directive side impact conditions.

Table 2: Mid Size Vehicle NHTSA S.I. Model Results

SID Responses	NHTSA Req	Model	Margins
TTI	85	64	25%
Pelvic Acc (G)	130	83	36%

Table 3: Mid Size Vehicle European Directive S.I. Model Results

EUROSID1 Responses	European Req.	Model	Margins
HPC	1000	254	75%
Upper Rib Deflection	42 mm	38	9.5%
Mid Rib Deflection	42 mm	37	12%
Lower Rib Deflection	42 mm	36	14%
V*C	1.20	0.82	31%
Abdomen Load (KN)	2.50	1.0	60%
Pubic Symphosis Load (KN)	6	4.0	33%

As stated earlier the TTI in NHTSA side impact and rib displacement in European side impact are the critical variables that in most cases controls compliance. Based on model predictions, it is seen that for European condition there is only 9 to 14 percents margin for the rib displacement requirement, where as in NHTSA condition there is a 25% margin in TTI for this vehicle.

Results from prototype tests of similar vehicles are now shown in Tables 4 and 5, in this case focussing on TTI and rib displacements, respectively.

Table 4: NHTSA S.I. Test Data

Tests	SID Resp	NHTSA Req.	Test Data	Margin
Test 1:	TTI	85	55	35%
Test 2:	TTI	85	61	28%
Test 3:	TTI	85	53	38%

Table 5: European Directive S.I. Test Data

EURO-SID1 Resp.	European Req.	Tests PT1 PT2 PT3	Margins (%) PT1 PT2 PT3
Upper Rib Deflection (mm)	42	38 47 28	7 -12 33
Mid Rib Deflec-tion (mm)	42	33 40 46	22 5 -10
Lower Rib Deflection (mm)	42	30 32 37	28 23 12
Average Rib Deflection (mm)	-	34 39 37	19 7 12

Again, it is confirmed based on test data of similar vehicles that rib deflections in European test conditions generally have much smaller margins against the corresponding requirement. The average rib deflection has a 7-19% margin and in 2 cases there is even a negative margin. On the other hand, TTI in three NHTSA tests has a larger margin of 28 to 38%.

DUMMY RESPONSES IN SUB COMPACT VEHICLE WITH STRUCTURAL ENHANCEMENTS

Next, the performance of a sub-compact vehicle designed to meet or exceed European test requirements are discussed. Considerable amount of structural enhancements in terms of bodyside reinforcement and other counter measures are added to the base design to be able to meet or exceed the European side impact requirements. The model responses in Table 6 shows that all rib deflections are meeting or exceeding the European requirements in addition to other parameter such as, HPC, V*C etc..

Table 6: Sub-Compact Vehicle European S.I. Model Response

EUROSID1 Response	European Req.	Model Resp.	Margin
HPC	1000	191	80%
Upper Rib Deflection (mm)	42	32.5	22%
Mid Rib Deflection (mm)	42	27.5	34%
Lower Rib Deflec-tion (mm)	42	28.3	32%
V*C	1.20	0.46	61%
Abdomen Load (KN)	2.50	1.61	35%
Pubic Symphosis Load (KN)	6.0	3.2	46%

The same vehicle was then analyzed to simulate the NHTSA side impact condition. Both pelvic acceleration and Thoracic Trauma Index (TTI) predictions are shown in Table 7 against corresponding NHTSA requirements. Based on model responses, it is seen that the sub-compact vehicle designed for European Directive side impact is already meeting the FMVSS 214 requirements with very large margins.

Table 7: Sub-Compact Vehicle NHTSA S.I. Model Responses

SID Responses	NHTSA Req.	Model Resp.	Margin
Pelvic Accelera-tion (G)	130	51	61%
Thoracic trauma Index (TTI)	85	64	24%

It is thus seen that both the mid size and sub compact vehicles provide similar conclusions in terms of front seat occupant responses in NHTSA and European test conditions. Both the model and test data indicate that for a specific vehicle designed with common structure and other common contents, it may be easier to meet or exceed the requirements in US test condition than the European Directive test condition. In other words, a new vehicle with common structure and restraint system may meet or exceed the US requirements once it is designed to meet or exceed the more demanding European Directive requirements.

UNIFIED US/European APPROACH:
NHTSA SIDE IMPACT WITH EUROSID1 DUMMIES

There is one other major difference between US and European side impact conditions. In the NHTSA test, there is an additional SID dummy located at the rear seat on impacted side of the vehicle. The European test condition thus, can not provide evaluation of side impact protection for a rear seated occupant. To meet this challenge and to explore commonizing other aspects of NHTSA and European test conditions, a third test set up is being looked at in this study. This test condition uses a NHTSA test up with NHTSA barrier face impacting at 27 degree crabbed angle at 33.5 mph. However, instead of US DOT SID dummies, it uses two EUROSID1 dummies in front and rear seats on the impacted side, in same seating position exactly as in NHTSA condition.

A simulation of this hybrid test up is also performed in this study. The structural responses such as door and B-pillar velocities are then compared with US and European test conditions as in Figure 10.

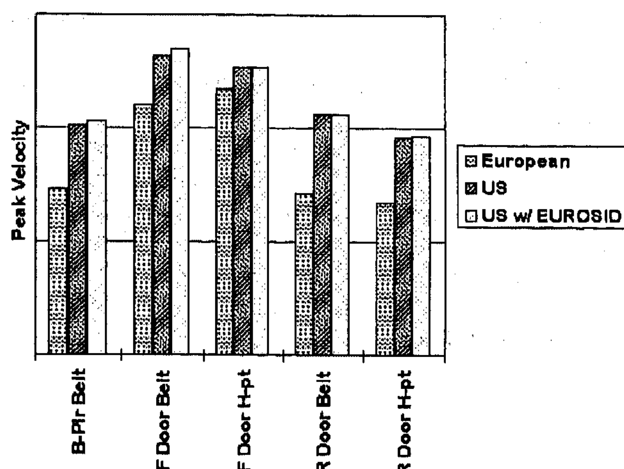


Figure10: Maximum Door and B-Pillar Velocities

As expected the velocity histories are very close to NHTSA condition.

Next, the front occupant responses are compared with the European test condition in Table 8;

Table 8: Comparison Of Front Dummy Responses

EUROSID1 Responses	European Req.	European SI Resp	US w/ EUROSID
HPC	1000	254	61
Upper Rib Deflection (mm)	42	38	32
Mid Rib Deflection (mm)	42	37	41
Lower Rib Deflection (mm)	42	36	42
V*C	1.20	0.82	0.82
Abdomen Load (KN)	2.5	1.0	2.1
Pubic Symphosis Load (KN)	6	4.0	4.7

It is seen due to higher intrusion of NHTSA barrier at lower door level, lower torso responses such as, abdomen and pubic symphosis loads are higher in unified US/European set up than the European Directive condition. For the same reason, the mid rib and lower rib deflections are much higher than the upper rib deflection.

Further investigation of this unified approach is needed based on model and prototype tests of different size and weight class vehicles before any conclusion can be drawn if this is a viable alternative.

CONCLUSIONS

The following conclusions are derived based on comparisons of structural and dummy responses of a mid size and a sub-compact vehicle simulating both NHTSA and European Directive dynamic side impacts:

Both the model and test data indicate that for a specific vehicle designed with common structure, it may be easier to meet or exceed the requirements in NHTSA test condition than in European Directive test condition. In other words, a new vehicle with common structure may meet or exceed the NHTSA requirements once it is designed to meet or exceed the more demanding European Directive requirements.

The European barrier face absorb more internal energy than the NHTSA deformable barrier.

The flatter European barrier face collapses the outer door panel completely, whereas the NHTSA barrier

primarily crushes the lower portion of the door.

The three major energy absorbing structural components in both NHTSA and European side impact tests are the impacted body side, front door and rear door, in order of magnitude.

The unified US/European test condition produces very similar structural responses, such as, intrusion and bodyside velocities of the impacted vehicle as seen in NHTSA side impact conditions. However, front dummy responses are considerably different.

The unified US/European test procedure might be a potential solution to simplify side impact test modes for global vehicles. However, this finding based on analytical data of one mid size vehicle needs to be further verified using model and test data for car and truck vehicles of different size and weight class.

ACKNOWLEDGMENT

The CAE contributions of Rabin Bhojan and Matthew Makowski of VSCAE, Ford Motor Company are greatly acknowledged.

REFERENCES

1. N.K. Saha, S.K. Mahadevan, D.E. Midoun, and J.S. Yang, "Finite Element Structure-Dummy System Model for Side Impact Simulation," ASME Winter Annual Meeting, 1991.
2. Prasad, P., Low, T.C., Chou, C.C., Lim, G.G and Sundarrajan, "Side Impact Modeling using Quasi-Static Crash Data," 1991 SAE Congress and Exposition.
3. S.M. Calso and N.K. Saha, "Analytical Comparison of NHTSA and European Dynamic Side Impact Performance," International Radioss Users's Conference, France, June 1995.
4. Midoun, D.E., Rao, M.K. and Kalindindi, B., "Dummy Models Crash Simulation in Finite Element Programs," Proceedings of the 1991 ASME Annual Conference.
5. Low, T.C., and Prasad, P, "Dynamic Response and Mathematical Model of the Side Impact Dummy," Prod. of 34th Step Car Crash Conference, 1990.
6. Midoun, D.E., Abramson, E., Rao, M.K. and Kakinada, R., "Development of a Finite Element Based Model of the Side Impact Dummy", SAE paper 930444.
7. A. Tilakasiri and P.D. DuBois, 1990, "A Simulation of an Automobile Side Impact with Finite Element Dummies Included in the Analysis," Proceedings of the Fifth International Cray Supercomputer Science and Engineering Symposium, London.
8. Daniel, R.P., P. Prasad and Yost, C.D., "A Biomechanical Evaluation of the Ford Side Impact Body Block and the SID/APR Side Impact Dummies," SAE 840882, SAE Intl. Congress, 1984.
9. A. Tilakasiri and D. Johnson, "Super Computer Simulation of Vehicles and Occupants for Front and Side Impact Protection at Ford," EAEC 5th International Congress, Strasbourg.

970132

A Method of Mapping Pre & Post NASS-93 Injury Descriptions to Enable Multi-Year Data Comparisons

E. A. Garthe and N. K. Mango
Garthe Associates

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

The National Highway Traffic Safety Administration's (NHTSA) National Accident Sampling System Crashworthiness Data System (NASS/CDS) collects detailed information on a sample of all police-reported motor vehicle crashes in the United States that involved passenger vehicles towed from the scene due to damage.[1]

This system's occupant injury definitions and severity levels, which are based on the Association for the Advancement of Automotive Medicine's (AAAM) Abbreviated Injury Scale (AIS), are periodically updated. The last major revision, in 1993, changed the injury identifier of all injuries, added injuries, deleted injuries and changed the body region and severity of a substantial number of injuries.[1] These changes confound the direct combining or trending of pre-1993 and 1993 and later NASS data.

This paper reports on the results of using a conversion method to equate pre-NASS-93 and NASS-93 and later injury descriptors and severities. The conversion maps each listed injury description and aspect in the NASS-93 coding manual to the injury descriptions and aspects listed in the NASS-88 coding manual.

This "full" conversion allows analyses of severity of injuries, frequency of injuries, and comparisons of injuries on an injury by injury basis as well as by category, body region, aspect or lesion by adjusting the NASS-93 and later data to be consistent with pre-NASS-93 data.

The paper discusses the conversion results and illustrates how the conversion can assist organizations that are analyzing data or creating standards based on multi-year NASS data.

Introduction

The (NASS/CDS) collects detailed information, such as exterior and interior vehicle damage, restraint use, occupant injuries and the possible sources of these injuries for about 5,000 crashes annually. Information about injuries is collected using methods based on the AAAM's AIS. Weighting factors permit the prediction of national estimates from this sample.[1].

The most recent 8 years of NASS occupant injury data were collected using two different injury system revisions. The "85 Revision" (based on the 1985 revision of the AIS or AIS-85) was used for data collected during the years 1988 through 1992.[2] For purposes of this paper, this system's coding manual will be referred to as the NASS-88 manual and the injuries collected during **1988-1992** will be referred to as NASS-88 data. The "90 Revision" of occupant injury data (based on the 1990 revision of the AIS or AIS-90) was used for data collected during the years 1993 to date.[1] For purposes of this paper, this system's manual will be referred to as the NASS-93 manual and the injuries collected during **1993-to date** will be referred to as NASS-93 data.

Due to the sample size, NASS CDS data in the years 1993 and later may not contain sufficient numbers of serious injuries for certain analyses. The use of the conversion adjusts the data between the systems to permit comparisons or trending over all years from 1988-to date. The additional data included using the conversion increases the data sample by about 4 times from about 55,000 (1993-94) to about 210,000 (1988-1994) injuries.

Tables are presented to show the conversion's effect on the NASS-93 manual's codes and on injuries by severity level for the years 1993-1994 when converted to NASS-88 severities. Examples are provided to illustrate how the conversion can help prevent misleading results during data analysis.

DESCRIPTION OF THE NASS-88 SYSTEM

The NASS-88 system is based upon, **but is quite different from** the AIS-85 and was used for NASS data collected 1988-1992[2,3]. There are two parts to the NASS-88 Occupant Injury Classification (OIC). The leading part of the code, the "pre-dot" is a four character alphabetic injury descriptor, which includes injury body region, aspect, lesion and system/organ identification characters. The body region and aspect characters can include "wild cards" (such as + . /) and other characters[1] that indicate the code can be expanded to a selection of body regions or aspects.[2] The fifth character in the code is a single numeric digit that specifies the severity level based on the AIS-85.[2] NASS-88 uses severities 1-6 like AIS-85, but uses 7 instead of 9 for an injury of unknown severity. The severity numbers correspond to the levels of injury shown in Table 1, with AIS-1 representing a minor injury and AIS-6 representing a maximum injury.

Table 1. Severity Numbers and Descriptions

AIS Severity Number	Description
1	Minor
2	Moderate
3	Serious
4	Severe
5	Critical
6	Maximum[2]
7	Injury of Unknown Severity
9	Unknown Injury
0	Not listed as an injury[3]

Differences between NASS-88 and AIS-85

While NASS-88 is based on AIS-85, the organization of the coding manual, the injury descriptions, code structure and severities are substantially different. The NASS-88 manual contains substantially more chapters (15 vs. 7) and injuries than AIS-85. NASS-88 also specifies 20 body regions, over three times the number used by AIS-85. The AIS system, regardless of the revision, collapses its injury list into 6 body regions as defined by the Injury Severity Score (ISS).[4,5]. The ISS is defined as the sum of the squares of the highest AIS score in each of the three most severely injured ISS body regions (Head or Neck, Face, Chest, Abdomen or Pelvic Contents, Extremities or Pelvic Girdle and External).[5,10] The NASS-88 manual is not organized to easily use ISS. The authors believe that the conversion discussed in this paper is the first time that ISS Body Regions have been assigned to NASS-88 and NASS-93 codes on a system wide basis to permit the computation of ISS for any combination of injuries.

NASS-88 also includes an "aspect" characteristic that is recorded as part of every NASS-88 data code, and has no equivalent in the AIS-85 coding structure. Aspect is used in NASS-88 to describe the location of an injury (left, right, anterior, posterior, whole area, etc).[1,2] NASS-88 defines which aspects are permitted for each code on a code by code basis, on occasion using the same aspect "wildcard" placeholder in different locations in the coding manual with different meanings. This makes the expansion of the manual for aspect difficult. The use of aspect changes the meaning of some codes with similar injury descriptions in AIS-90 and AIS-85.

NASS-88 and AIS-85 do not always recognize the same injury or assign the same severity to injuries described in the same way. For example, NASS-88 places the code for pituitary injury with brain stem injuries (hypothalamus, medulla, midbrain, pituitary, pons and thalamus) with a severity level of AIS-5 or AIS-6, depending on the precise injury. In contrast, AIS-85 did not list pituitary injury in its definition of brain stem injuries (hypothalamus, medulla, midbrain, pons) nor did it list a separate code for pituitary injury elsewhere in manual.[2,4] This particular difference in the systems was changed later in both NASS-93 and AIS-90, a pituitary injury code is listed separately from brain stem injuries and is assigned a severity of AIS-3.[1,5] Creating a conversion by code matching without referring primarily to the injury descriptions can misclassify this type of match. This is why the author's conversion is based on matching the full injury descriptions.

Unlike the AIS systems, the NASS-88 system does not use a unique injury code to identify a single unique injury description. In NASS-88 two or more injury descriptions often share a common injury code. Further, an injury description may be repeated several times in different locations of the coding manual, a practice also not used in AIS-85 or AIS-90. Skin injuries are an example, many of which are replicated between the Skin Chapter and the individual body regions. As a result, while there are 5,698 fully expanded NASS-88 injury descriptions, there are only 3,724 unique codes to describe them. Also unlike AIS-85, a single injury "pre-dot" code can have more than one AIS severity associated with it. In these cases, as the severity changes, the definition of the injury changes.[2] Example 1 illustrates an example of this effect.

---

[1]Note: The use of the "." in the predot code required NASS-88 to use the dash as the separator between the pre and post code parts.

[2]Note: NASS-88 also listed this level as "virtually unsurvivable"

[3]Note: While not defined in NASS, the injury severity of zero is used in this paper to indicate a NASS-93 injury that has no equivalent listed injury in the NASS-88 system (hence the level '0' indicating no severity attached to the injury code).

Example 1.
CODE STRUCTURE FOR NASS-88 CODE: H_UB

Body region:	H = head (NASS region)
Aspect:	_ = right, left, anterior, posterior, superior, inferior, whole region or unknown (one only)
Lesion:	U = unknown
System/Organ:	B = brain
Severities:	**5 = critical**, 4 = **severe**

If the severity assigned is **AIS-5** (critical) H_UB is used to represent - diffuse brain axonal injury (white matter shearing injury) as well as cerebellum and cerebrum hematoma /hemorrhage -- epidural or subdural >100cc. However, if the severity is **AIS-4**, H_UB represents the injuries cerebellum and cerebrum hematoma/hemorrhage -- epidural or subdural less than or equal to 100cc or unspecified as well as hematoma/hemorrhage -- intracerebral or intracerebellar (including petechial and subcortical). The code "pre-dot" (H_UB) is the same in both cases. As a result of this type of code structure, the combination of pre-dot and severity is necessary to describe an injury in NASS-88 in the most accurate way. Users of AIS-85 and AIS-90 can be confused by NASS-88, since those systems uniquely identify injuries with the pre-dot part of the code alone and do not replicate codes or re-use injury identifiers.

DESCRIPTION OF THE NASS-93 SYSTEM

Like NASS-88, NASS-93 was based upon the AIS, in this case AIS-90. However, unlike the NASS-88 system, NASS-93 shares much of the organization and structure of the base AIS-90 manual. There are three parts to the NASS-93 injury code. The "pre-dot" part of the code is a six character injury descriptor, which includes injury body region, type of structure, specific structure, and level identification characters. The second part of the injury code is a single numeric digit that specifies the severity level based on the AIS-90.[1] NASS-93 uses severities 1-6 like AIS-90, but uses 7 instead of 9 for an injury of unknown severity. The severity levels are shown in Table 1. The third part of the code is the injury aspect (left, right, etc) and is represented by a single numeric digit. The coding manual uses "wildcards" for body region and aspect characters to indicate the injury description can be expanded to cover a range of body regions or aspects.[1]

Differences Between the NASS-93 and AIS-90

The first part of the injury code in both systems is a 6 digit injury descriptor. This descriptor is all numeric in AIS-90, but the NASS-93 coding manual contains some "wildcards" for body region. NASS-93 changes the injury code for some injury descriptions compared to AIS-90, skin injuries in particular. NASS-93 also changes some injury descriptions for the same injury code used in AIS-90. In the cases where NASS-93 uses body region "wildcards" in the injury code, these must be expanded before coding the injury. AIS-90 does not use any "wildcards". A review of the differences between NASS-93 and AIS-90 conducted by the authors has shown that about 250 (19%) of the original 1315 AIS-90 codes differ in some manner between the two systems (**before** the codes are expanded for injury aspects). The second part of the injury code, AIS severity, is similar between NASS-93 and AIS-90. Codes that share the same injury description almost always have the same severity. While AIS-90 has only two code parts, NASS-93 adds a third part to every code to represent injury aspect. AIS-90 has no direct equivalent, although a few AIS-90 codes specify aspect in their *descriptions* (e.g. bilateral lung injuries). The range of aspects defined in NASS-93 for use with a specific code change on a code by code basis. The aspect code selected for an injury is indicated by an additional digit at the end of the code.[1] The aspects are the same as those defined in NASS-88, but are represented in NASS-93 by numbers rather than letters as in NASS-88.

Once properly expanded for body region, the injury descriptor in NASS-93 uniquely identifies the injury, as it does in AIS-90. AIS-90 contains 1315 injury descriptions and codes and there is no expansion required. NASS-93 contains, after expansion for body region and aspect, 3176 codes. The number of injury descriptions and unique codes in NASS-93 are identical at 3176, so there are no re-used codes. Because the injury descriptor and severity are independent quantities, it is possible to revise and update severities in NASS-93 without altering the injury definitions, a process not possible in the NASS-88 system.

ILLUSTRATION OF DIFFERENCES BETWEEN NASS-93 AND NASS-88

Table 2 shows an example of the differences in NASS-88 and NASS-93 injury codes for the same injury.

Table 2

Injury Description	NASS-88	NASS-93
Right Femur Fracture, NFS <12 years old	TRFS-3	851802.2-1

In this example the TRFS-3 code in NASS-88 indicates a thigh body region, right aspect, fracture of the skeletal system, with the injury description "Right femur fracture-Condyle, Head, Neck, Trochanters, Shaft, Or Intertrochanteric--With Or Without Sciatic Nerve Involvement". The -3, included as part of the code identity, specifies an AIS-3 level (serious) severity. The NASS-93 injury code 851802 specifies a lower extremity, skeletal femur fracture NFS, age < 12 years old. The ".2" specifies the severity as AIS-2, (moderate) while the following -1 specifies right aspect.[1,2]

NASS-88 AND NASS-93 SYSTEM DIFFERENCES THAT CAN CONFOUND ANALYSES OF INJURY DATA

The following differences between the NASS-93 and NASS-88 systems can confound analyses unless adjustments are made:

- severity changes - numerous decreases and increases in injury severity occur due to revisions in the coding system
  - without conversion, changes in severity related to coding system revisions may confound severity changes related to safety interventions
- injury description changes - text descriptions change in some way for many injuries
- changes in code identifier format: NASS-88 uses an alpha code and NASS-93 uses a numerical identifier
  - the lack of a common identifier between the systems prevents direct comparison of the data
- addition of codes - for example, a list of 51 intracranial blood vessel injuries, ranging from AIS-3 to AIS-5 in severity, was added to NASS-93
- deletion of codes - for example, some injuries involving major respiratory difficulty (e.g., hyoid fracture and vocal cord injury, both AIS-5) were deleted from NASS-88
- aspect changes - valid aspects change for some injuries, for example bilateral may be listed as a valid aspect for a particular injury in NASS-93 but not for the same injury in NASS-88
- body region changes - NASS-93 uses 9 and NASS-88 uses 20 body regions
- age adjustments - NASS-93 added injury codes that are age adjusted for severity (e.g., certain types of brain, rib cage, femur and burn injuries)

CONVERSION PROCESS

The first step in the conversion process was to fully expand the NASS-93 and NASS-88 manuals. This expansion was needed to describe as precisely as possible every injury in each system, using the written text, valid body region and valid aspect expansion. This step was crucial for improving the precision of the matching process. The following example illustrates how this improves match precision.

Table 3

Fully Expanded Descriptions:

NASS-93		NASS-88	
Injury	right injury	Injury	right injury
	left injury		left injury
	bilateral injury		(right + left)

Using the fully expanded code for each injury, it is possible to make a precise match of the bilateral injury in NASS-93 to an equivalent (right + left) injury in the **same occupant** in NASS-88. This type of match would not be possible with unexpanded or partially expanded code descriptions. Without this expansion, there would be no match for bilateral injury and the (right + left) injuries in NASS-88 would be counted as two injuries instead of one bilateral injury, confounding the injury totals.

The full expansion of the NASS-93 manual to account for valid body regions (e.g., head, neck, upper extremity, etc) and multiple matches results in a total number of 1396 codes, unexpanded for aspect. Further expansion for aspects (e.g., right, left, bilateral, etc.) results in 3,176 unique injury codes, in contrast to the total number in the current manual (1,297). A total previously reported by the author represents an "unexpanded" total of 1,292 codes, [3] prior to the addition of 5 new codes in 1996 by NHTSA [6].

A complete listing of NASS-88 injuries also was created by expanding the injuries in the printed manual to account for valid body regions, aspects and severity upgrades (e.g., add severity of +1 if open fracture). The expansion of the NASS-88 manual required a complex multi step process. This was done to obtain all the proper unique injury descriptions. The first step was to expand the original 755 injury codes in the manual to 1271 injury codes by expanding the "wildcards" for body regions and adding the listed severity upgrades. Considering each unique description, without aspect expansion, there are 2193 entries. Further expansion of the "wild card" aspect specified for each code, results in 5,698 unique injury descriptions, but only 3,724 unique NASS-88 alpha codes plus severity to describe these descriptions.

The distribution of injury codes for NASS-88, AIS-85, NASS-93 and AIS-90 by severity level is shown in Table 4. This table shows that the percentage of injury codes listed in the manuals that were AIS-3 or higher (excluding AIS-7) decreased from 62.8% in AIS-85 to 50.6% in AIS-90.[3] However, the opposite is true between the unexpanded NASS-88 and NASS-93 systems, where the same AIS-3 and higher injuries increased from 37.7% in NASS-88 to 47.9% in NASS-93. These opposite trends illustrate the differences that exist between the AIS and NASS systems. Because lower severity injuries tend to have more aspects than the higher severity injuries, when the NASS systems are expanded for aspect, the AIS-3 and above percentages (excluding AIS-7) shift down for the NASS-88 and NASS-93 expanded systems to 32.7% and 38.9% respectively. While the systems' distribution of injury codes by severity does not directly correspond with the distribution of injuries coded using the systems, this table does illustrate the differences between the AIS and NASS systems, and also points out the effect on severity of system aspect expansion. Compared to the conversion results shown later, Table 4 also illustrates how it is not possible to deduce the effect of a conversion by looking at a system's severity distributions alone.

A systematic and comprehensive line by line review of all 3,176 NASS-93 expanded injury classifications was conducted and the appropriate NASS-88 injury

TABLE 4												
	AIS-85, AIS-90, NASS88 AND NASS93 CODE SEVERITY DISTRIBUTIONS With and without Aspect Expansion for NASS Systems											
	AIS85		**AIS90**		**NASS88**		**NASS93**		**NASS88**		**NASS93**	
Severity	**No Aspects**				**Aspects Not Expanded**				**Aspects Expanded**			
0	0	0.0%	0	0.0%	1	0.0%	1	0.1%	1	0.0%	1	0.0%
1	166	13.6%	246	18.7%	528	24.1%	293	21.0%	1538	27.0%	955	30.1%
2	289	23.6%	399	30.3%	798	36.4%	422	30.2%	2139	37.5%	971	30.6%
3	374	30.6%	343	26.1%	462	21.1%	357	25.6%	1213	21.3%	771	24.3%
4	217	17.7%	156	11.9%	199	9.1%	148	10.6%	415	7.3%	248	7.8%
5	151	12.3%	144	11.0%	136	6.2%	142	10.2%	200	3.5%	193	6.1%
6	27	2.2%	22	1.7%	30	1.4%	22	1.6%	38	0.7%	22	0.7%
7	0	0.0%	0	0.0%	39	1.8%	11	0.8%	154	2.7%	15	0.5%
***9**	0	0.0%	5	0.4%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Totals	**1224**	**100.0%**	**1315**	**100.0%**	**2193**	**100.0%**	**1396**	**100.0%**	**5698**	**100.0%**	**3176**	**100.0%**

*Note: AIS-90 and AIS-85 use 9 for "Injury, Unknown Severity'

descriptor(s) and severity code(s) for each NASS-93 injury were assigned by an experienced AIS user with input from expert NASS and AIS coders. Using the fully expanded (numeric or alphanumeric) codes and complete text description for each injury, matches were made to the most precise code(s) available in NASS-88. The matches were made to the most appropriate code(s) and severity level(s), factoring in the injury aspect, age of injured occupant, combinations of injuries and multiple injuries. The conversion of the expanded NASS-93 into the expanded NASS-88 manual results in a "3 dimensional" conversion that accounts for body region, aspect and severity. In contrast, a conversion that uses unexpanded codes or focusses on one of the three variables is "1 dimensional" in nature. The full conversion contains over 30,000 description to description matches.

The conversion was structured to convert injury severity and body region data in either direction; however, this paper reports on the results of converting NASS-93 injuries to NASS-88 descriptions and severity levels. This direction was chosen by the authors because preliminary analysis indicates that better match precision is possible to obtain by converting from the system with generally more detailed injury description information (NASS-93) to the one of generally less detailed information (NASS-88). However, the authors plan further analysis of the bidirectional capabilities of the conversion.

Since NASS is based on AIS, the conversion was designed to be consistent with assumptions used to develop other AIS-related conversions, in particular those between AIS-90 to AIS-85 (completed by the authors) for use with hospital injury severity data and NASS-93 to AIS-85 for use in an earlier unexpanded conversion to NASS injury severity data.[7] A limitation of this earlier conversion developed by the author was that it assigned AIS-9 (injury, but of unknown severity) to NASS-93 codes that did not exist in the earlier system.[7] This particularly affected the injuries in the head body region because NASS-93 added a section of intracranial blood vessel injury codes (n=51) that did not exist in NASS-88. These injuries range in severity from AIS-3 to AIS-5, but were coded in the data as AIS-7 or injury of unknown severity by coders using the NASS-88 manual.[7]

In this conversion, these codes can be treated as AIS-7 for frequency analyses but as either AIS-7 or special "reference" codes for severity analyses. Whenever possible, "reference" severity codes were assigned in cases like these, using the injury code and severity for a related (but not matchable) injury in NASS-88. An example would be an intracranial injury involving a hemorrhage used as a reference code for an unmatchable intracranial blood vessel laceration. Although these "reference" codes are "flagged" and should not be used in analyses in the same manner as a "matched" code, they preserve more information about the injury severity than assigning AIS-7. In cases where no reference code was identified the general AIS-7 code for the body region or an AIS-0 (no injury) code was assigned.

The conversion described in this paper provides a method for adjusting for these confounding factors, so that the NASS data from 1988-1992 and 1993 and later can be used together for a variety of injury severity, body region and frequency analyses. In addition, it links NASS-93 and NASS-88 in a unique way so that, for the first time, data queries involving injury descriptions, codes, aspects or severity can be made in a consistent manner from one system to the other, **in either direction.** However, this paper presents data on the "historical" conversion only, which assigns NASS-88 codes to the current NASS-93 injuries. The conversion also provides an enhanced ability to "filter" the NASS data sets for "invalid" codes used in either data base, since it

TABLE 5
ORIGINAL AND CONVERTED NASS SYSTEMS
Without Aspect Expansion

Severity	NASS88		NASS93		NASS93 CONVERTED	
0	1	0.0%	1	0.1%	5	0.4%
1	528	24.1%	293	21.0%	214	15.3%
2	798	36.4%	422	30.2%	382	27.4%
3	462	21.1%	357	25.6%	407	29.2%
4	199	9.1%	148	10.6%	193	13.8%
5	136	6.2%	142	10.2%	160	11.5%
6	30	1.4%	22	1.6%	27	1.9%
7	39	1.8%	11	0.8%	8	0.6%
Totals	2193	100.0%	1396	100.0%	1396	100.0%

maintains a listing of all codes, aspects and severities for each NASS system.

CONVERSION RESULTS

Table 5 summarizes the distribution of severity in the unexpanded NASS-88 and NASS-93 systems, and compares them with the NASS-93 system after conversion. The conversion process maps all possible matches to each NASS-93 code and therefore may contain several code matches of differing severity. In these cases the tables present a summary number representing the severity of all the codes matched. Since the conversion preserves all the code matches, it is possible to look at the precise codes, **and even the individual descriptions** matched to a given injury and obtain the severity of all injuries. The 30,000+ matches involved in the conversion, however, require the simplified representation of the conversion results shown.

Table 5 shows that the percentage of codes that are AIS-3 and higher (excluding AIS-7) is 37.7% for NASS-88, 47.9% for NASS-93 but increase to 56.4% for the converted NASS-93 codes. A higher percentage of converted codes are AIS-3 and above than either of the base systems. The conversion creates a **net** gain of 118 codes of AIS-3 or higher severity. This result is not perhaps surprising in perspective, as the revisions in the AIS have tended in general to lower severity for some injuries over time.[1,3] This is to be expected as a result of advances in medical technology.

Table 6 shows the percentage distributions of codes by severity for the expanded NASS-88, NASS-93 and converted NASS-93 systems. The percentage of codes that are AIS-3 or higher in severity is 32.7% in the expanded NASS-88 system and 38.9% in the expanded NASS-93 system. When the NASS-93 system is converted, the percentage of codes AIS-3 or higher increases to 48.5%. The conversion creates a **net** gain of 306 codes of AIS-3 or higher severity. The effect of the conversion is an increase in both the proportion and number of AIS-3 and higher codes.

Changes in Severity by NASS Body Region

The conversion also provides a method to reconcile different body regions that may be assigned an injury in NASS-88 and NASS-93. As noted previously, there are 9 and 20 body regions in NASS-93 and NASS-88, respectively. Some injuries are distributed into more specific body regions (e.g., upper extremity injuries are split into upper arm, elbow, forearm and wrist/ hand/ fingers) while others change body region (e.g., diaphragm injuries move from the Thorax to the Abdomen body region). The changes in NASS body region assignments affect analyses using Maximum AIS (MAIS) by NASS body regions. (MAIS is the highest single AIS code in a patient with multiple injuries and has been used by investigators to describe overall severity.)[5]

TABLE 6
ORIGINAL AND CONVERTED NASS SYSTEMS
Full Aspect Expansion

Severity	NASS88		NASS93		NASS93 CONVERTED	
0	1	0.0%	1	0.0%	5	0.2%
1	1538	27.0%	955	30.1%	705	22.2%
2	2139	37.5%	971	30.6%	918	28.9%
3	1213	21.3%	771	24.3%	931	29.3%
4	415	7.3%	248	7.8%	347	10.9%
5	200	3.5%	193	6.1%	235	7.4%
6	38	0.7%	22	0.7%	27	0.9%
7	154	2.7%	15	0.5%	8	0.3%
Totals	5698	100.0%	3176	100.0%	3176	100.0%

The conversion also addresses problems which may arise when using ISS with NASS data. ISS is sensitive to changes both in severity number and changes in the allocation of specific injuries to ISS Body Regions. Without correction by the conversion, the relocation of injuries to different ISS Body Regions caused by the revisions of the systems will affect the ISS for that occupant record.

MAIS has been more widely used with NASS data than ISS. However, prior to this conversion there was no specification in the NASS systems of which injuries were assigned to the six ISS body regions. **Without ISS body region assignment it is not accurate to compare NASS injury ISS data to hospital ISS injury data because body regions are defined differently.**

Both AIS and ISS are being used to evaluate emergency medical care and state-wide trauma systems, [8,9] so the need for ISS and Body Region comparisons between NASS and AIS data may increase. The

conversion provides a tool for identifying and adjusting for differences in body region and severity for individual injuries or groups of injuries between the NASS systems.

IMPACT OF THE CONVERSION METHOD ON TRENDED OR COMBINED INJURY DATA

Sample Case 1 illustrates how the differences in NASS-88 and NASS-93 discussed in previous sections can affect the accuracy of combined or trended data and how the conversion adjusts for these affects. A sample motor vehicle accident injury is shown as it would be described in data coded with NASS-88 and NASS-93.[1,2] The NASS-88 column indicates how the injury would be classified by severity in data collected from 1988 to 1993, while the NASS-93 data shows how the data would be classified by severity in the 1993 and later data. The converted data column shows that adjustments to NASS-93 can make it comparable to data coded in NASS-88. In Sample Case 1, the conversion generates a MAIS and ISS that are the same as the NASS-88 coded data.

Sample Case 1 also shows the use of additional qualifying information such as age of patient or blood loss from injury that is used in NASS-93 but not in NASS-88.

Sample Case 1

An 11 year old child sustains a fractured femur, a dislocated knee and a LeFort III fracture of the maxilla (with ≤ 20% blood loss) in a motor vehicle accident.

Injury Description	NASS-88	Unadjusted NASS-93	Converted NASS-93 To NASS-88
ISS Body Region - Extremities or Pelvic Girdle:			
Right femur fracture, NFS, (age <12)	TRFS-3	851802.2-1	TRFS-3
Right knee dislocation, NFS.	KRDJ-3	850806.2-1	KRDJ-3
MAIS for Body Region:	3	2	3
ISS Body Region - Face:			
LeFort III fx of maxilla (with blood loss ≤ 20% by volume), aspect is central	FCFS-4	250808.3-4	FCFS-4
MAIS for Body Region:	4	3	4
MAIS Overall	4	3	4
ISS	$3^2 + 4^2 = 25$	$2^2 + 3^2 = 13$	$3^2 + 4^2 = 25$

Qualifying information can change the level of severity of an injury that otherwise would have the same severity between the systems.

Although the difference in maximum AIS between the two systems is only 1 severity point for each ISS Body Region and overall, the difference in overall severity, as calculated using ISS is 12 points. Depending on the criterion used, this difference can mean the severity of the NASS-88 coded case would be classified as major trauma, while the severity of the NASS-93 coded case might not be. (ISS greater than or equal to 13 or 16 criteria have been used in studies to identify major trauma patients.[9])

ISS has been shown to correlate with mortality.[10,11]. The survival rate for a patient under age 55 with blunt trauma and ISS=13 is approximately 25% better than that for a patient with ISS=25.[12] A researcher observing that the older NASS-88 coded data shows a more severe injury and the more recent NASS-93 coded data shows a less severe injury might attribute the reduction to the effects of safety interventions, if these were present in the vehicle with the lower severity score. **The difference in ISS and the survival rates observed in this example are "phantom" and are entirely attributable to the changes between the NASS-88 and NASS-93 systems.** The conversion can help adjust for these "phantom" results.

Since the ISS uses the highest severity from the three most severely injured ISS Body Regions, cases where injury codes change ISS Body Regions between the systems can also impact the ISS just as the severity changes did in Sample Case 1.

Sample Case 1 illustrates the impact of the revisions between NASS-88 and NASS-93 on MAIS and ISS. The ISS value can be substantially increased or decreased when severity numbers change by only 1 point. In special situations resulting from the conversion, other major changes in ISS values can occur. For example, a NASS-93 injury code of 7 that converts to AIS-3 could have a large impact on the ISS. The change from a 7 would allow the computation of an ISS for the occupant (as long as none of the other injuries present was a 7).

Injury codes that convert from a severity of 1-6 to a 0 ("no injury") also impact ISS and MAIS. These represent injuries present in NASS-93 that are not listed as injuries in NASS-88. For example, NASS-93 lists 8 injury codes as "other trauma", whereas NASS-88 doesn't list any in this category.[1,2] All of these codes (which are related to either inhalation injury or hypothermia) were converted to AIS-0.

TABLE 7

YEARLY NASS FULL ASPECT INJURY DATA, 1988-1994

Severity	NASS88 1988 Full Aspect Data		NASS88 1989 Full Aspect Data		NASS88 1990 Full Aspect Data		NASS88 1991 Full Aspect Data		NASS88 1992 Full Aspect Data	
0	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
1	26152	75.7%	22991	73.5%	25950	78.2%	20647	77.4%	24928	75.0%
2	4641	13.4%	4489	14.3%	4273	12.9%	3515	13.2%	4692	14.1%
3	2171	6.3%	2307	7.4%	1840	5.5%	1513	5.7%	2414	7.3%
4	656	1.9%	648	2.1%	416	1.3%	391	1.5%	563	1.7%
5	310	0.9%	259	0.8%	193	0.6%	155	0.6%	211	0.6%
6	155	0.4%	144	0.5%	104	0.3%	91	0.3%	84	0.3%
7	476	1.4%	447	1.4%	426	1.3%	361	1.4%	344	1.0%
Totals	34561	100.0%	31285	100.0%	33202	100.0%	26673	100.0%	33236	100.0%

Total, 1988-1992 = 158957 Injuries

Severity	NASS93 1993 Full Aspect Data		NASS93 1994 Full Aspect Data	
0	0	0.0%	0	0.0%
1	22230	75.2%	19441	74.5%
2	3910	13.2%	3630	13.9%
3	2030	6.9%	1844	7.1%
4	668	2.3%	585	2.2%
5	348	1.2%	319	1.2%
6	86	0.3%	101	0.4%
7	276	0.9%	171	0.7%
Totals	29548	100.0%	26091	100.0%

Total, 1993-1994 = 55639 Injuries

CONVERTED

Severity	NASS93 1993 Full Aspect Data		NASS93 1994 Full Aspect Data	
0	1	0.0%	1	0.0%
1&2	25257	85.6%	22191	85.1%
3	2633	8.9%	2473	9.5%
4	891	3.0%	804	3.1%
5	391	1.3%	350	1.3%
6	130	0.4%	132	0.5%
7	220	0.7%	126	0.5%
Totals	*29523	100.0%	*26077	100.0%

*Total, 1993-1994 = 55600 Injuries

*Note: 39 injuries not defined in NASS93 manual filtered out, 25 in 1993, 14 in 1994

Because these injury codes were assigned severities of 1-5 in NASS-93, **their conversion to AIS-0 in NASS-88 can effect the ISS by as much as 25 (5^2) points.**

To determine the effects of severity adjustment from NASS-93 to NASS-88 on the NASS data base, all the NASS-93 injury codes contained in the data for years 1993 and 1994 were converted. This converted NASS data was then compared to unadjusted data. Table 7 shows the conversion's affect on the NASS data. The distribution of injuries by severity by year is shown for each year of NASS data 1988-1994, compared to the converted data for 1993 and 1994.[13,14,15,16, 17,18,19] Table 7 shows that the percentage distributions of injuries by severity have been very steady over the years. Using unadjusted data for the most recent three years, the percentage of injuries that is AIS-3 or higher in severity (excluding AIS-7) ranged from 9.8% in 1992 to 10.9% in 1994.

However, when the conversion is used to adjust the data collected using the NASS-93 system, (1993-1994), a shift in injuries by severity levels occurs. The converted 1993 data show that 13.7% of all injuries are AIS-3 or higher in severity (excluding AIS-7), with a net gain of 913 injuries in this category. The converted 1994 data show that 14.4% of all injuries are AIS-3 or higher in severity (excluding AIS-7), with a net gain of 910 injuries in this category. Table 8 summarizes the percentage change of injuries between converted and unconverted NASS-93 data for the combined years 1993-1994.

For both the 1993 and 1994 converted data, the combined percentages of 1 and 2 level injuries declines slightly compared to the unconverted data, about 3% for each year. For 1993, the conversion increases the

TABLE 8 SUMMARIZED NASS INJURY DISTRIBUTIONS								
Severity	NASS88 1988-1992 Combined Full Aspect Data		NASS93 1993-1994 Combined Full Aspect Data		CONVERTED NASS93 1993-1994 Combined Full Aspect Data		CHANGE CONVERTED vs. UNCONVERTED Data	
0	0	0.0%	0	0.0%	2	0.0%	2	
1&2	142278	89.5%	49211	88.4%	47448	85.3%	-1763	-3.6%
3	10245	6.4%	3874	7.0%	5106	9.2%	1232	31.8%
4	2674	1.7%	1253	2.3%	1695	3.0%	442	35.3%
5	1128	0.7%	667	1.2%	741	1.3%	74	11.1%
6	578	0.4%	187	0.3%	262	0.5%	75	40.1%
7	2054	1.3%	447	0.8%	346	0.6%	-101	-22.6%
Totals	158957	100.0%	55639	100.0%	*55600	100.0%	*-39	

*Note: 39 injuries not defined in NASS93 manual filtered out

number of AIS-3 injuries by almost 30% (from 2030 to 2633). For the 1994 data, the conversion increases the number of AIS-3 injuries by 34% (from 1844 to 2473).

Another interesting aspect of the data is the change in AIS-5 injuries when NASS-93 was adopted in 1993. (Shown in Table 7) There was an increase of nearly 65% in the number of AIS-5 injuries recorded between 1992 and 1993. An examination of the data reveals that this may be the result of changes in the coding of certain AIS-5 (critical) injuries. These injuries existed in NASS-88, but as severity AIS-3. The change in the coding system had the affect of moving these injuries from AIS-3 to AIS-5. These same codes appear to also cause the relatively moderate 11% increase in AIS-5 injuries in the converted NASS-93 data for 1993 and 1994. Injuries at severities 3, 4 and 6 increase several times this amount. Because the codes in question are equivalent to AIS-3 in NASS-88, the conversion correctly assigns these injuries to the 3 severity, thus greatly reducing the number of AIS-5 injuries in the converted data. Thus the same affect that causes the shift in AIS-5 level injuries between 1992 and 1993 also produces the lower increase in AIS-5 injuries in the converted data.

CONCLUSION

NHTSA, through the NASS/CDS has collected highly valuable data on a sample of automobile crashes over the past decade. However, the occupant injury data for the past 8 years is collected in two different formats, one for the five years ending in 1992 and the second for the years 1993 to date. This paper discusses a method to convert the 1993 and later data to be compatible with the earlier 5 years of data. The paper covers the following points regarding this conversion.

1. Because the NASS system is a sampling system, only a relatively small number of cases are available for each year (about 5,000). The cases are expensive to collect because of the depth of information included, and are therefore very valuable. It is unlikely that large numbers of additional cases can be easily or quickly collected per year, due to budgetary and other constraints.

2. The relatively small number of cases means that certain types of injuries (including severe ones such as head and spinal trauma) may have very low occurrences in each year. The occurrences may be so low as to confound statistical analysis. Of all the defined severe or higher injuries (AIS 4, 5, 6) n=421 of 478 (88%) have 10 or less occurrences in the 1993-94 data. A large percentage of defined injury codes of all severities n=1722 of 3176, 54% have 0 (zero) occurrences in the 93-94 data.

3. Additional cases that are available in the occupant injury data collected prior to 1993 are unavailable for use because this data is not compatible with the data collected for the years 1993 to date. This incompatibility is due to major differences in both the injury descriptor codes and the basis used to assign the injury severity.

4. The conversion discussed converts the injury descriptions and normalizes the severities to allow data from the years 1988-1992 and 1993 and later to be combined or trended. This expands (by about 4 times) the number of injuries available for analysis. The injury pool for the 1993 and 1994 years of data (about 55,000 injures) can be augmented, using the conversion with the five earlier years of data (about 160,000 injuries) to create a sample of 215,000 injuries.

5. The conversion "normalizes" the severities of the injuries in the two systems to a common baseline. When applied to the 1993-1994 NASS injury data, the conversion produces an overall increase of about 30% in the percentage of AIS-3 and higher injuries (excluding AIS-7) when the data are converted to be consistent with NASS-88. Although this conversion makes many changes (increases and decreases in severity) the net effect of the conversion on 1993-1994 injury data is an adjustment for the reduction in severity scores that has occurred in the current system as a result of advances in medicine and revisions to the severity scoring system. Without this normalization it is possible to confuse reductions in severity that are a result only of the scoring system used with the real effects produced by safety interventions.

6. A case example is given to show the effect of the conversion. The case illustrates on a "micro" scale how combining the original and converted data can increase the number of serious (AIS-3 or higher) injuries available for study. The case also illustrates the effect of the conversion's "normalization" of severity, and how the conversion adjusts what otherwise would be a substantial difference in the overall patient severity score (for the example shown an ISS of 13 vs 25, and a maximum AIS of 4 rather than 3). The conversion difference in the example was enough to change the classification of the injuries between major and non-major trauma, solely on the basis of the severity scoring system artifacts.

The authors believe that the conversion is an important tool for adjusting a substantial number of differences in severity, body region and aspect between the NASS-93 and NASS-88 systems. In addition, it can be used as a method to identify injuries that were described differently between the two systems and as an enhancement to the data filtering process.

ACKNOWLEDGMENTS

The authors wish to thank the Insurance Institute for Highway Safety for its interest in this area of work and gratefully acknowledge the information or assistance provided by Elaine Petrucelli, Evelyn Benton, Ruth Isenberg and Charlie Compton.

ABOUT THE AUTHOR

Elizabeth A. Garthe is President of Garthe Associates; she and her colleague, Nicholas K. Mango work with national organizations, academic centers and trauma centers on projects involving injury classification systems, including the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), the Abbreviated Injury Scale (AIS) and the NASS Occupant Injury Classification (OIC). Ms. Garthe received the John States award from the AAAM in 1981 for her paper titled "Compatibility of the ICD-9-CM with AIS-80" and contributed to the development and update of a conversion of ICD-9-CM injury rubrics to AIS-90 severity. Contact information: Garthe Associates - Tel: 617-631-1553; FAX: 617-631-2146.

REFERENCES

1. National Highway Traffic Safety Administration. 1993. National Accident Sampling System 1993 Crashworthiness Data System Injury Coding Manual. Washington, DC: National Highway Traffic Safety Administration. Editor: Calspan Corporation.

2. National Highway Traffic Safety Administration. 1988. National Accident Sampling System 1988 Crashworthiness Data System Injury Coding Manual. Washington, D.C.: National Highway Traffic Safety Administration.

3. Garthe, EA, "Comparison of the AIS-85 and AIS-90 with NASS-93," Paper presented at the 1996 Society of Automotive Engineers Annual Conference: Detroit, MI. (SAE Technical Paper No. 960095).

4. Association for the Advancement of Automotive Medicine. 1985. The Abbreviated Injury Scale, 1985 revision. Des Plaines, IL: Association for the Advancement of Automotive Medicine.

5. Association for the Advancement of Automotive Medicine. The Abbreviated Injury Scale, 1990 revision. Des Plaines, IL: Association for the Advancement of Automotive Medicine.

6. National Highway Traffic Safety Administration. 1993. National Accident Sampling System 1993 Crashworthiness Data System Injury Coding Manual. Washington, DC: National Highway Traffic Safety Administration. Editor: Calspan Corporation. Revised manual - obtained 3/96 from Ruth Isenberg.

7. Garthe, EA, Ferguson, S, Early, N: ""A Method for Converting Injury Severity in NASS-93 (AIS-90) to NASS-88 (AIS-85)". 40th Annual Proceedings, Association for the Advancement of Automotive Medicine, Des Plaines, IL, 477-493; 1996.

8. MacKenzie, EJ: "Injury Severity Scales: Overview and Directions for Future Research," *Am. J. Emergency Med*, 2: 537-549, 1984.

9. MacKenzie, EJ, Steinwachs, DM, and Ramzy, AI. "Evaluating Performance of Statewide Regionalized Systems of Trauma Care," *J. Trauma*, 30: 681-688, 1990.

10. Baker, SP, O'Neill, B, Haddon, W, Jr, et al.: "The Injury Severity Score: A Method for Describing Patients with Multiple Injuries and Evaluating Emergency Care," *J. Trauma*, 14: 187-196, 1974.

11. Baker, SP, O'Neill B: "The Injury Severity Score: An Update." *J. of Trauma* 16:822, 1976.

12. Moore, E, Mattox, K. and Feliciano, D., ed., Trauma, 2nd edition, Appleton and Lange, Norwalk, Connecticut, 1991.

13. National Highway Traffic Safety Administration. 1989. National Accident Sampling System 1988 Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration.

14. National Highway Traffic Safety Administration. 1989. National Accident Sampling System 1989 Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration.

15. National Highway Traffic Safety Administration. 1990. National Accident Sampling System 1990 Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration

16. National Highway Traffic Safety Administration. 1991. National Accident Sampling System 1991. Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration

17. National Highway Traffic Safety Administration. 1992. National Accident Sampling System 1992. Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration

18. National Highway Traffic Safety Administration. 1993. National Accident Sampling System 1993. Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration

19. National Highway Traffic Safety Administration. 1994. National Accident Sampling System 1994 Crashworthiness Data System. Washington, DC: National Highway Traffic Safety Administration.

970133

Rotating Wheels - Their Impact on Wind Tunnel Test Techniques and on Vehicle Drag Results

G. Wickern, K. Zwicker, and M. Pfadenhauer
Audi AG

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The question of the proper simulation of wheel rotation has not so far been a major concern. Within the scope of an examination of the influence of wheels and tyres on aerodynamic drag it will be shown that their contribution to the overall drag value - whether they are rotating or not - is of about the same magnitude as the proportion of the rough underbody. Therefore the question of the importance of the simulation of wheel rotation is posed.

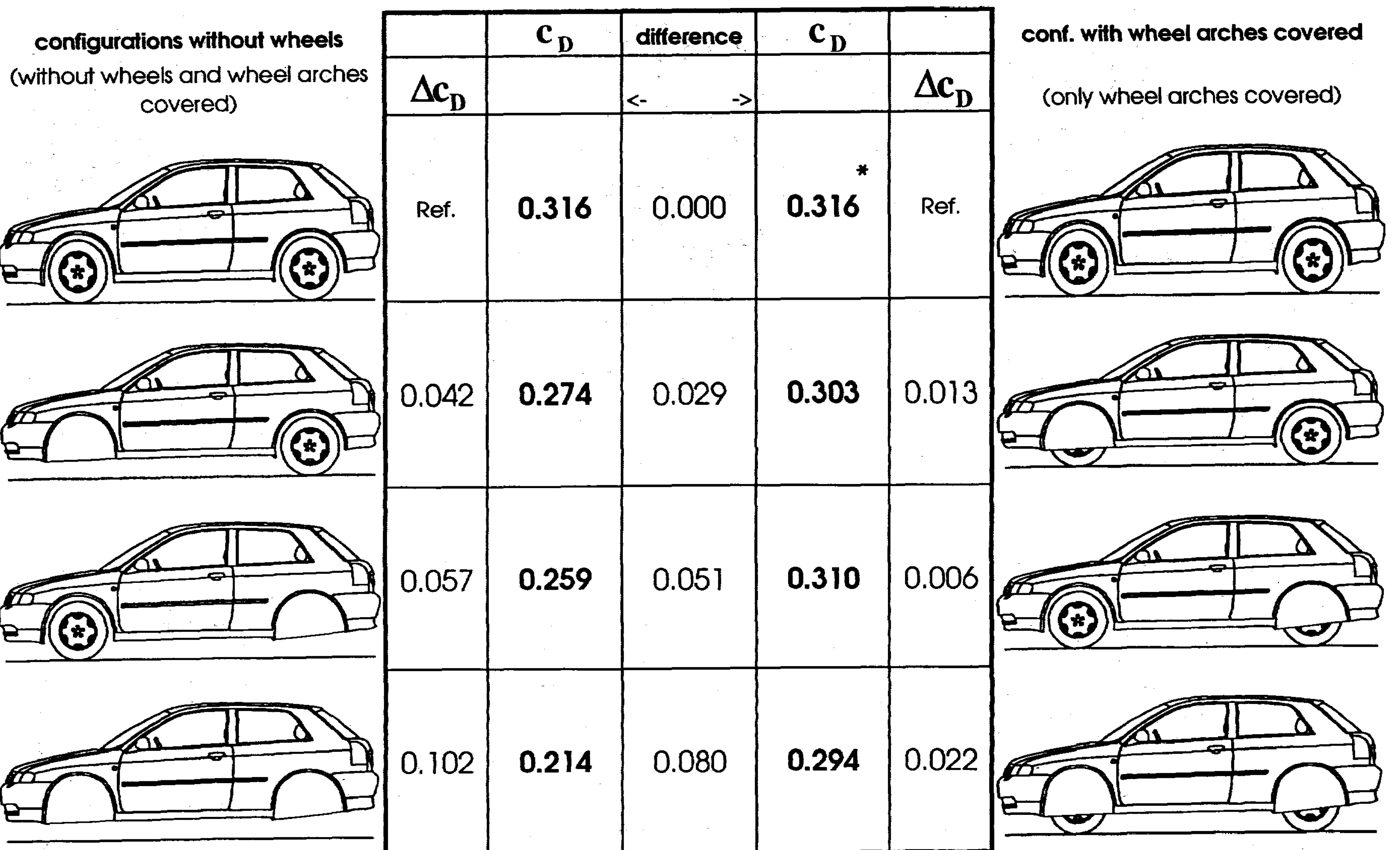
This paper discusses how a measurement with a better simulation can look like and what the major changes in the flow field are. In particular a new physical quantity, which has to be determined, the so-called "fan moment" is introduced. . The problems that arise in the determination of the fan moment of the wheels and hence in the required isolation of the rolling resistance, are described in detail. This is done for a test set up with full width moving belt and measurement via internal balance and sting support. An alternative set up with separate wheel drives and external balance is discussed briefly.

In a presentation of vehicle test results it is shown that the simulation of wheel rotation has a significant effect on the aerodynamic resistance. The dependence of the effect on the wheel/ tyre configuration is demonstrated in several examples. Finally, using the experimental results from a special test rig with a single wheel, the manner in which the tyre contact area and the yaw angle effects the forces at the rotating wheel is shown.

INTRODUCTION

For a long time, the influence of wheels and tyres on vehicle aerodynamics was an aspect of aerodynamic development only in terms of tyre width and flush wheel covers. The area of concentration was on optimization of the basic vehicle body. Nevertheless, since a level has been reached in the optimization of the visible outer shell that – considering stylistic matters – is difficult to improve upon, aerodynamic engineers have given their attention to the underside of the vehicle to a greater degree. In the course of exploiting the optimization potential of the underside of the vehicle, the conventional method of measuring aerodynamic values – with a stationary vehicle located in a fixed position in a wind tunnel – has been questioned. For a better simulation of the actual conditions on the road, several different groups have tried out the installation of a moving belt underneath the vehicle, which modifies the underbody flow to correspond better to the real situation. The first moving belt measurements with a production car (see Mercker and Knape [1]) showed that not just a stationary ground surface but also stationary wheels result in inadequate simulation. The simulation of rotating wheels, however, puts new demands on wind-tunnel measurement technology, which go well beyond difficulties in the simulation of the moving floor as an influence on the boundary layer.

Here the question arises as to what degree it makes sense to take these additional measures and to consider the percentagein drag of the wheels and tyres as equivalent to the percentages of the vehicle's shape and underbody. Furthermore, we should discuss what the fundamental options are for determining this percentage, and what degree of accuracy of simulation must be demanded, for example, how important the simulation of the tyre contact area could be.



Δc_D	c_D	difference <- ->	c_D	Δc_D
Ref.	**0.316**	0.000	**0.316***	Ref.
0.042	**0.274**	0.029	**0.303**	0.013
0.057	**0.259**	0.051	**0.310**	0.006
0.102	**0.214**	0.080	**0.294**	0.022

* Values have been measured with stationary wheels in a different wind tunnel an d have been shifted to same level as left column for ease of comparison

Fig. 1: Test series to determine the influence of wheel and wheelarches measured with Audi A3 with realistic underbody, tyre dimension 195/65 R 15

THE PROPORTION OF THE WHEELS IN THE OVERALL DRAG OF A PRODUCTION VEHICLE

The significance of the wheels' contribution can be demonstrated with the results of a few basic experiments that were conducted by Audi on this issue. The results from Mercker et al. [2] served as the starting point. They measured a production vehicle without wheels and with covered wheel wells and thus deduced the percentage of the wheels and wheel wells on the aerodynamic drag. They determined a percentage of approximately 33 % of the overall aerodynamic drag for the vehicle, but in a configuration with a smooth underbody. This experiment was repeated by Audi with a production car with realistic underbody (Audi A3); additionally investigations were made with different variations to determine the influence of the front and rear wheels and of the wheels and wheel arches, at least in terms of order of magnitude.

Figure 1 shows the summary of the experimental results. In a series of experiments, the vehicle was measured over a moving belt in the following states while suspended from a rear-located support:

- initially with all four wheels in position (wheels rotating)
- once with only the rear wheels installed and the front wheel arches blanked off
- once with only the front wheels installed and the rear wheel arches blanked off
- finally with no wheels and all wheel arches blanked off.

As the results show, the wheels exert a considerable influence on overall drag, with this being initially distributed almost uniformly between the front and rear wheels.

However, the proportion attributable to the wheels still contains an element due to the wheel arches, which it would not be logical to allocate to the wheels themselves. In order to determine this proportion the sheet metal covers placed over the wheel arches in the first series of experiments were cut away so that they could be installed with the wheels in position, but would seal the gap between wheel and wheel arch. In a second series of experiments (now in an conventional wind tunnel), the vehicle always stood on all four of its wheels, but measurements were taken with the covers installed over the wheel arches in a similar manner to the first series of experiments. It was found that the proportion accounted for by the open wheel arches is distinctly smaller than the proportion from that part of the wheel which is not covered by the body (this result was also obtained in similar form by Mercker et al [1]). If the effects at the front and rear are compared, a common experience is confirmed, namely that sealing off the wheel arches at the front is much more effective than at the rear. At this point it should be noted, that the c_D - values obtained in the second test series (with fixed ground) to determine the influence of the wheel wells were adjusted to the same level as the results of the first series for configuration 1. This was done for ease of comparison. Strictly spoken, only the Δc_D of the second series are valid, and they are exact only for the case with stationary wheels.

A Δc_D of 0.08 or a 25 % of the total vehicle drag remains for the protruding part of the wheel. This Δc_D of 0.08, interestingly enough, corresponds precisely to the proportion which Hörner [3] established as long ago as 1951 for vehicles from the nineteen-thirties. However, in those days the complete vehicle's aerodynamic drag coefficient was $c_D = 0.75$. According to Mercker et al. [2], the Δc_D for the wheels was also in this order of magnitude for a car with flush underbody.

It is also noticeable in these results that after substraction the wheel-arch proportion, the rear wheels exhibit a higher level of drag. This may be due to the rear-end outline of the vehicle investigated (squareback), and should therefore not be regarded as generally valid without further investigation.

The proportion of the vehicle's aerodynamic drag which originates at the wheels is not necessarily identical with the aerodynamic force acting directly on the wheels. We assume for the purpose of estimating an order of magnitude that the interference drag (caused by the wheels at the body) is low. Then if the wheels' proportional drag is related to the sum of the wheels' frontal areas (only the area not covered up) a c_D value of approximately 1 is obtained. In other words, the wheels' aerodynamic quality is about as unfavourable as a vertical plate exposed to the airflow.

The results of the experiment shown in *Fig. 1* can be summed up as follows:

- The proportion of the wheels in the total drag of a production vehicle can easily amount to 25 %. The cause of this high figure is the unfavourable shape of the wheel when exposed to the airflow.

Furthermore the following additional conclusions can be drawn:

- Optimization of the wheels must be given the same importance as improving the underbody of the vehicle, since considerable potential is available here not only in terms of the absolute value for the proportional drag but also the poor aerodynamic quality as a starting level.
- Accordingly, simulating the rotation of the wheels must be given the same priority as simulating the movement of the ground surface beneath the vehicle.

However, the correct measurements with the wheels rotating cause certain difficulties, which will be discussed in the following sections.

BASIC CONSIDERATIONS FOR WIND-TUNNEL TESTS WITH ROTATING WHEELS

In order to measure a vehicle with rotating wheels correctly in the wind tunnel, we must first analyse which forces and moments act additionally on the vehicle in case of rotating wheels in contrast to a conventional test set up. *Fig. 2* illustrates these forces.

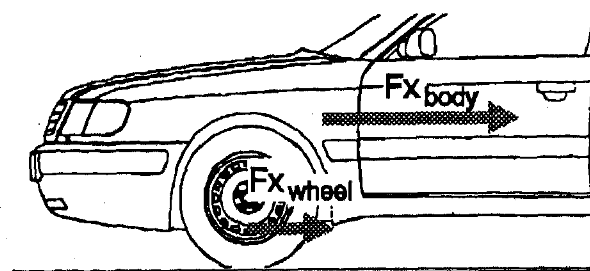
Considering first a stationary wheel, the airflow forces can be divided into an element acting on the body and one acting on the wheel. The force acting on the wheel is normally applied off-centre, below the axle in the case of a wheel partly shielded by the wheel arch.

If the wheel is set in rotation by a suitable simulation technique, the flow field around the wheel changes, first of all because of the moving wheel surface (which creates different boundary conditions for the boundary layer). Secondly, in case that air is passing through the wheel this air is given a rotation by the wheel spokes (acting similar to the blades of a fan). As a result, the magnitude and point of attack of the aerodynamic forces on the wheel change. The change in airflow around the wheels, however, also gives rise to an indirect effect on the body, that is to say the interference drag. Conventional measuring techniques are unable to separate these two partial forces, but if the wheels are independently suspended this ought to be feasible.

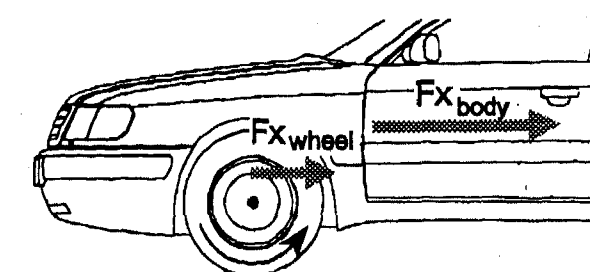
A third value has to be taken into account in the case of rotating wheels: the moment round the wheel axis. As has already been shown on the stationary wheel, the airflow force does not act at the centre point of the axle. However, this is the only point at which forces can be introduced to the vehicle's structure. If the force is displaced to the centre point of the wheel, an additional moment is caused; if the wheel is stationary, this acts in the direction of forward travel. If, in other words, the wheel were to be attached to the axle and supported on the ground without friction, the wind load would start it rotating in the forward travel direction. This supporting effect of the airflow force is reversed, however, at full wheel speed on account of the fan-blade effect of the wheels. The moment reverses its sign and the vehicle's engine has to supply more power in order to maintain wheel rotation at a constant speed. The necessary drive moment for this is referred to below as the "fan moment". The resultant force on the wheel, in view of the sign change in the airflow -force moment when the wheel is rotating, is displaced to an effective line above the axle (*Fig. 2*). (This change in the sign of the moment will normally occur in all cases which are of technical interest, though cases can certainly be envisaged in which it does not occur, possibly because the fan-blade effect is small.)

The fan moment can be regarded as being a combination of three elements:

the aerodynamic force on the wheel acting off-centre



forces at the stationary wheel



forces at the rotating wheel

Fig. 2: Forces acting on the rotating wheel in comparison with those at the stationary wheel

- the fan-blade effect, which gives a rotational component to the flow through the wheel
- the skin friction of the rotating wheel in relation to the oncoming flow field.

All three are related to the square of the speed: The aerodynamic force rises with the square of the flow velocity v (primarily pressure forces); the fan-blade effect causes a moment rising with the square of angular velocity, i.e. speed of the vehicle and at last, the skin friction is scaled with v^n where $n \approx 2$ (a detailed discussion of the skin friction term will be given later).

The wheel's fan moment can therefore be converted into a contribution to the aerodynamic drag coefficient (a Δc_D). Since the contribution to the vehicle's driving resistance as a result of the fan moment is caused by air friction, can only be determined in the wind tunnel and can only be optimised by aerodynamic measures, it would appear logical to add this amount to the c_D value in practice. On the other hand the c_D value is by definition a nondimensional force, and there might be applications where the pure force is of interest only. Therefore a two term representation of these two aerodynamic driving resistance portions seems to make

sense. For a complete description of the driving resistance as needed for example in performance calculations both drag force and fan moment have to be taken into account. Nevertheless, for ease of comparison with the amount of total drag the share of wheel ventilation will be expressed in terms of Δc_D (instead of a moment coefficient) in this paper.

However, the technical problems of determining the fan moment are considerable. This is discussed below with reference to measuring techniques already being state of the art.

MEASURING SET-UP WITH AN INTERNAL BALANCE AND THE WHEELS DRIVEN BY A MOVING BELT WHICH SIMULATES THE ROAD SURFACE

As already mentioned in the introduction, the classic procedure when simulating movement of the road surface beneath the vehicle is to perform the measurements on a moving belt. The vehicle is supported on a bar ("sting support") from the rear or the top (see *Fig. 3*) and the forces measured with an internal balance. The test set-up used in this investigation has already been described in detail by Mercker et al. [1, 2], so in here reference will only be made to special features of measurement with rotating wheels. In this test procedure the wheels are allowed to rest on the moving belt with a low load (< 50 N) and thus forced to rotate. To minimise friction losses it is customary to remove the drive shafts and the brake pads or linings, and to rework the wheel bearings.

Although it is the belt which supplies the power needed to overcome the fan moment, this force is evident again at the internal balance on which the model and therefore the wheel are supported. The balance therefore measures aerodynamic drag as an integral value containing all the elements discussed above. Unfortunately the measured values also include the wheels' rolling resistance, which must therefore be determined and subtracted. As a common solution, a tare measurement ("zero measurement") with the wheels rotating is suggested, but this gives rise to fresh sources of error and will therefore be discussed in detail in a separate section.

ELIMINATION OF ROLLING RESISTANCE BY TARE MEASUREMENTS (INTERNAL BALANCE)

When measuring with rotating wheels, the problem referred to here occurs when the question of correct tare measurement arises. In order to examine this, we should first recall the drag forces which occur when the actual measurement on exposure to the airflow takes place (see *Fig. 4*).

1 Possibly mass forces

2 Rolling resistance of the wheels, since these are driven by the belt and therefore measured by the balance as a reaction force. This consists of
- tyre rolling resistance
- wheel bearing friction

3 Aerodynamic force on the body

4 Aerodynamic force on the wheels

5 Force needed to overcome the fan moment of the wheels when exposed to the airflow (also a reaction force, as **2**)

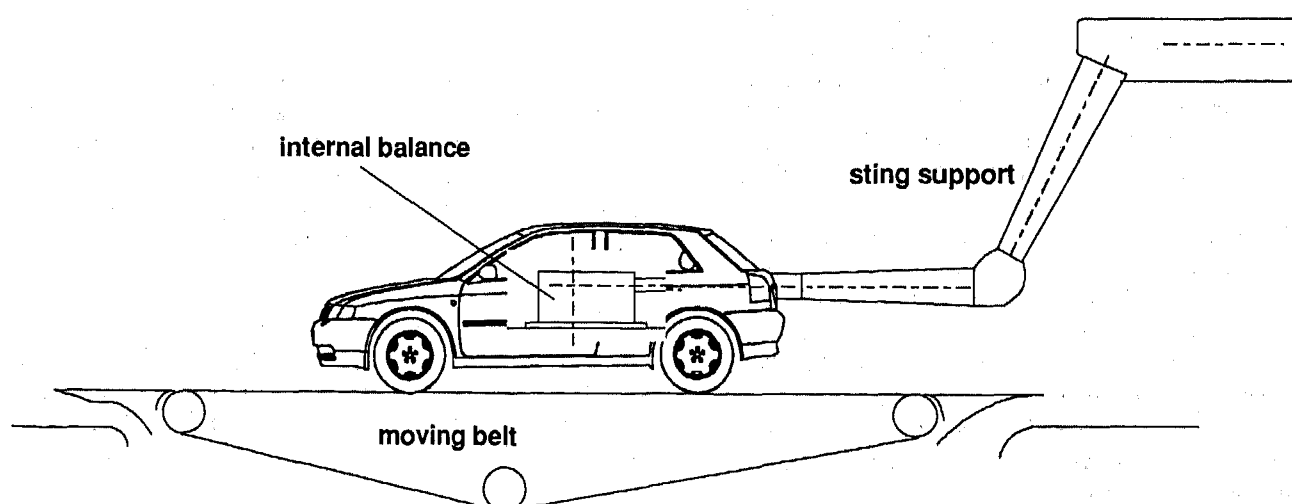


Fig. 3: Set-up for measurements with internal balance (wheels driven by moving belt)

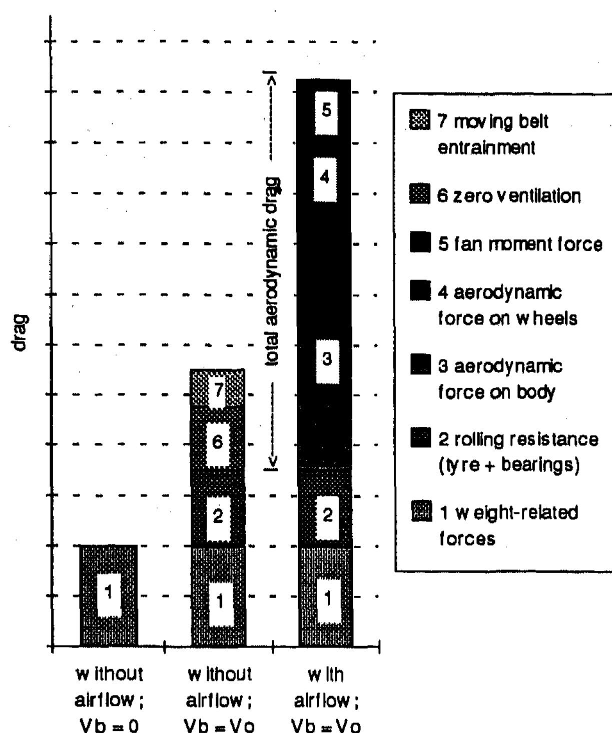


Fig. 4: Forces in the x direction when measured with the internal balance

In addition to forces **3** and **4** the total aerodynamic force must also contain the component from fan moment **5** (see section 2), so that the sum of 3, 4 and 5 must be separated from 1 and 2. This is done in the conventional way with a tare measurement. By analogy to the measurement with stationary wheels, it can be performed without airflow and with the moving belt stationary (V_b = 0). It is common practice, however, to eliminate rolling resistance by performing the tare measurement without airflow but with the moving belt in operation ($V_b = V_0$).

The forces measured in each case here can also be seen from *Fig. 4*. Although the tare measurement with moving belt in operation, as often used today, takes the rolling resistance correctly into account, additional terms arise which have nothing to do with determining the aerodynamic forces:

6 Zero ventilation (fan moment of the wheels without the airflow)

7 Forces on the vehicle body caused by the entrainment effect of the moving belt (boundary layer)

A tare measurement with the belt stationary, however, merely takes into account the forces generated by the vehicle's weight **1** and therefore also supplies an incorrect value.

The magnitude of these errors was determined by the following experiment, which also provides a means of separating the fan moment and rolling resistance components:

For the vehicle to be measured (Audi A3) a tare measurement at V_b = 0 was carried out, so that any weight-related forces **1** were eliminated. Next, the forces without airflow were determined as a function of moving belt speed from V_b = 0 to V_0 (here 140 km/h), in other words a variation of the tare method within the limits of the methods described above. Accordingly, this measurement contains only components from rolling resistance **2**, zero ventilation **6** and moving belt boundary layer effect **7**. Two cases were investigated:

- first a standard summer-tread production-car tyre on a wheel with a series-production wheel cover (see *Fig. 14.1* in the Annex)
- second a tyre with no tread but otherwise with a similar external shape (a "slick") and a smooth, full-diameter wheel cover (see *Fig. 14.2* in the Annex) to represent the ideal situation in terms of zero ventilation).

Fig. 5 shows the measured drag force in both cases.

Below 10 km/h probably non-linear incipient rolling processes occur (in the moving belt system, in the tyre and bearings) which need not be examined here. Above this speed the forces run in a very similar manner in the x-direction. Assuming that zero ventilation **6** changes with the square of wheel speed and therefore with V_b^2, and that the force **7** generated by belt speed also changes with V_b^2, component 6 + 7 can be separated off. This assumption is correct within the observed levels of accuracy to the extent that zero ventilation is primarily made up of a friction component at the surface of the wheel and tyre and a through-flow component.

The moment as an integral wall shear stress value increases with an exponent of 1.8 in the limiting case represented by a rotating disc in still air (see Schlichting [7] and Annex), whereas the moment from the through-flow component rises according to the square of the peripheral speed, in a manner analogous to rotating fluid dynamic machines.

The remaining force after substraction of the quadratic part in both cases is the same in magnitude within the limits of measurement accuracy, and remains more or less constant when plotted against belt speed from V_b = 20 km/h upwards (this limit undoubtedly depends on the vehicle being measured). As an initial approximation, this proportion of the total value can be identified as the tyres' rolling resistance plus wheel-bearing losses, that is to say as wheel rolling resistance **2**.

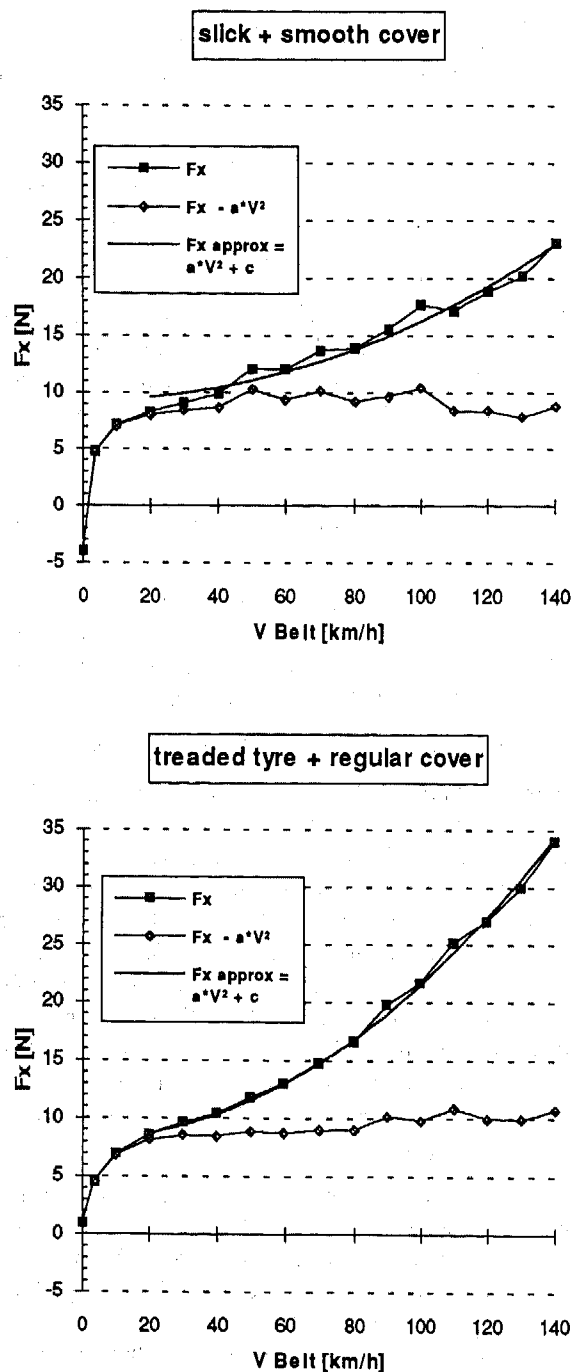


Fig. 5 : Force in x direction without airflow as a function of belt speed. Upper: plots for wheel with treaded tyre and regular wheel cover; lower: "slick" with smooth wheel cover. See text for further explanations

This means that the force in the x-direction is made up approximately from squared and constant elements. This can be shown by a formula such as $F_x = a\ V_b + c$ with an approximation of the measured values, for example by help of a least-square-of-error method; if applied to the two cases under investigation, the following results are obtained:

- The constant component is about the same in both cases

  F_{Roll} = app. 9 N

  This element is not taken into consideration when the tare measurement takes place with the belt stationary, that is to say the measured aerodynamic drag is too high by this amount, equivalent to Δc_D = 0.005 at V_o= 140 km/h (A=2.06 m²)

- The squared element (zero ventilation plus moving belt) depends on the wheel/tyre combination and must be determined afresh for each change of wheel. For the cases investigated, two typical results for V_b = 140 km/h are obtained:

  $F_{\text{zero ventilation}} + F_{\text{belt boundary layer}}$ =

  = 14 N ("slick" with smooth wheel cover)

  and

  = 24 N (treaded, with regular cover)

  When a tare measurement is carried out with the belt moving, this is accordingly the amount by which the aerodynamic drag will be too low; here it is equivalent to Δc_D = 0.007 or 0.012.

  The proportion caused by the moving belt's boundary layer effect is, in our experience, relatively small (app. 2 ... 3 N, measured with all four wheels lifted off the belt).

The values stated above show that the forces which occur cannot be neglected when undertaking aerodynamic drag measurements. It is therefore desirable to examine the forces relevant to the tare measurement more closely. The measurement of forces as a function of V_b, as described here, is a practicable method of separating rolling resistance from the fan-moment component.

The above reflections can be summarised as follows:

- c_D in the case of moving belt measurements with an internal balance is quoted as larger than previously by the amount represented by the zero ventilation and the belt boundary layer effect, when the tare measurement took place with $V_b = V_o$ The magnitude of this error is approx. Δc_D = + 0.010 (here between 0.007 and 0.012)
- If a different wheel/tyre is fitted, the change in zero ventilation must be taken into account as appropriate, since the wheels will otherwise be allocated a value

in the calculation which is incorrect by this amount (here $\Delta c_D = 0.005$)

Tare measurement at low belt speed, which can be carried out with the above procedure, is a simple and practicable solution to the problems of improving drag measurement as described here.

In order to check the magnitude of the zero ventilation, measurements were conducted using a different measuring technique and without the vehicle being present, on the tyre test rig operated by the Continental Tyre company. The above values for the two wheel combinations were confirmed to a most satisfactory extent.

Zero ventilation is by definition the fan moment without incoming airflow. As *Fig. 4* shows, separating the fan moment with airflow from the other aerodynamic forces is not possible in a set-up as shown in *Fig. 3*, since here only the integral value of the relevant forces can be determined. For correct separation the force which the rotating wheel exerts on the belt when exposed to the airflow would have to be determined on its own. Therefore no data is available on the amount of the fan moment with airflow for the configuration investigated here. Moreover, to our knowledge neither in the literature any data is available except for the work of Deubert [8], who published results of Mercedes-Benz. In the reported case the fan moment was measured in a conventional wind tunnel and the wheel was driven by an external electrical motor via torque-measuring shaft. The tyre dimension (195/65 R15) was exactly the same, whereas the wheel designs and the vehicle (MB W 124) were of course different. Nevertheless these results give us useful hints. If we compare the results for the zero ventilation of the treaded tyre with regular cover shown in *Fig. 5* with the summer tyres on steel wheels in the Mercedes-Benz results in *Fig. 7*, we find quite similar Δc_D -values. Thus the zero ventilation value maybe is a workable approximation if an initial estimate of the magnitude of the fan moment is required. Another interesting result is, that for the steel wheel with cover the fan moment for front and rear is nearly identical. For the more open (light) alloy wheel the results differ in a similar way as the results for the wheel arches (blanked off or not). Due to suction peak on the outside of the front wheels a larger effect is obtained there compared to the rear wheels. Not surprising is that the winter tyre with a more coarse tread has a larger fan moment than the summer tyre.

The magnitude of the fan moment is of special interest when comparing results of a test with full width moving belt and internal balance with results from test set-ups, where the fan moment is not included in the measured drag force. A such a set-up will be discussed in the next section.

wheel location	wheel configuration	Δc_D for single wheel
front wheel	steel wheel with cover and winter tyre	**0.0049**
"	steel wheel with cover and summer tyre	**0.0036**
"	alloy wheel with winter tyre	**0.0056**
rear wheel	steel wheel with cover and winter tyre	**0.0048**
"	alloy wheel with winter tyre	**0.0045**

wheel configuration	Δc_D for all wheels (2x front, 2x rear)
steel wheel with cover and winter tyre	**0.019**
steel wheel with cover and summer tyre	**0.014***
alloy wheel with winter tyre	**0.020**

*rear wheels taken as identical as front on the base of steel wheel + winter tyre results

Fig. 6: Results published by Mercedes Benz (see Deubert [8]) on the fan moment of wheels with tyres 195/65 R15

MEASUREMENTS WITH AN EXTERNAL BALANCE AND SPECIAL WHEEL DRIVE UNITS

In order to avoid the complexity of suspending the vehicle from a support arm, an alternative method of simulating rotating wheels was developed, with the wheel drive elements forming part of the weighed system. The forces can then be measured in the conventional manner, using an external floor-mounted balance. To complete the simulation, a moving belt can be installed between the wheels. An example for this type of facility is the Pininfarina wind tunnel, as described by Cogotti [4].

In the beginning of rhe development of this testing method the vehicle was placed on additional supports connected to the balance, and drove its own wheels by way of its all-wheel drive system (see for example Wickern [5]; Mercker, Soja and Wiedemann [6]). The balance plates were equipped with plain rollers or brush systems to prevent air from flowing under the rotating wheels. As a refinement the drive can be transferred to the balance plates, resulting in the set-up shown in *Fig. 7*. The drive is no longer part of the test vehicle, but is still within the weighed area of the balance, so that no power bypass can falsify the measurements. Compared with internal-balance measuring,

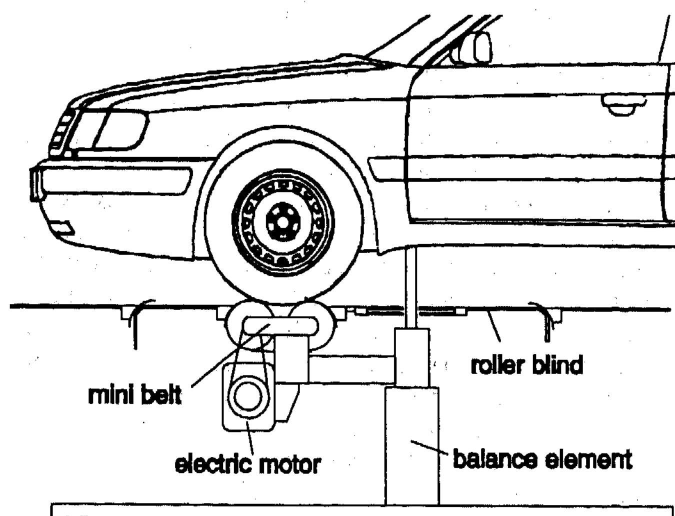


Fig. 7: Set-up for measurement with external balance (wheel drive integrated into balance plates)

this method has the fundamental advantage that rolling resistance cannot influence the results, since the wheel drive forces are absorbed internally by the balance frame. The integral aerodynamic drag value does not include the proportion for wheel fan moment, since this is also part of the wheel drive forces. The wheel drive forces must therefore be measured separately and then divided into a fan element and a rolling resistance element in the same way as for the moving belt measurement. This can be seen as a drawback, since an additional measuring device is necessary. On the other hand measurements with full wheel load should be possible without loosing accuracy for the aerodynamic forces (except for the fan moment).

TARE MEASUREMENTS WITH EXTERNAL BALANCE

In the case of measurements with an external balance, the problems of registering the fan moment are quite different in nature, on account of the difference in the testing set-up. The drag forces recorded during measurement (see *Fig. 8*) are in this case

- **1** possible weight-related forces
- **3** aerodynamic force acting on the body
- **4** aerodynamic force acting on the wheels

Rolling resistance **2** does not cause any problems here, since these are internal forces which are not measured. To this extent the tare measurement is also unproblematical, since it contains no foreign components.

However, once again fan moment **5** cannot be determined, so that the integral value for the measured drag is fundamentally falsified (too low by the fan moment value). In order to determine the fan moment, the wheel drive force at the wheel contact points must also be measured. Since the wheel drive units are installed on the weighed section of the balance, the drive force must be measured as an internal force in the balance structure by means of separate sensors. There are various ways to achieve this:

- with an additional load cell to measure the force in the x-direction on the balance plate (with low-friction mounting) with wheel drive
- the power needed to drive the wheel is determined from the motor reaction torque, so that the force at the wheel contact point can be calculated
- by means of strain gauges mounted directly at the axles of the wheel drive rollers, all forces at the wheel can be determined (drive force plus supporting force -> determination of wheel lift forces)

None of these three posibilities is available at present in an existing facility. Therefore we can not give any measured data for the vehicle under investigation here. In a future wind tunnel with the wheel drive force determined in the described way is once again the sum of the fan moment and the rolling resistance, and must be divided up into these two components by a suitable method, as in the case with the internal balance.

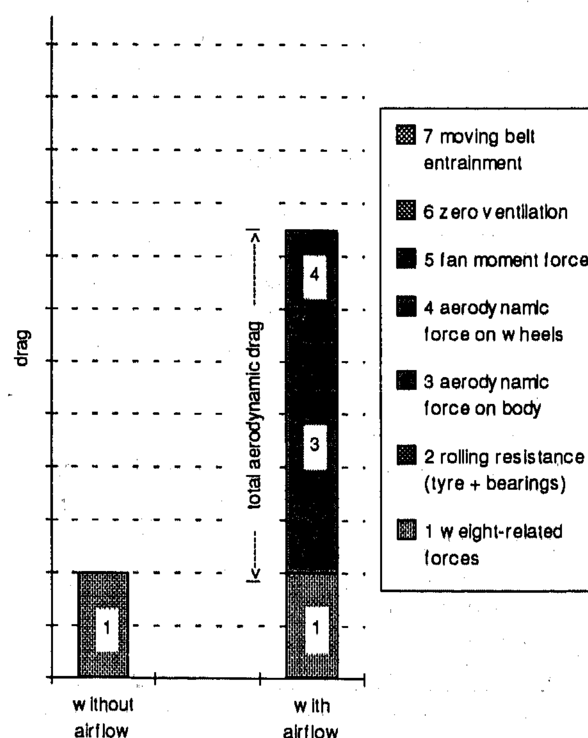


Fig. 8: Forces with the external-balance measuring set-up

The process of separating the rolling resistance encounters technical measurement problems at the full wheel load, which is desirable as a means of showing the correct patch. The question is to what extent simulation of the realistic wheel load is necessary for the measurement of the fan moment alone. If not, the fan moment could be determined in a "second shot" without wheel load.

MEASUREMENTS ON DIFFERENT WHEEL/TYRE COMBINATIONS ON THE VEHICLE

With the internal-balance experimental set-up already described, measurements were carried out on an Audi A3 with different wheel/tyre combinations. Measurements were taken both with rotating wheels and moving belt, and with stationary wheels on fixed ground for purposes of comparison. For the rotating-wheel measurements, two different forms of tare measurement were compared. One was the tare measurement at low belt speed as discussed before, the other the frequently used form of tare measurement with belt moving at $V_b = V_0$ (here 140 km/h). This supplies information on the influence of wheel rotation and of the type of tare measurement used.

The following parameters were investigated:

- Wheel design
- Tyre tread
- Cooling air

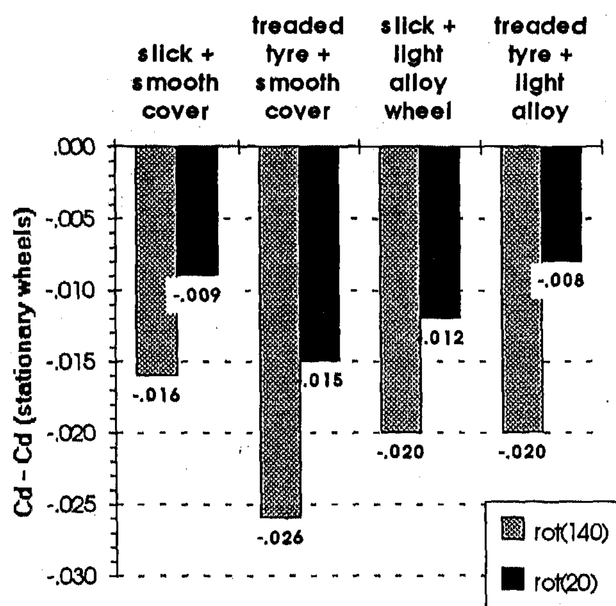


Fig.9a: Assessment of different wheel/tyre combinations evaluated by different measurement setup:
- stationary wheels and fixed ground (reference)
- rotating wheels, tare measurement at 20 km/h
- rotating wheels, tare measurement at 140 km/h

In each case the tyres were of size 195/65 R15 on rims of size 6 J x 15. The wheels investigated are shown in *Figs. 14.1* and *14.3* in the Annex.

The results (see *Fig. 9a*) indicate first of all that wheel rotation and the form of tare measurement affect the level of the measured aerodynamic drag coefficient: the absolute value c_D is always larger in case of stationary wheels compared to rotating wheels (it is the effect of the wheels and not of the moving belt, since moving belts increase drag). This effect has already been noted by various observers (see [1], [2], [6]). When measuring with a tare measurement at 20 km/h, however, the difference is distinctly smaller. Hence, with a tare measurement at 140 km/h, a considerably lower c_D is stated.

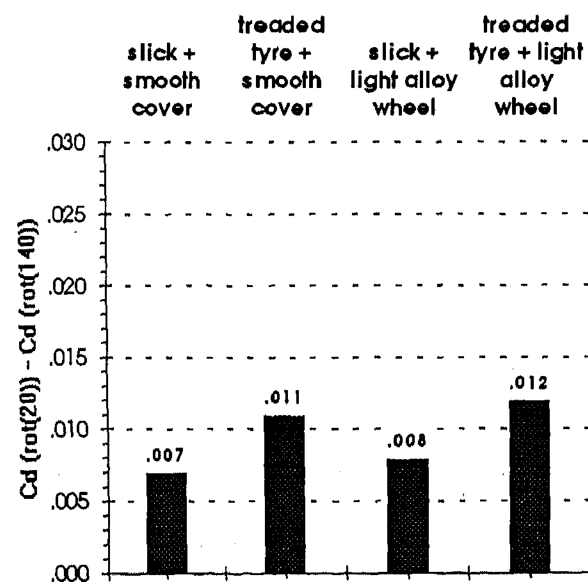


Fig. 9b: Difference in assessment of wheeld/tyre combinations usibg different tare measurements
- tare mesurement with 20 km/h
- tare measurement with 140 km/h

referred to a measurement with stationary ground and wheels

The nature of the tare measurement also reveals a significant difference in the Δc_D of different wheel/tyre combinations:

$$\Delta c_{D\ rot(20)\,/\,rot\,(140)} = \text{between } 0.007 \text{ and } 0.012$$

This difference, according to what has already been discussed, is quite exactly the off-set caused by the zero ventilation value of the wheels plus the boundary-layer effect of the moving belt. Accordingly, the difference is at its lowest when the wheel exhibits low zero ventilation ("slick" plus smooth wheel cover), with the opposite

applying when a treaded tyre and a light-alloy wheel are used.

Broken down according to the stated experimental parameters, the results are (see *Fig 9c*):

- **Wheel design:** measurements with rotating wheels illustrate the difference between the individual wheels more clearly than if the wheels are stationary. However, the wheel cannot be assessed in isolation but only in conjunction with the tyre:

Δc_D (alloy wheel/smooth cover) =

= + 0.014 (treaded tyre)

or = + 0.007 ("slick")

- **Tyre tread:** Measurements with rotating wheels yield the surprising result that the treadless tyre on a wheel with a smooth wheel cover performs worse than the equivalent tyre with tread:

Δc_D (treaded / "slick") (smooth cover) = - 0.005

This is not the case with the light alloy wheel:

Δc_D (treaded / "slick") (alloy wheel) = + 0.002

Here too, therefore, the tyre cannot be assessed without the wheel.

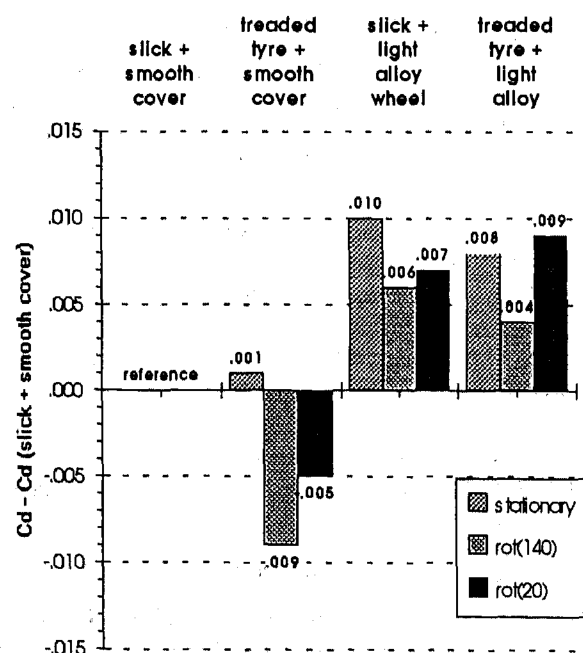


Fig. 9c: Assessment of different wheel/tyre combinations depending on the measurement technique and tare measurement; reference is the slick tyre on wheel with smooth cover

- **Cooling air:** It had been found by Mercker et al.[1,6], that cooling air is closely related to the aerodynamic drag at the wheels. In our case (see *Fig. 9d*) the cooling air flow does not have any significant effect on the assessment of the two wheel/tyre combinations:

Δc_D (light alloy wheel/smooth cover) =

= + 0.014 (cooling open)

= + 0.013 (cooling closed)

Nor does this result if the tare measurement is varied. This initially seems to contradict the measurements obtained by Mercker et al. [1,6], but we should have in mind that this result is certainly vehicle-specific and secondly only one integral value for the drag change at wheels and body was measured. Individual drag values could certainly be subject to some kind of influence. This problem is discussed in more detail by Pfadenhauer [9].

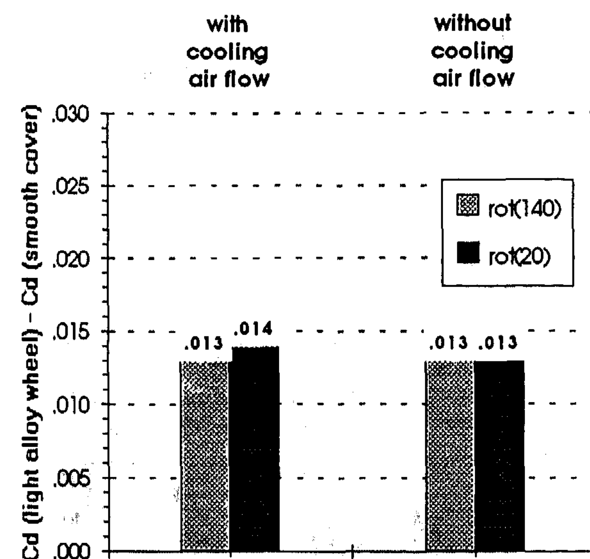


Fig. 9d: Assessment of different wheels with and without airflow. Measurement with rotating wheels and ground and different tare measurements

Taking these measurements as an example, we can see the extent to which the nature of the measurement influences the assessment of various wheels. Measurement with the wheels rotating represents none the less a definite improvement in the simulation of road conditions. With a change in tare measurement, a further correction of the same magnitude is obtained.

Another, possibly important aspect is the occurrence of deformation corresponding to tyre load at the contact point or "patch". When the transition from stationary to rotating wheel takes place, the simulation worsens in this respect. This will be examined in the following chapter.

INFLUENCE OF THE TYRE CONTACT PATCH ON DRAG VALUES (PRINCIPLE TEST ON A SINGLE WHEEL)

The simulation of a realistic tyre contact patch in wind tunnel tests with rotating wheels has been quite unsatisfactory so far. To simulate it in conjunction with rotating wheels caused problems of a technical nature. High friction losses and the associated thermal loads reduced the operating life of the moving belts when operating at full wheel load to such an extent that in many cases an inadequate simulation of the tyre contact patch was given preference instead, in order to maintain the progress of the test. In a fundamental test conducted in the FKFS wind tunnel in Stuttgart, it proved possible to undertake single-wheel tests with the wheel rotating under full load. The influence of the tyre contact patch on wheel drag was examined in relation to the incoming airflow and to wheel rotation.

To a better understanding of the flow conditions at the front and rear wheels of a complete vehicle first the flow field in a horizontal plane underneath the car is discussed (see Fig. 10). This will help us to apply the result of a single wheel test to the much more complicated situation at a real car.

The situation at the front wheels is dominated by the displacement flow from the nose of the vehicle and the wheel, which causes the airflow to strike the front wheel at an angle. The lower part is exposed freely to the flow without interference with wakes of other parts upstream. The local yaw angle was investigated on several vehicles, and is always approx. 15 degrees. No significant influence oft vehicle dimensions on the local yaw angle at the front wheels was detected.

At the rear wheels the flow arrives at a smaller angle than at the front. Flow losses from the front-wheel outflow and losses at the underfloor of the vehicle distinctly reduce the velocity at this point.

The complex flow patterns at the rear wheels seemed to be less suitable for investigations into the basics of wheel drag. However, giving priority to the front wheels should not be taken as implying that the rear wheels make less of a contribution to aerodynamic drag than the front wheels.

To drive the wheel, a miniature moving belt developed by FKFS was used (see J. Potthoff, R.-G. Fiedler [10]). This belt permits the simulation of wheel rotation at full wheel load. It is possible in this case to investigate the effects of tyre contact patch formation and the associated dynamic deformation of the rotating tyre on wheel drag.

The experiments were made with a standard light-alloy wheel with perforated wheel disc, size 6 J x 15, with a 195/65 R 15 tyre. The c_D values stated refer by way of an example to a vehicle with 2.0 m² frontal area.

The test setup for determining forces and moments consists of a vertical post and a cantilever arm. Since post and arm are part of the weighed system, the airflow forces acting on the mounting were determined by calibration measurement before the experiment was undertaken. For this measurement the wheel, as an interference body, was

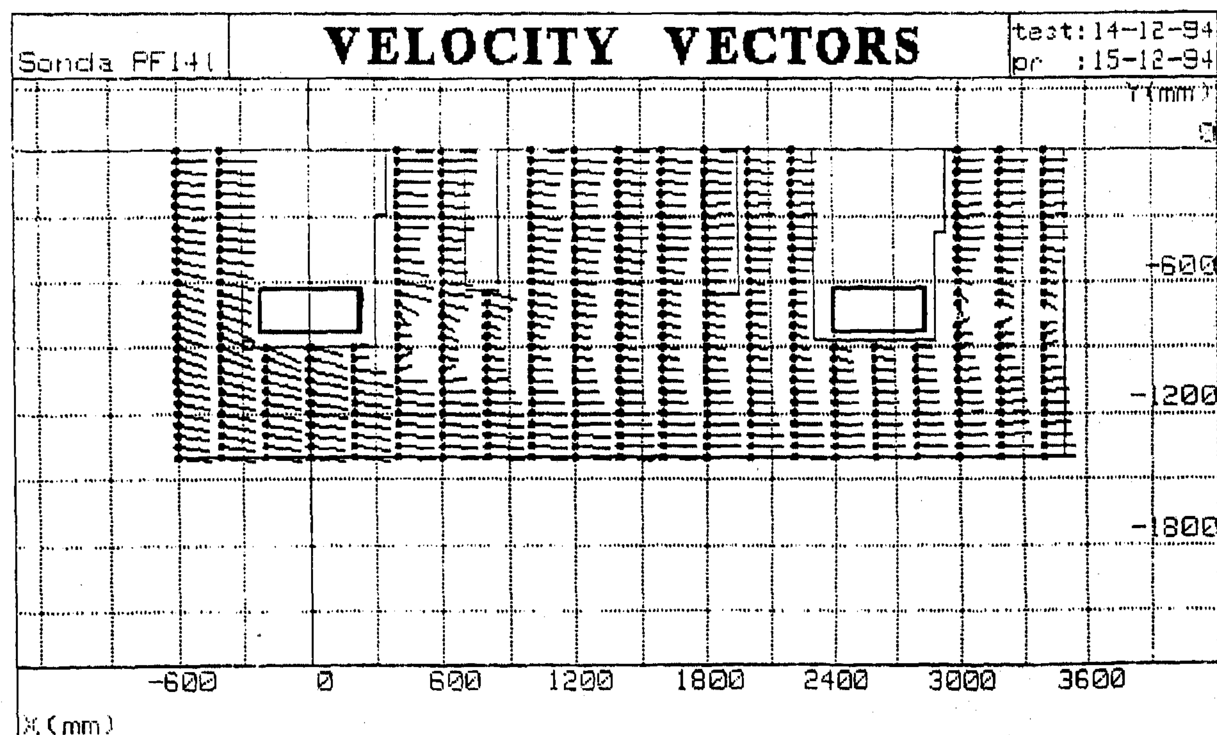


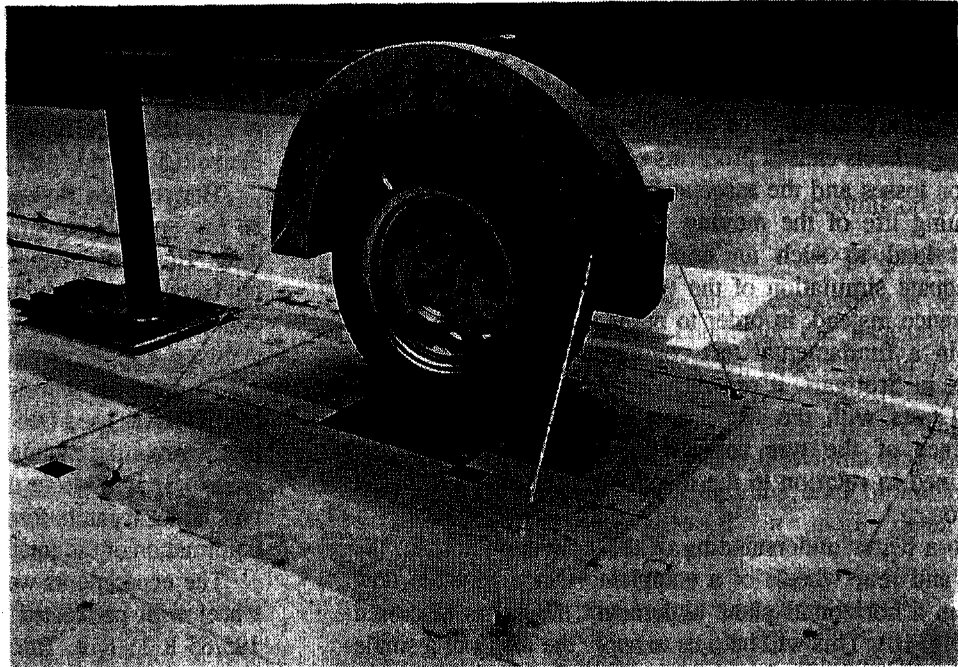
Fig. 10: Flow zone speed vectors on the underside of the vehicle, in a plane parallel with the road surface at height Z = 100 mm.

decoupled from the weighed section in force terms. Interference of the rotating wheel on the mounting was not determined separately, so that this force element is present in the evaluations for the rotating wheel.

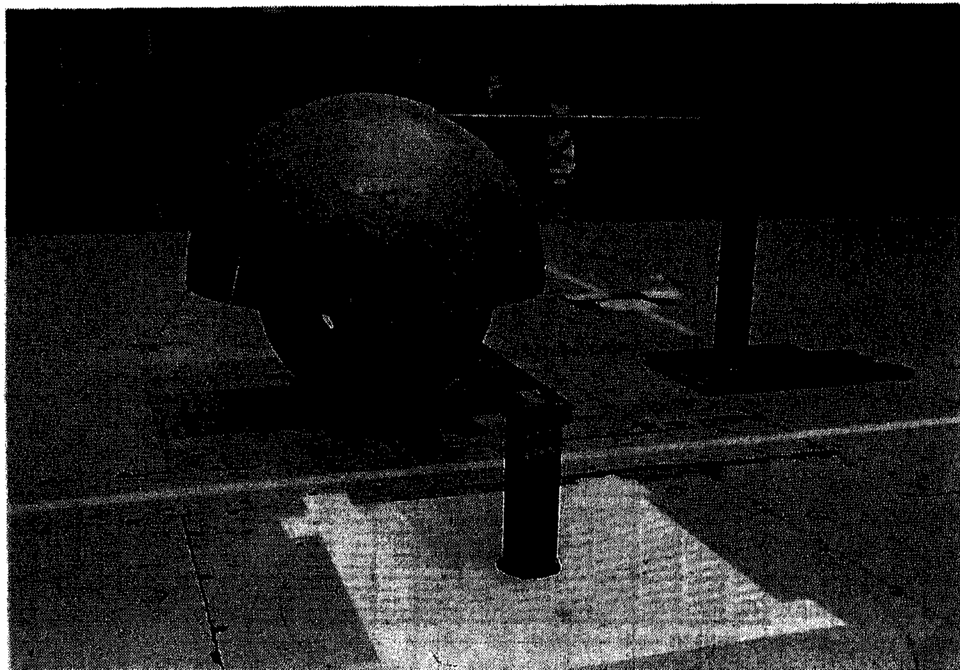
A wheel well simulator was developed to represent the flow conditions in the wheel well. It was a double-walled structure with its inner contour modelled on the wheel well of a standard production vehicle. The outer contour has the configuration of a quarter-sphere segment, in order to ensure almost constant incoming flow conditions when the inclined flow incidence angle is varied, regardless of the outer contour. The wheel well simulator is decoupled from the weighed section in terms of force transmission, and acts exclusively as an interference body.

The wheel load applied to create the correct tyre contact patch (preload and unspring mass) was 2,620 N. When the load is applied to the wheel, the moving belt is lowered by a certain amount between the two reversing rollers (see *Fig. 7*). The belt mainly yields only in the zone represented by the tyre's contact patch. However, with the belt lowered between the rollers and a load of this magnitude, the contact patch simulation is sufficient for an initial estimation.

Variation of the yaw angle is achieved by rotating the turntable carrying the complete experimental set-up. A positive flow incidence angle is one on to the inside of the wheel and describes the realistic flow conditions for the car.



Experimental set-up (view on open side of wheel well)



Experimental set-up (view on closed side of wheel well)

Fig. 11: Experimental set-up: principle tests on a single wheel at the FKFS.

At this time it was not possible to assess fan moment by the setup used, so that the values quoted for this test do not include the proportion of fan moment contributed by the moving belt . It is also assumed that the exact form taken by the tyre contact patch has no influence of fan moment or on the interference effect between wheel and body.

In order to present the results correctly to derive the influence of the tyre contact patch, it is first necessary to ge

briefly into some detail regarding the pattern of drag at the stationary and rotating wheel in relation to the yaw angle (as presented in *Fig 12*).

Without yaw (β=0°) the drag coefficient is reduced by approx. 60 % when the wheel is rotated. As the yaw angle is increased up to β=25°, drag increases at both the stationary and the rotating wheel. However, the increase is distinctly steeper at the rotating wheel, so with the yaw angle exceeding 10° the rotating wheel exhibits higher drag than the stationary wheel. The cause of this increase in drag at the rotating wheel is more intensive flow seperation at the side of the tyre. These results are discussed in detail by M. Pfadenhauer [9].

Since it has until now not been possible to show the tyre contact patch adequately, assessment of wheel drag has suffered from various uncertainties. Use of the miniature moving belt, however, has permitted investigations with realistic tyre deformation (both contact-patch and dynamic deformation) when the wheel is rotating.

At low yaw angles, the tyre contact patch results in an increase in drag. If wheel rotation is simulated, this influence is reduced by an almost constant offset of Δc_D < 0.001. As the yaw angle increases, the tyre contact patch's drag-increasing effect is lost. Neutral drag occurs, depending on wheel rotation, at approx. 12 degrees (rotating) or 15 degrees (stationary). At angles greater than 25 degrees, which however are not relevant to vehicles, the presence of the tyre contact patch always has the effect of causing a slight reduction in drag.

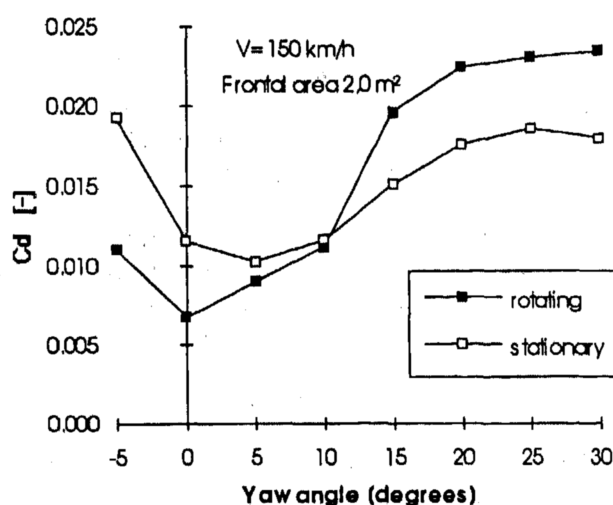


Fig. 12: Relationship between drag at the stationary and rotating wheel and the yaw angle. (The drag coefficients referred to a frontal area of 2.0 m² → complete vehicle).

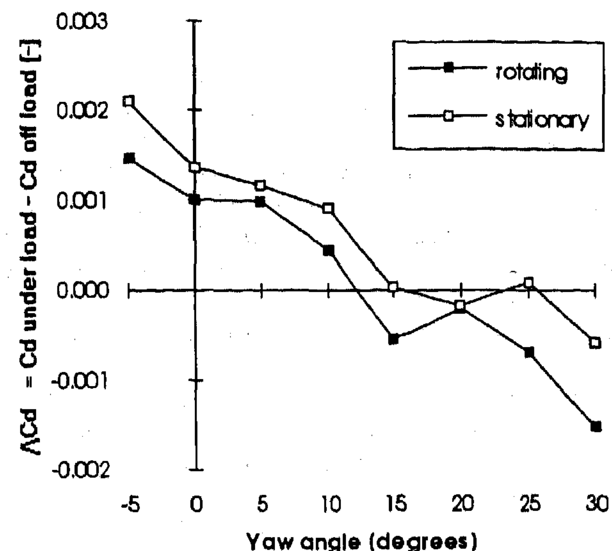


Fig. 13: Influence of the tyre contact patch for the stationary and the rotating wheel in relation to the yaw angle.

In order to estimate the influence on the vehicle's drag coefficient caused by the tyre contact patch, we need to know the flow angles at the front and rear wheels. From *Fig. 13* we can deduce the influence of one wheel on total drag.

At the front wheels, with an yaw angle of 15 degrees, the change in drag due to the contact patches of both front wheels is Δc_D < +0.001 stationary and Δc_D = -0.001 rotating.

The tyre patches at the two rear wheels change the vehicle's overall drag at an yaw angle of 3 degrees by Δc_D = +0.003 for stationary wheels and Δc_D = +0.002 for rotating wheels.

The influence of the tyre contact patch on the drag coefficient of the complete vehicle can therefore be estimated as

Δc_D = +0.003 for stationary wheels and

Δc_D = +0.001 for rotating wheels.

From these results we can conclude that realisation of the tyre contact patch has only slight effects on the vehicle's total drag coefficient, and that this influence is weakened still further by the rotating wheels. In view of this, simulation of the tyre contact patch, which involves considerable technical effort with rotating wheels in particular, can be regarded of minor importance. Therefore it can be neglected when determining a vehicle's aerodynamic drag.

SUMMARY

An experiment devoted to establishing fundamental values was able to demonstrate that the wheels and tyres can account for 25 % of a vehicle's total aerodynamic drag, with the wheel wells causing much less than the wheels. If this proportion is to be correctly measured with the technical facilities at our disposal, it is in our opinion necessary to simulate rotation of the wheels. In principle, measurements can be taken either with an internal balance and the wheels driven by a moving belt, or with an external balance and special wheel drive units.

In the first of these arrangements, the force generated by the wheels' fan moment is included in the total drag value. Since this force is aerodynamic in character and also increases approximately in accordance with the square of the speed, we feel that it is correct in physical terms to convert it into a Δc_D and add it to the vehicle's aerodynamic driving resistance. If measurements are taken with an external balance and rotating wheels, the aerodynamic drag value will accordingly be too small and will need to be completed by the amount represented by the fan moment. Until now, we have been unable to determine the fan moment in isolation. However, it is suspected to exert an influence on the drag of the same magnitude as the zero ventilation which we have measured.

Depending on whether internal- or external-balance measurement is undertaken, either the zero ventilation or the fan moment must be separated from the wheels' rolling drag. We have described a practical means of doing this when measuring with an internal balance, with the rolling resistance eliminated by means of a tare measurement at low belt speed. For the external-balance measuring method this method could also be applied as appropriate, but the necessary technical facilities must first be made available.

With the aid of examples we were able to show that the measuring method has a decisive influence on the assessment of typical types of wheels. In addition, these measurements confirmed the importance of recording the aerodynamic drag of the wheels and tyres correctly.

A principle test with a single wheel on a moving belt demonstrated that in the case of rotating-wheel measurements the so far unavoidable simplification by reducing the wheel load has only a slight effect on the measured results. Accordingly, aerodynamic drag measurement with rotating wheels under reduced load is a much better compromise than testing with stationary wheels.

BIBLIOGRAPHY:

[1] E. Mercker, N. Breuer, H. Berneburg, H.J. Emmelmann: On the Aerodynamic Interference Due to the Rolling Wheels of Passenger Cars. SAE Technical Paper 910311, 1991

[2] S.F. Hoerner: Fluid-Dynamic Drag. 1. Aufl., Eigenverlag 1951

[3] E. Mercker, H.W. Knape: Ground Simulation with Moving Belt and Tangential Blowing for Full-Scale Automotive Testeing in a Wind Tunnel. SAE Technical Paper 890367, 1989

[4] A. Cogotti: Ground Effect Simulation for Full-Scale Cars in the Pininfarina Wind Tunnel. SAE Technical Paper 950996, 1995

[5] G. Wickern: The Effect of Moving Ground on the Aerodynamic Drag of a Production Car. SATA Conference, South Bend, 1991

[6] E. Mercker, H. Soja, J. Wiedemann: Experimental Investigation on the Influence of Various Ground Simulation Techniques on a Passenger Car. Conference on Vehicle Aerodynamics, Proceedings Roy. Aer. Soc., Loughborough University of Technology, 1994

[6] H. Schlichting: Boundary Layer Theory, McGraw-Hill, 1979

[8] M. Deubert: Bestimmung des Luft- und Lüfterwiderstandes sich drehender PKW-Räder als Teil des Fahrwiderstandes. Diplomarbeit FH Köln, Fachbereich Fahrzeugtechnik, 1989

[9] M. Pfadenhauer: Konzepte zur Verringerung des Luftwiderstandsbeiwertes von Personenkraftwagen unter Berücksichtigung der Wechselwirkungen zwischen Fahrzeug und Fahrbahn, sowie der Raddrehung. Diplomarbeit, Bericht FLM 95/33, Lehrstuhl für Fluidmechanik, TU München

[10] J. Potthoff, R.-G. Fiedler: Simulation der Raddrehung bei aerodynamicen Untersuchungen an Kraftfahrzeugen im Windkanal. Tagung Aerodynamik des Kraftfahrzeugs, Haus der Technik, Essen, 1995

NOMENCLATURE

A	reference (frontal) area	[m²]
a,c	coefficients for approximation	
c_D	drag coefficient	[-]
c_M	moment coefficient	[-]
F_x	drag force	[N]
F_{Roll}	rolling resistance	[N]
M	drag moment	[Nm]
n	exponent for scale effect	
R	radius of disk/wheel	[m]
Re	Reynolds number	[-]
v	vehicle speed	[m/s]
v_0	free stream velocity	[m/s]
v_b	belt speed	[m/s]
ν	fluid viscosity	[m²/s]
ρ	fluid density	[kg/m²]
ω	circumferential velocity	[Hz]

Indices

stat. measured with stationary wheels and fixed ground

rot(20) measured with rotating wheels and moving ground, tare measurement with 20 km/h

rot(140) measured with rotating wheels and moving ground, tare measurement with 140 km/h

ANNEX

1.Evaluation of drag moment of a rotating disk in still air

The proportion of the skin friction for a rotating wheel in still air can be approximated by a formula given by Schlichting [7] for rotating disks in turbulent flow (i.e. the case of a turbulent boundary layer):

$$2M = 0.073 \cdot \rho\omega^2 R^5 \left(\nu/_{\omega} R^2\right)^{\frac{1}{5}}$$

where M is the moment for one side of the disk, R is the radius of the disk, ω the circular frequency and ν and ρ are fluid properties.

This can be reformulated nondimensional:

$$c_M = 0.146 \cdot Re^{-\frac{1}{5}}$$

with a Reynoldsnumber based on disk dimensions. Recast in terms of Δc_D this gives for a single wheel:

$$\Delta c_D = 0.146 \cdot Re^{-\frac{1}{5}} \cdot \frac{R^2}{A_M}$$

Calculating the Δc_D for four wheels for the case investigated here yields to:

$$\Delta c_D = 0.002$$

Thus the the major part of the fan moment is not produced by pure skin friction.

2. Additional figures

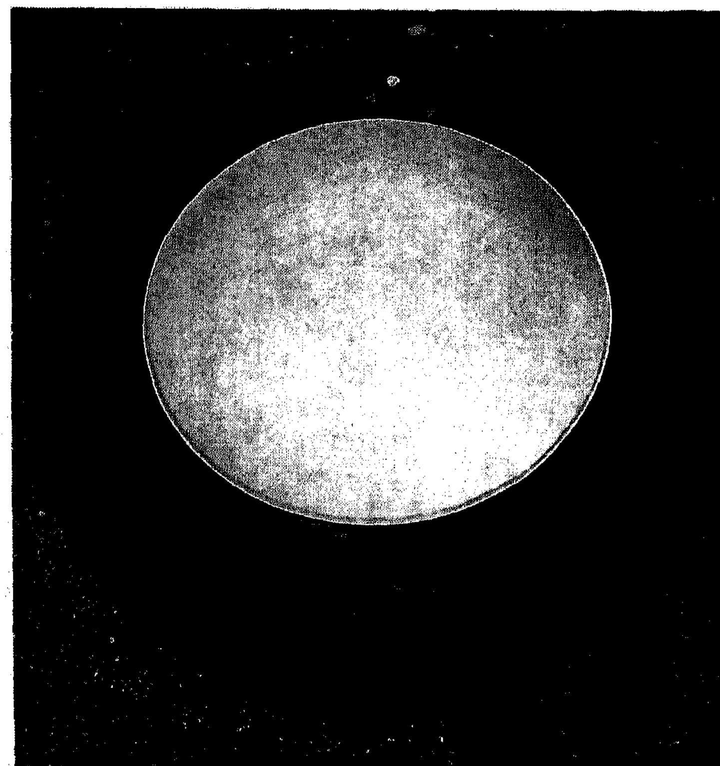


Fig. 14.1: Wheel with full-diameter smooth cover



Fig. 14.2: Wheel with regular series-production cover

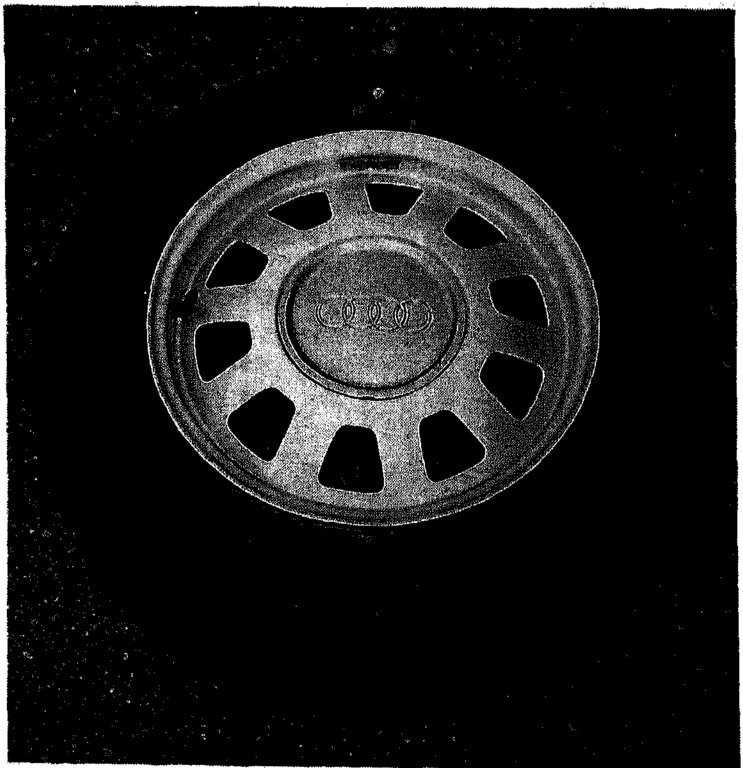


Fig. 14.3: Light alloy wheel

970135

Assessing the Effects of Shear and Turbulence During the Dynamic Testing of the Crosswind Sensitivity of Road Vehicles

Roger Macklin and Kevin Garry
Cranfield Univ.

Jeff Howell
Rover Group

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

With increasing speeds and the anticipated reduction in weight of modern cars, the issue of crosswind sensitivity is becoming increasingly important. In a previous paper by the same authors, the normal method of testing such aerodynamic characteristics at model scale, using static models at yaw to the freestream, was compared with dynamic testing, in which the model is propelled across a 'gust' simulated by a wind tunnel. A direct comparison using a similar gust profile for both static and dynamic tests was made with the conclusion that the simple static test technique was underestimating the true transient loads.

Further tests have been carried out, on a generic squareback (or estate) model, during which the effect of varying both the vertical velocity profile and the turbulence intensity within the gust was considered. Although these gust characteristics were not necessarily representative of an atmospheric boundary layer they give a good indication of the differences that might be expected between tests done using a static model at yaw and the more accurate dynamic simulation.

The data show a considerable variation in aerodynamic load on the model as the vertical velocity profile was altered but less variation with turbulence intensity. An attempt is made to determine, a simple relationship between a reference velocity, representative of the shape of the gust profile and the resultant forces and moments.

NOMENCLATURE

x,y,z	body axis system
u'	turbulence intensity
h_m	model height, 0.189 m
ψ	yaw angle
U_R	resolved velocity
m	turbulence parameter power index, see Bearman[3]
Δ_{skew}	correlation factor, see Macklin[7]
C_y	Sideforce coefficient
C_L	Lift coefficient (subscript F & R, front and rear axle)
${}^xL_u$	longitudinal turbulence length scale
M_z	aerodynamic moment (subscript z & x, yaw and roll)
U_0	freestream velocity
l	model length, 0.6 m

INTRODUCTION

This paper will begin by summarising previous work on the aerodynamic characteristics of vehicles in crosswinds conducted using the crosswind track facility at Cranfield University. It will then go on to detail work done to assess the impact of the nature of the crosswind gust by introducing both turbulence and changes in the vertical velocity profile.

The methodology behind scale model crosswind testing was detailed in an earlier paper by Macklin et al[1]. Here the advantages and disadvantages of static-static, static-dynamic, and dynamic-static testing techniques were considered, the later being adopted in the present investigation as being the most practical way of investigating changes to the gust characteristics.

In the previous paper[1] the crosswind track technique was evaluated by comparing data from three generic passenger vehicle configurations with data from the same models tested under conventional static-static conditions over a fixed ground plane and range of yaw angles. During

these tests the vertical velocity profile was deliberately kept uniform to ensure that the effective profile seen by the moving model did not contain any 'skew'.

The results presented by Macklin et al[1] show that at yaw angles below 15 degrees, the conventional static-static tests give a conservative measure of the model forces and moments when compared to the dynamic-static data from the crosswind track. At yaw angles greater than 15 degrees (measurements were made up to 45 degrees) the magnitude of the dynamic coefficients was generally greater than the corresponding static data. This was particularly evident in peak values of yawing moment.

It was recognised during this programme that dynamic tests with a uniform gust profile (ie without skew) would be the 'worst case' scenario and that the introduction of skew would most likely give lower aerodynamic coefficients. The degree to which the gust velocity profile is modified is therefore important and the current paper analyses the impact of five different profiles on the transient aerodynamic characteristics of a generic squareback vehicle configuration. In addition the issue of which velocity, or dynamic pressure, should be used to non-dimensionalise coefficients, when considering a non-planer incident flowfield, is addressed.

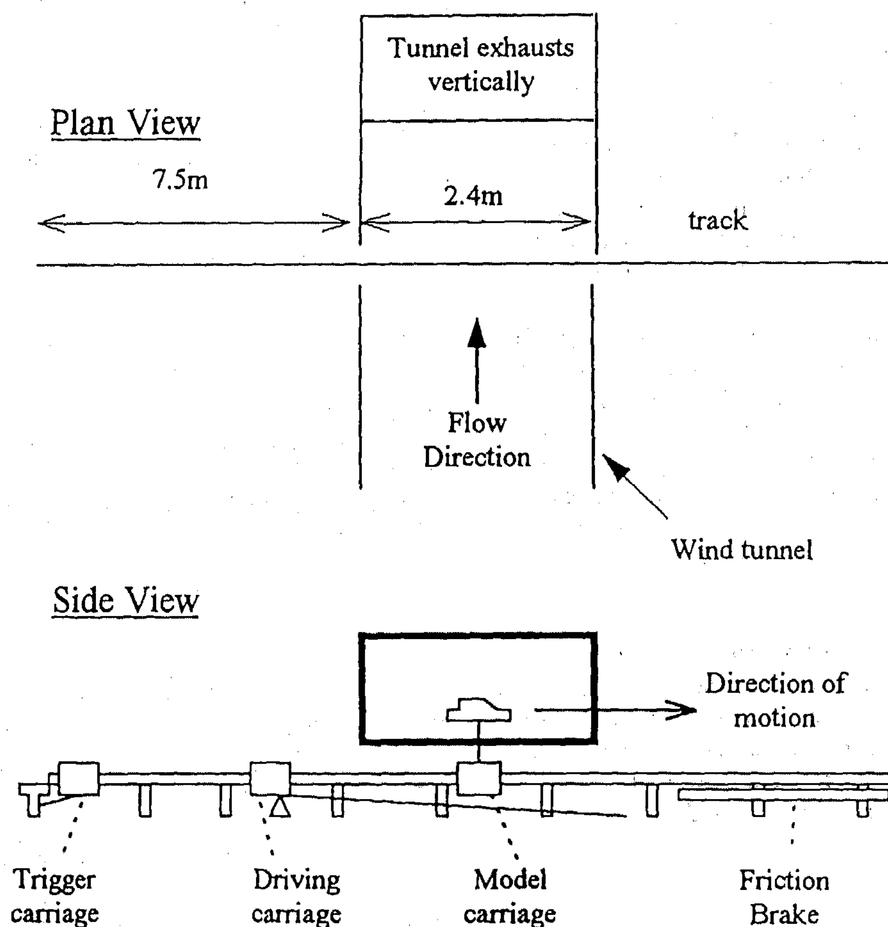


Figure 1
Schematic layout of the crosswind track and 2.4m x 1.2m boundary layer wind tunnel.

EXPERIMENTAL FACILITY

The crosswind track runs perpendicular to the 2.4m x 1.2m boundary layer wind tunnel and is shown schematically in Figure 1. An outline of the essential features of the facility is given here for convenience, while a more detailed description of the facility is given by Macklin et al[1]. The model under test (typically a 1/5th scale passenger car) is propelled across the wind tunnel working section, together with it's internal 5 component strain gauge balance, on a carriage, having been accelerated by a bungee driven catapult. The model speed is regulated to within 0.5 m/s by the release position of the driven carriage and is typically 12-18 m/s. The model carriage also carries a data acquisition system and power supply which, after being activated by track mounted LED's, records balance data and model position which are down loaded to a PC at the end of each run for analysis.

Aerodynamic coefficients, with the exception of drag are evaluated from a balance calibration matrix. The signals are filtered to remove track induced noise and a 'wind-off value' signal is subtracted such that the resulting data relates only to the influence of the crosswind gust.

The necessarily lightweight model is intended to represent an estate configuration (or squareback) and is shown in Figure 2 together with the force/moment sign convention and reference point. Table 1, reproduced from ref.[1] shows the standard deviations of the coefficients at zero and 17 degrees yaw for the squareback model over 9 runs.

Gust Characteristics

The difficulty of changing the vertical velocity profile of the gust or wind tunnel flow, without changing the turbulence characteristics was recognised. A series of tests were therefore carried out using gusts of varying turbulence intensity and scale but with a uniform velocity profile. The turbulence characteristics of these gusts, generated by introducing rectangular grids at the upstream end of the wind tunnel working section, are shown in Figure 3. A total of four turbulence levels were evaluated corresponding to freestream turbulence intensities nominally in the range 1% to 3.5%. Turbulence intensity was seen to increase considerably at heights below that of the model mid-height. Turbulence length scale, evaluated on the basis of the auto correlation function, is given in Figure 3.

THE INFLUENCE OF TURBULENCE

In his review of the effects of turbulence on bluff bodies, Bearman[2] identified three basic mechanisms: (i) accelerated transition to turbulent flow within shear layers (ii) enhanced mixing and entrainment and (iii) distortion of freestream turbulence itself by the mean flowfield. The overall effect of changes in turbulence were seen to be the result of more than one of these three basic mechanisms and each is effected differently by the two principal characteristics of the turbulence - intensity and scale.

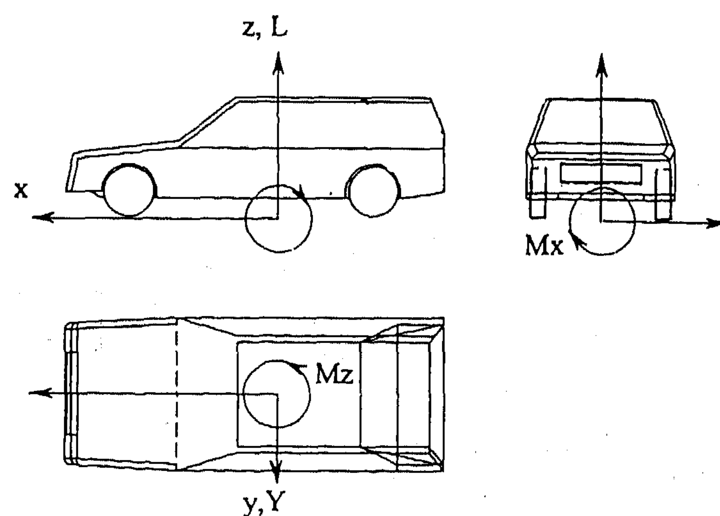


Figure 2
Schematic diagram of the 'squareback' model together with the axis and force and moment sign convention.

In an attempt to define a single turbulence parameter, Bearman[3] used a 'power index' (*m*) which has subsequently been used by a number of researchers

$$\left(\frac{u'}{U}\right)\left(\frac{L}{D}\right)^{m}$$

The significance of the ratio of the turbulence length scale (L) to some characteristic dimension of the flowfield (D) is reflected in the fact that the power index (*m*) varies according to the problem being considered.

The data showing the influence of turbulence on ground vehicles, and particularly on trains and HGV's, suggests that there are two main areas of interest. The first, usually referred to as 'micro' effects relate to the mechanisms discussed in the previous paragraph, and influence the position of separation and reattachment lines, together with the size and structure of vortices in the wake. The second 'macro' effect relates to the concept of 'aerodynamic admittance' where variations in the forces and moments are directly related to the variation in the instantaneous velocity.

Whilst micro turbulence effects have been shown by Baker[4] to be significant for trains, particularly to the C_y vs ψ and C_L vs ψ characteristics, it is recognised that the relative length scale of turbulence is of the same order of magnitude of the vehicle's wake. This will not be the case for a typical passenger car and indeed Bearman[2] shows that there is very little effect of changes in turbulence, either scale or intensity, on the base drag of a series of car body shapes.

Influence of turbulence in crosswind gust

Dynamic force and moment coefficients were recorded for the squareback vehicle configuration under four different turbulence profiles without skew. No attempt was made to correctly scale atmospheric turbulence.

The response of the vehicle is given in figure (4) and, with the exception of the data set corresponding to the turbulence generated by the horizontal bars, there appears to be very little effect of turbulence within the range considered.

The possibility of a turbulence parameter correlating the data was explored by adopting the form used by Bearman[3] - with the power index (*m*) taken to be unity - and evaluating the mean turbulence intensity of the profile over four different heights:

$z = 0$ to $1.5\ h_m$
$z = 0$ to h_m
$z = 0$ to $0.64\ h_m$
$z = 0.64\ h_m$ to $1.5\ h_m$

($z = 0.64\ h_m$ was considered because it corresponds to the approximate height of the centroid of the model side area. $z = 1.5\ h_m$ was included to allow for the influence of the flow above the vehicle on boundary layer growth).

Examples of force and moment coefficients at various yaw angles are given for two of the above turbulence parameters in figure 5. These data appear to support the initial conclusion that for this model and range of turbulence, there is little or no effect of turbulence or the dynamic force and moment data.

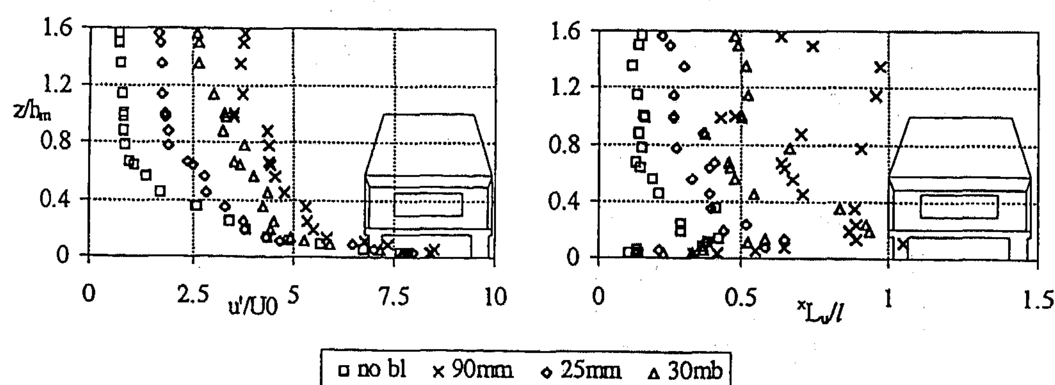


Figure 3
Profiles of turbulence intensity and non-dimensionalised longitudinal length scale for the four gust cases considered

THE INFLUENCE OF VELOCITY SKEW

A number of researchers have considered the effect of skew in crosswind tests on trains and HGV's. There seems to be little if any data relating to passenger cars. The impact of wind fences on the crosswind characteristics of HGV's has been investigated by Coleman[5] and as would be expected, due to the shielding effect, sideforce and lift coefficients were seen to fall compared to the static-static case. Baker[4] produced data from dynamic-static tests on train models which show inconsistencies depending on the force or moment coefficient being considered as a result of simulating either level surroundings or an embankment.

The five different vertical velocity profiles generated for this investigation are shown in Figure 6. These profiles were generated by placing roughness elements ('gravel','egg' and 'lego') on the floor of the wind tunnel upstream of the crosswind track and as a result the turbulence intensity and longitudinal length scale profile in the resultant gust will also be changed. No attempt was made to simulate a specific atmospheric boundary layer. In fact the turbulence profiles associated with the ground boards - in particular the 'egg' and 'lego' boards - show higher values of $u'(z)/U_\infty$ than created for the turbulence tests. (This creates a problem in so far as any conclusions arrived at on the basis of the turbulence data may not necessarily be extrapolated to predict results from the skew tests. As a consequence variations in the coefficients may not be entirely attributable to the change in the skew profile).

The effect of the five different velocity profiles on the aerodynamic characteristics of the squareback model in the yaw angle range 0 to 50 degrees is shown in Figure 7. The influence, compared to both the zero skew and static-static data appears to be greatest for front and rear axle lift and yawing moment. Sideforce and rolling moment do not seem to be influenced to the same extent and this is consistent with the results of Baker[4] for sideforce, but not for lift.

When considering the data from tests of this type the issue of which velocity, or dynamic pressure, to use to non-dimensionalise the coefficients becomes important. Clearly the use of freestream conditions in a boundary layer that may be many times the height of the vehicle does not seem appropriate. Previous work on buildings by Wacker[6] considered the use of a mean velocity evaluated from an integral of the profile over the height of the building, this was seen to significantly reduce the difference between aerodynamic data for different profiles.

As in the case of the mean turbulence characteristics, a number of different height ranges were chosen for the current study to evaluate an effective reference velocity:

(i) $0 < z < 1.5\ h_m$, takes into account the flow on the roof

(ii) $0 < z < 0.6\ h_m$, concentrates on flow over the bonnet and therefore influences front axle lift

(iii) $0.6\ h_m < z < 1.5\ h_m$, concentrates on flow over the backlight area and therefore influences rear axle lift.

In addition, single point reference velocities taken at:

(iv) $z = h_m$

(v) $z = 0.6\ h_m$

and

(vi) $z = \delta^*$ for convenience when working with full scale atmospheric boundary layers.

It should be remembered that changing the reference velocity in dynamic tests changes both the yaw angle (ψ) and the resolved velocity (U_R), while in the static tests it only effects (U_R).

If the alternative non-dimensional velocity is known in terms of U_∞ we can use a factor, B, such that

$$U_B = \frac{U_{ref}}{U_\infty}.U_\infty = B.U_\infty$$

the adjusted yaw angle is then given

$$\psi_B = \tan^{-1}(B.\tan\psi)$$

and the dynamic coefficients should be adjusted

$$C_{F_B} = C_F\left[\frac{1+\tan^2\psi}{1+B^2\tan^2\psi}\right]$$

while the static data is simply

$$C_{F_B} = C_F.B$$

An example of the impact of changing the reference velocity is given in Figure 8 for the squareback model. In this case aerodynamic data normalised using both the freestream dynamic pressure and that obtained from an integral over the height of the model, ie $0 < z < h_m$. It is obvious from Figure 8 that any improvement in collapse of the data will be difficult to quantify for the various reference velocities to be considered and consequently a quantitative measure, Δ_{skew}, was devised.

Δ_{skew} is a measure of the variation between datasets as a percentage of the unaltered data based on the standard deviation of the datasets. As a result

i) if adjustment of the non-dimensionalising velocity improves the data collapse the Δ_{skew} will be +ve. If all the data collapses onto one line, $\Delta_{skew} = 1$.

ii) if adjustment of the non-dimensionalising velocity does not improve the data collapse the Δ_{skew} will be -ve

The influence, in terms of Δ_{skew}, of each of the seven different non-dimensionalising velocities or aerodynamic coefficients, is given in Figure 9. It is seen that no one alternative reference velocity universally improves the collapse of the data. This is obviously a disappointing result, but it may well be that the most 'ideal' non-dimensionalising velocity was not included in those evaluated, since clearly a freestream value is not ideal. However, it is not possible from this data set to predict what the best value would be for passenger cars.

CONCLUSIONS

Dynamic tests using a scale model traversing a wind tunnel on a crosswind track are recognised as being an effective means of investigating the influence of both freestream turbulence and skew on the aerodynamic characteristics of passenger cars. The results from a series of tests on a generic squareback scale model vehicle suggest that, within the range of turbulence intensity and scale considered there was little effect of turbulence on the vehicle's aerodynamic characteristics. The possibility of a single turbulence parameter correlating the data was explored but proved inconclusive. Further tests in which the mean vertical velocity profile of the gust was varied to introduce 'skew' into the effective velocity profile confirmed that the 'non skew' profiles give conservative results in terms of crosswind stability assessment, particularly at yaw angles below 15 degrees. Analysis of several alternative reference velocities did not reveal a suitable position or integral that gave a universal collapse of the data. The vehicles lift characteristics were better correlated with a reference velocity taken as the mean velocity over the height of the vehicle.

REFERENCES

1. A R Macklin, K P Garry, J P Howell *Comparing Static and Dynamic Testing Techniques for the Crosswind Sensitivity of Road Vehicles* SAE SP-1145/960674 (1996)

2. P W Bearman, T Morel *Effect of Freestream Turbulence on the Flow around Bluff Bodies.* Prog Aero, Sci. Vol. 20 pp. 97-123 (1983)

3. P W Bearman *An Investigation of the Forces on Flat Plates in Turbulent Flow.* NPL Aero. 1296 (1969)

4. C J Baker *Ground Vehicles in High Cross Winds Part 2: Unsteady Aerodynamic Forces.* Jn of Fluids and Structures Vol 5. pp. 90-111 (1991)

5. S A Coleman *The Aerodynamics of Ground Vehicles in Cross Winds* PhD Thesis University of Nottingham, UK (1990)

6. J Wacker, R Friedrich, E J Plate, F Darsch *Drag and Lift on Rectangular Bluff Bodies Immersed in Deep Boundary Layer Flow over Rough Surfaces* Int. Colloquium on Bluff Bodies and its Applications, Kyoto Japan pp.113-123 (1988)

7. A R Macklin *Measuring the effects of crosswinds on cars.* PhD Thesis, Cranfield University, UK (1996)

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of both The Engineering and Physical Sciences Research Council under agreement 93300043 and Rover Group plc.

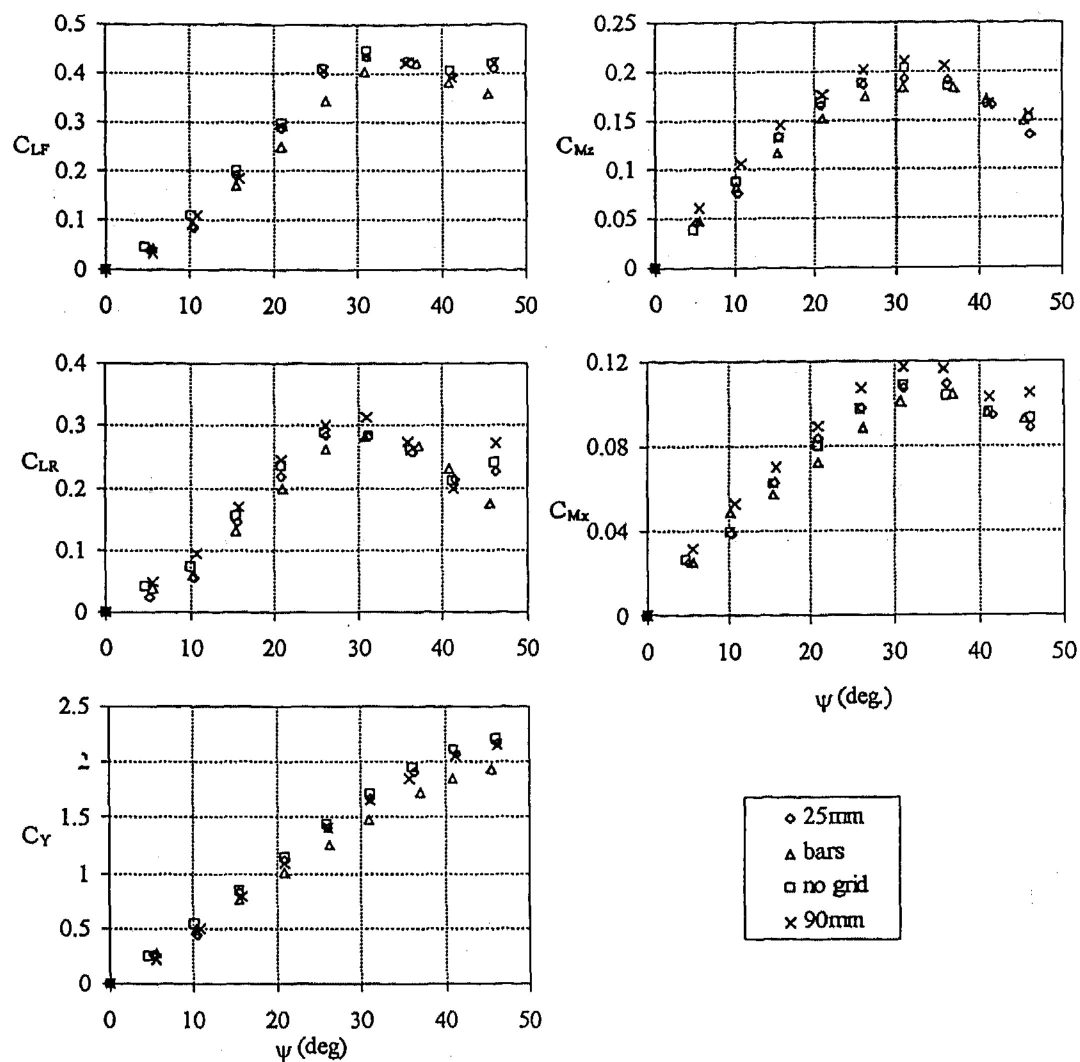


Figure 4
Force and moment coefficients versus yaw angle for the squareback model configuration under various turbulent gust conditions

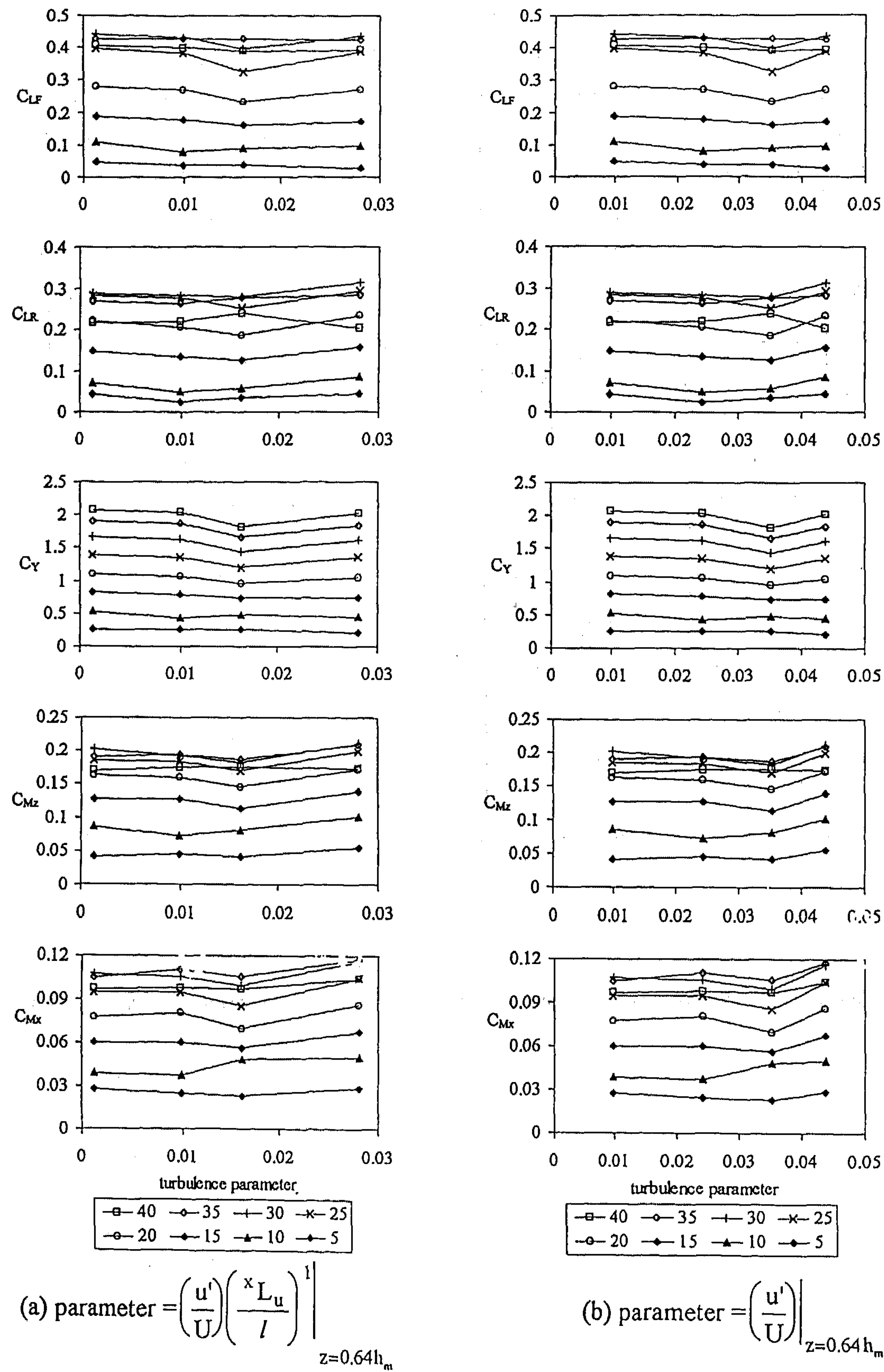


Figure 5
Force and moment coefficients plotted against two different turbulence parameters for various yaw angles

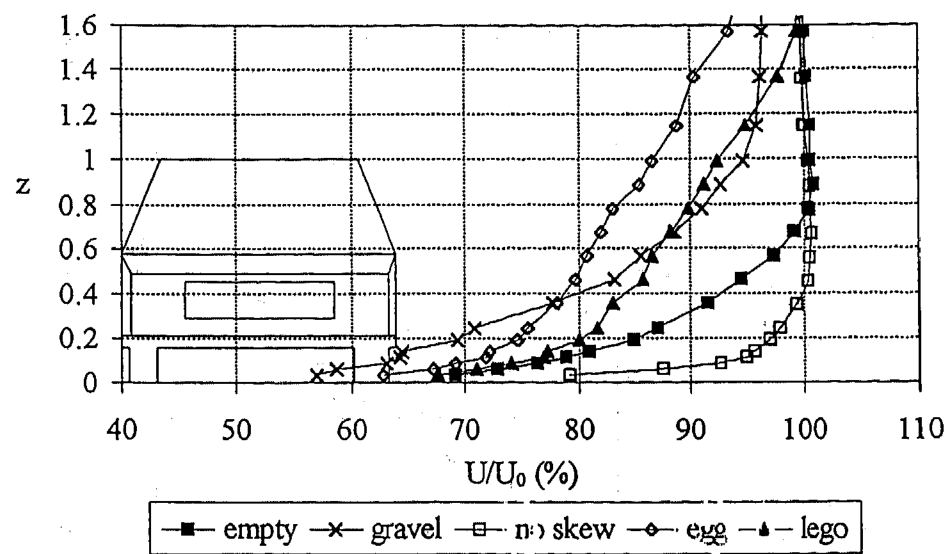


Figure 6
Vertical non-dimensional velocity profiles for the five skew gust cases considered

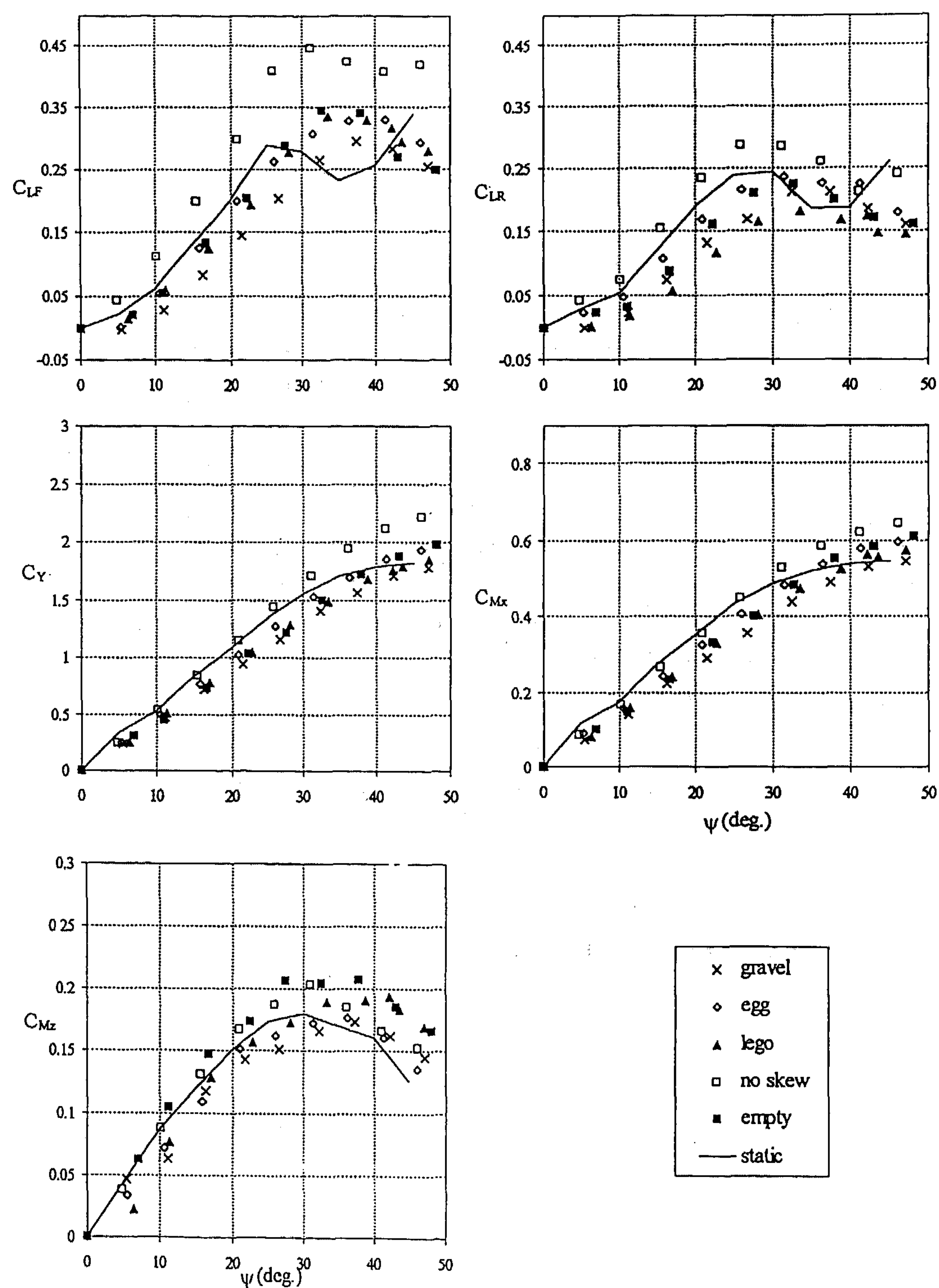


Figure 7
The effect of different skew profiles on force and moment coefficients for the squareback model configuration compared to the static data, ref[1]

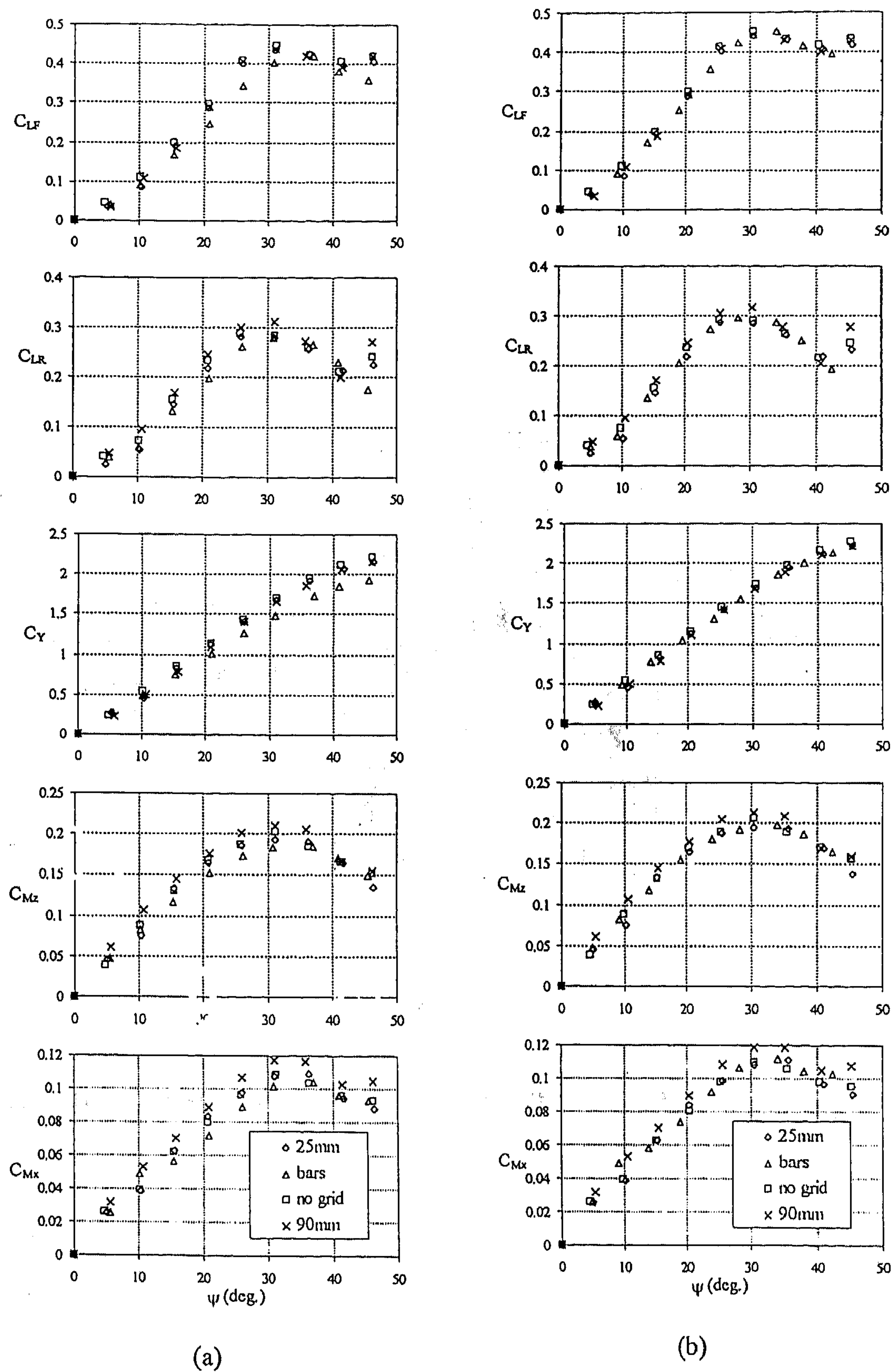


Figure 8
Comparison of force and moment data for different skew profiles non-dimensionalised by (a) freestream dynamic pressure and (b) dynamic pressure integrated over the range $0<z<h_m$

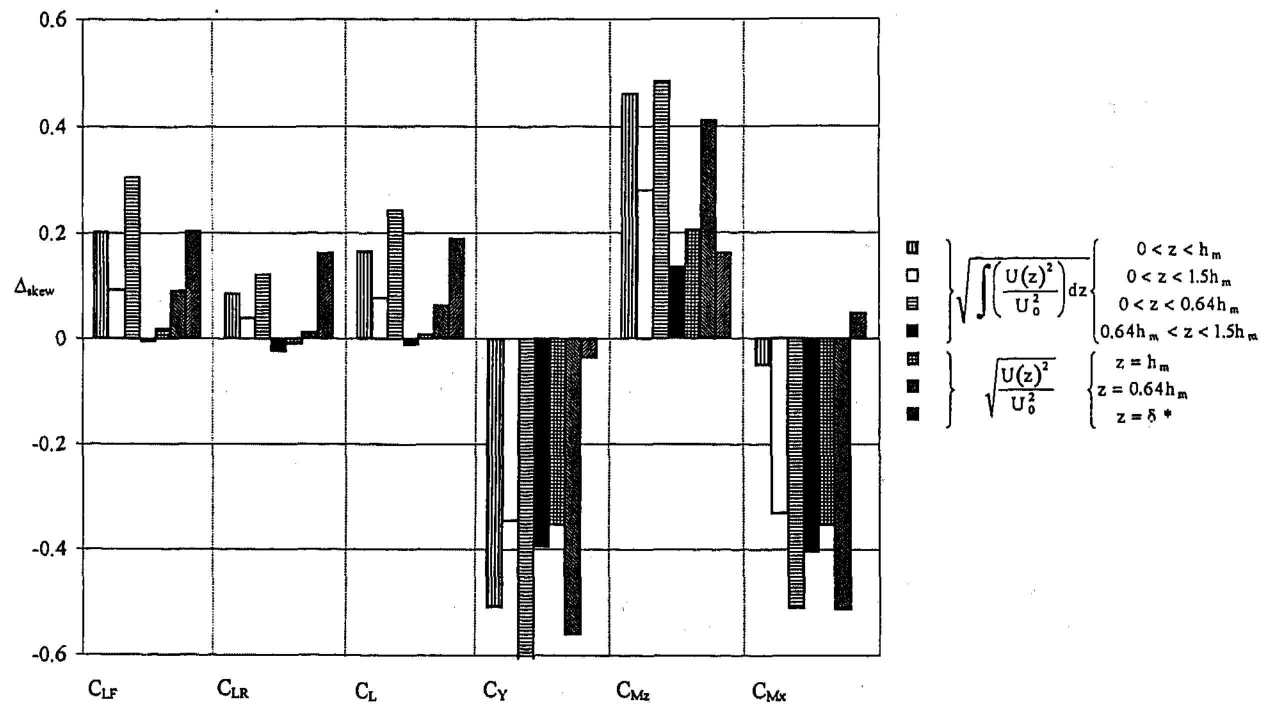


Figure 9
The effect on correlation factor, Δ_{skew}, of a number of alternative methods for evaluating a reference velocity suitable for skew velocity profiles

970136

Contemplation of Nozzle Blockage in Open Jet Wind-Tunnels in View of Different 'Q' Determination Techniques

Edzard Mercker
German-Dutch Wind Tunnel

Gerhard Wickern and Jochen Weidemann
AUDI AG

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper deals with the correction of aerodynamic interference effects taking place between the nozzle of an open jet wind tunnel and a test model. In order to deduce correct aerodynamic coefficients these interference effects have to be allowed for in the determination of the correct wind tunnel speed. In open jet wind tunnels basically two different methods are used to determine the tunnel speed. One is the so-called nozzle-method, utilizing the pressure difference down the nozzle to determine the nozzle exit velocity or tunnel speed. The other procedure is the so-called plenum-method, where the pressure difference between the settling chamber and the surrounding plenum chamber of the test section is measured and used. In this paper it is shown that both methods yield a systematic error, since the velocity distribution in the nozzle differs from the velocity distribution in an unbounded stream measured at the same distance from the model. In order to compensate for this effect a correction factor is derived with the aid of a simple potential flow model.

INTRODUCTION

When comparing aerodynamic results measured in different open jet wind tunnels it is important to consider interference effects imposed upon the test results by the test section, the nozzle and the collector. On the assumption of small perturbation theory E. Mercker and J. Wiedemann [1] identified four different interference effects occurring in open jet wind tunnels. Neglecting higher order terms, they derived an overall correction factor by summing up the interference resulting from the proximity of the model to the nozzle and the flow collector, the limited extent of the jet dimensions and the horizontal buoyancy effect due to the static pressure gradient often present in empty test sections.

Applying the correction formula to a number of data sets, measured in a large number of wind tunnels, they showed that the correction procedure works for scale models as well as for full scale passenger cars of various sizes. On average a standard deviation between 2 and 3 drag-counts (1 drag-count is CD = 0.001) was attained between the different wind tunnels after correction. Considering that all possible measuring errors and model set-up errors are included, this result can be regarded as satisfactory. Additionally, the results were compared with blockage-corrected data measured in closed wall wind tunnels. The agreement between the two types of wind tunnels was remarkably good.

However, the universal application of the correction formula of Mercker and Wiedemann is restricted to data which are obtained in wind tunnels where the test section reference speed is measured with the so called nozzle-method. This method uses the pressure difference between two cross-sections of the nozzle to determine the tunnel speed. A second method for stipulating the free stream speed is the so called plenum-method. In this method the pressure difference between the settling chamber of the wind tunnel and the plenum hall surrounding the open jet is measured. The plenum-method has not been considered in the correction procedure of Mercker and Wiedemann.

Since a model exposed to a flow generates a pressure gradient in the axial direction, the velocity

distribution in the nozzle of an open jet wind tunnel is affected and tends to deviate from the distribution of an infinite flow. Therefore, the reference velocity of the test section may be evaluated falsely. Apparently, independently of which method is used to determine the tunnel speed, both methods need to be corrected due to this nozzle effect.

This fact is in opposition to statements given in the published literature. For instance, A. Kuhn [2] and R. Künstner et al. [3] claim that either the nozzle-method or the plenum-method is the only correct way to determine the tunnel speed and both authors disregard completely any correction of the tunnel speed with respect to the nozzle effect.

The subject of this paper is to give a new detailed physical explanation of the two methods to determine the tunnel speed. Due to the fact that both methods entail a correction, the procedure of Mercker and Wiedemann is extended here to wind tunnels where the plenum-method is employed.

CORRECTION METHODOLOGY

The dominant effect of interference in wind tunnels is usually taken to be equivalent to a simple modification of the free stream velocity expressed by a perturbation velocity ε

$$U_{cor} = U_\infty + \Delta U = U_\infty (1+\Sigma\, \varepsilon) \qquad (1)$$

where U_{cor} denotes the velocity corrected for interference effects and $\Sigma\, \varepsilon$ is the algebraic sum of all perturbation velocities originating from different fluid dynamic interactions. Expressed in the form of the dynamic pressure q_{cor}, this leads to

$$q_{cor} = q_\infty(1+ \varepsilon_S + \varepsilon_C + \varepsilon_N)^2 \qquad (2)$$

where ε_S denotes the perturbation velocity due to jet expansion, ε_C represents the collector blockage effect and ε_N is the nozzle blockage factor. In this equation, it is assumed that the magnitude of a single interference effect is small and thus higher order terms due to coupling of different interference effects may be neglected. Therefore, the different effects can be simply superimposed and a total perturbation velocity comprising all interactions can be defined. Furthermore, the so-called "invariance under constraint" has to be assumed, which implies that the flow pattern about the model remains unaltered, even with interference effects present. Comparing, for instance, the velocity profiles in front of a vehicle in an infinite stream with those in a wind tunnel, it is assumed that the profiles remain self-similar. In other words, it is assumed that the distribution of the velocity is invariant under constraint and the velocity vectors deviate only by a constant scaling factor.

As mentioned above, this paper deals with the detailed analysis of nozzle interference effects or nozzle blockage. Therefore, for a detailed deduction of ε_S and ε_C as well as for the effect due to a static pressure gradient reference is made to [1]. In this context, it is noteworthy that the effect of the empty test section pressure gradient $\Delta C_{D\,HB}$ has to be treated differently and is introduced into the correction procedure before any dynamic pressure corrections are applied. Regarding the drag coefficient, the entire correction formula for open jet tunnels can be expressed as:

$$C_{D\,cor} = (C_{Dm} + \Delta C_{D\,HB})/(q_{cor} / q_\infty) \qquad (3)$$

Here $C_{D\,cor}$ denotes the corrected and C_{Dm} denotes the measured drag coefficient.

For convenience, the total correction formula comprising all necessary geometrical and fluid-dynamic properties will be given at the end of this paper.

DETERMINATION OF WIND SPEED

The wind speed in the test section of a wind tunnel is determined by measuring the static pressure difference Δp between the settling chamber p_{SC} and a second pressure located either somewhere downstream on the nozzle contour (p_N) or in the plenum chamber (p_P) surrounding the open jet. If the downstream pressure tap is installed in the nozzle contour, the method is called "nozzle-method". With the reference measured in the plenum chamber, the method is called "plenum-method". Fig.1 shows the general set-up for both kinds of measurements.

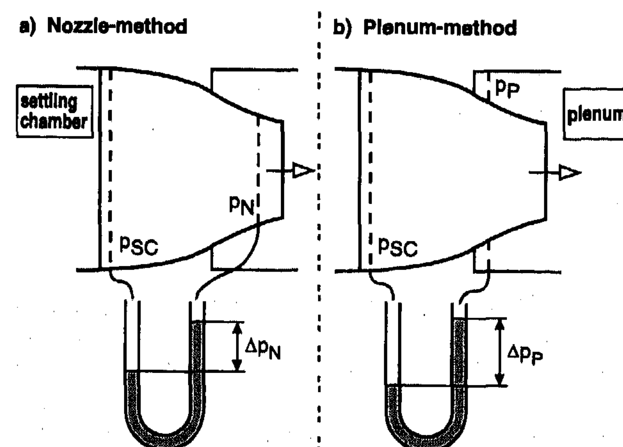


Fig. 1 General set-up for nozzle-method and plenum-method

The procedure for determining the dynamic pressure q_∞ of the jet is in principle the same for both cases.

The pressure difference of the nozzle (Δp_N) or plenum-method (Δp_P) is related to the pressure difference of a Pitot-static tube measured at the center of the jet of the empty test section:

$$q_\infty = \rho/2\ U_\infty = k\ \Delta p \qquad (4)$$

Usually, the numerical value of the nozzle factor k is greater than unity, because the wind speed in the settling chamber is not exactly zero and hence the local static pressure there is below the total pressure. Furthermore, due to the fact that the reference static pressure is measured at different locations, k is different for both methods. For a more accurate determination of k, it is recommended that the relation between Δp and q_∞ is approximated by a third degree polynomial. The reason for the slightly non-linear behaviour of k, especially in the low velocity regime, is observed in the Reynolds-number dependency of the nozzle wall boundary layer.

The decision, of which method should be used in a wind tunnel for determining the tunnel speed, is largely dependant upon the wind tunnel dimensions and the dimensions of the model. Due to the fact that an adverse pressure gradient in front of a model may extend into the nozzle, the measurement of the static reference pressure may be falsified with the nozzle method.

Therefore, for small wind tunnels used for full scale automotive testing, the plenum method is recommended, because the reference pressure is located in the surrounding plenum chamber which usually is ventilated to the ambient air. Thus, for this method the reference pressure is constant and independent of the model size and model position inside the test section. Strangely enough, for some existing automotive wind tunnels the situation is somewhat reversed. For wind tunnels with a large nozzle and often larger distance between pressure tap and test vehicle the plenum method is sometimes advocated, whereas for wind tunnels with small nozzle sizes the nozzle-method is often applied. As mentioned above, both methods need a correction because the presence of the model affects the nozzle exit flow to some extent.

However, it is important to realize, that once the pressure at the tap location (speed holes) is contaminated by the adverse pressure gradient originating from the test vehicle, a correction procedure is no longer meaningful and incorrect aerodynamic coefficients will result.

The problem may be by-passed by locating the static reference pressure tap at a more upstream position until the model influence has vanished. However, this may take place at the expense of a smaller pressure signal and thus a smaller resolution for the tunnel velocity. Fig. 2 shows experimental data measured in one of DNW's model scale wind tunnels. Depending on the model position in the test section, a different streamwise static pressure distribution is measured at the nozzle wall. In this example it is depicted that the pressure distribution converges and is independent of the model location if the downstream pressure tap is located at the position indicated by N.

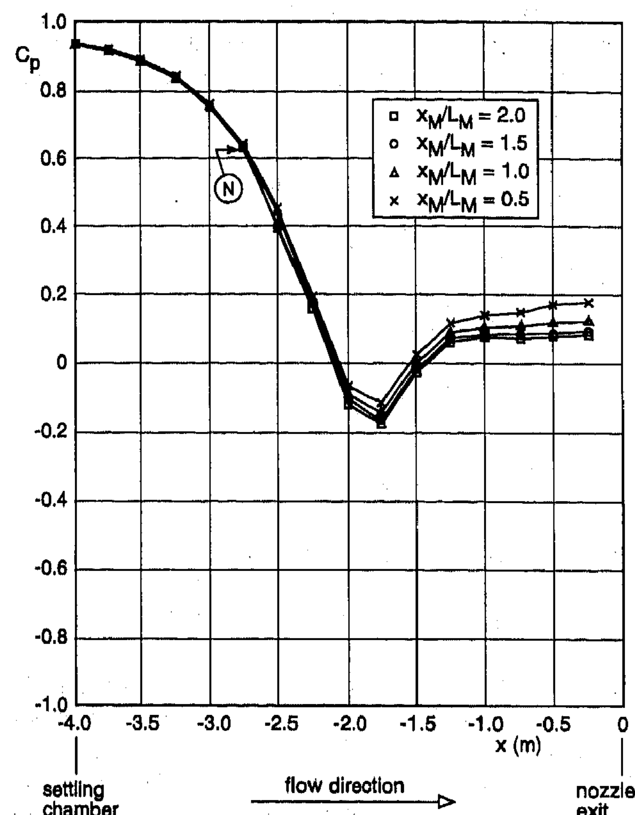


Fig. 2 Static pressure distribution in the nozzle for different model positions; L_M = model length; x_M = distance model center/nozzle

BLOCKAGE EFFECT FOR THE NOZZLE METHOD

As mentioned above, one way to determine the wind speed in an open jet wind tunnel is the nozzle-method, utilizing the pressure difference over the nozzle, see Fig 1a. Assuming that the model influence does not contaminate the downstream pressure reading, the volume flux through the nozzle is constant, regardless of where the vehicle is placed inside the test section and regardless of the geometrical dimensions of the model.

The test section length of an open jet wind tunnel is often limited due to jet stability reasons and due to a decreasing streamwise potential core length of the jet. Therefore, the vehicle is usually placed relatively close to the nozzle exit. As a consequence, the

adverse pressure region in front of the vehicle disturbs the flow inside the nozzle exit plane. In the next chapter this effect will be quantitatively determined with the aid of a simple flow model.

PRESSURE INDUCED BLOCKAGE EFFECT - With a constant volume flux through the nozzle, a blockage effect occurs due to a displacement effect of the flow, and locally a velocity higher than the calibrated tunnel speed is attained in the nozzle exit plane. This situation has to be compared with the flow of an infinite stream.

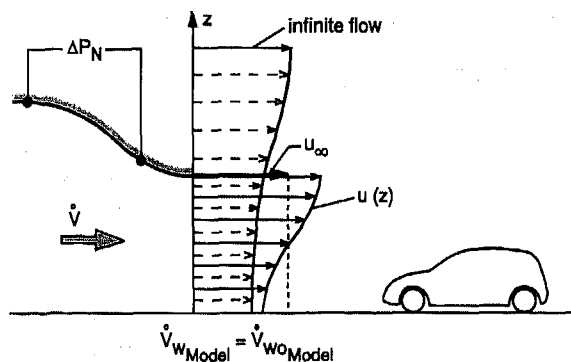


Fig. 3 Qualitative velocity distribution for nozzle-method. ($\dot{V}_{W_{Model}}$ = Volume flux with model; $\dot{V}_{WO_{Model}}$ = Volume flux without model)

Fig.3 shows qualitatively the velocity distribution in front of a vehicle for both cases at the same axial position. Whereas for an infinite flow the free stream velocity will be reached at a large lateral distance only, an appreciably higher velocity is attained at the edge of the nozzle exit for the wind tunnel flow. Consequently, the model experiences a higher aerodynamic force which is non-dimensionalised with a tunnel-q, that is too small. Therefore, a correction has to be applied to compensate for this effect.

Apparently, the drag correction of a model depends on the ratio of the momentum flux of an infinite flow I_∞ and an increased momentum flux I_N due to blockage. As momentum flux is proportional to the dynamic pressure of the flow, it can be written:

$$C_{D\,cor} = (C_{D\,m})_N\,(q_N/q_{cor}) = (C_{D\,m})_N\,/\,(1+\Sigma\,\varepsilon)^2 \qquad (5)$$

with $q_{cor} = q_N\,(1+\Sigma\,\varepsilon)^2$

The index N denotes that the quantities are measured with the nozzle method (at a later stage, when we are dealing with the plenum method the index will be changed to P). The index m denotes the uncorrected measured aerodynamic quantity.

In order to evaluate the nozzle blockage effect, we first look at the interference effects taking place at the nozzle. In a second step we evaluate the reaction of these effects upon the model which is placed in the test section further downstream.

Taking now the interference effect at the nozzle to be equivalent to a modification of the tunnel speed U_∞, a perturbation velocity ε_{QN} can be defined in the nozzle exit plane of the wind tunnel:

$$\varepsilon_{QN} = \Delta U/\,U_\infty = [\,U_\infty - (U_N\,)_N\,]\,/\,U_\infty \qquad (6)$$

where $(U_N\,)_N$ denotes velocity of the unconstraint flow, integrated over the fictitious nozzle cross sectional area A_N.

$$(\,U_N\,)_N = 1/A_N \int u(z)\,dA \qquad (7)$$

A simple potential-flow model consisting of a point source and a parallel flow is used to calculate the velocity distribution of an infinite stream at the nozzle location. The flow model will result in a half infinite body of revolution which represents the test vehicle plus its mirror image. The same approach has been followed by E. Mercker and J. Wiedemann [1] . Only the resulting equation for ε_{QN} will therefore be cited:

$$\varepsilon_{QN} = [A_M\,/(2\,A_N)][1 - x_S/(\,x_S^2 + R_N^2\,)^{1/2}] \qquad (8)$$

where

$$x_S = x_M - L_M\,/2 + [\,A_M/\,(2\,\pi)]^{1/2} \qquad (9)$$

denotes the distance from the nozzle to the source point. Furthermore, x_M and L_M denote the distance from the nozzle to the model center and the vehicle length, respectively. A_M denotes the frontal area of the model and A_N is the cross-section of the nozzle. Finally, R_N denotes the equivalent duplex nozzle radius (for symmetry reason the duplex lateral wind tunnel dimensions are used in the corresponding formulas):

$$R_N = (\,2\,A_N/\pi\,)^{1/2} \qquad (10)$$

As already mentioned in the paper of Mercker and Wiedemann [1], the suggested potential-flow model is a simple representation of the test vehicle. However, as we are dealing with far-field effects in a potential-flow environment, the model may be adequate, nevertheless. For trucks, busses and light vans with a blunt fore-body a more refined flow model may be advisable, however.

Beside the interference effect due to an adverse pressure gradient extending into the nozzle, two other interference effects occur. Again, the reason for these flow effects has to be seen in the interaction between model and nozzle:

1. Due to the solid walls of the nozzle the streamlines will be confined compared to an infinite flow.
2. Due to flow displacement about the model the flow will be deflected at the nozzle contour.

STREAMLINE CONSTRAINT CORRECTION - First, we look at the effect due to streamline constraint and, Fig. 4 illustrates this effect schematically for an arbitrary chosen body of revolution. It can be seen that the free expansion of the streamlines for an infinite flow is confined due to the fact that the flow is

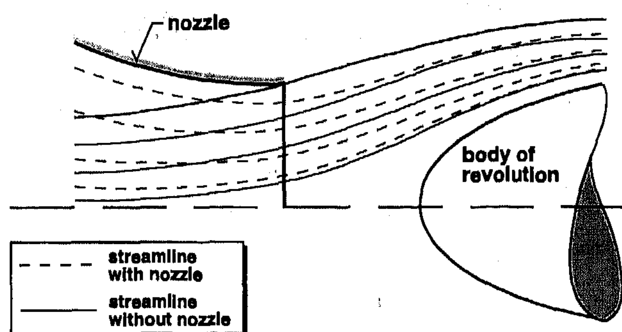


Fig. 4 Streamline constraint due to nozzle proximity

constrained by the nozzle.

It is important to realize that in principle the "invariance under constraint" is no longer fulfilled in this case. However, it will be assumed that this effect can be restricted to the flow-field at nozzle position and therefore the flow field about the model can be considered invariant under constraint

In order to quantify the interaction between nozzle and model a simple flow model shall be employed. For this purpose the nozzle-flow is replaced by a vortex ring superimposed on a parallel flow at the nozzle position. For illustration see Fig. 5.

Herein, the parallel flow represents the undisturbed stream conditions and the vortex flow represents the streamline constraint of the nozzle, which has to be seen in combination with the effect of the adverse pressure gradient in front of the vehicle, because flow constraint occurs only if an interference effect between model and nozzle is present. Therefore, it is assumed that the strength or circulation Γ of the vortex is proportional to the perturbation velocity ε_{QN} at the center of the vortex-ring cross-section. For the circulation Γ it can be written:

$$\Gamma = 2\, R_N\, U_\infty\, \varepsilon_{QN} \qquad (11)$$

where R_N denotes the radius of the nozzle.

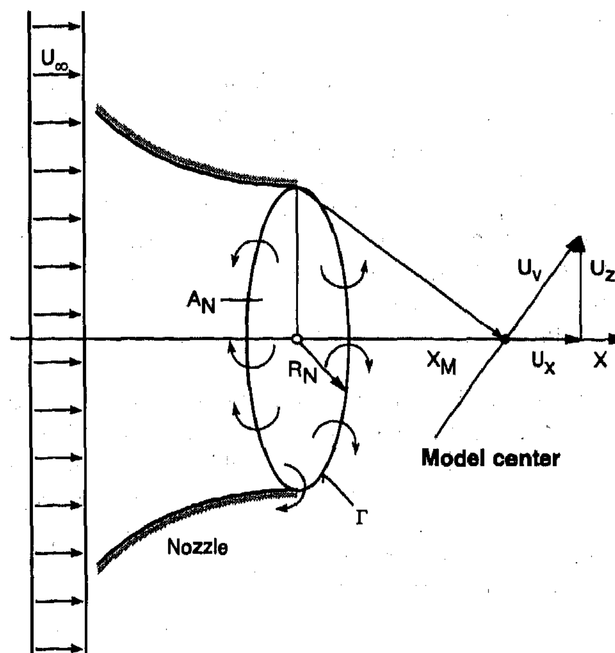


Fig. 5 Induced velocity U_V due to vortex ring with circulation Γ

Applying Biot-Savart principles, it is now possible to calculate an induced velocity U_V of the vortex flow at the model position. The streamwise component U_x of this velocity will be used for correction.

$$U_x = [\Gamma/2]\, [\, R_N^{\,2} / (\, R_N^{\,2} + x_M^{\,2})^{3/2}] \qquad (12)$$

The perturbation velocity at the model position x_M can then be calculated:

$$\varepsilon_N = \varepsilon_{QN}\, R_N^{\,3} / (\, R_N^{\,2} + x_M^{\,2})^{3/2} \qquad (13)$$

Thus, for the example of the drag coefficient, the following correction formula can be derived. With respect to Equation (2) and (3) it can be written

$$C_{D\,cor} = [\, (C_{D\,m})_N + \Delta C_{D\,HB}] \,/\, (1 + \varepsilon_S + \varepsilon_Q + \varepsilon_N\,)^2 \qquad (14)$$

It is noteworthy that ε_N is smaller in magnitude than ε_{QN}, depending on the distance x_M between model and nozzle. The reason for this has to be seen in the fact that the vortex ring induces a vertical velocity component U_z at the model position which tends to widen the jet. For continuity, the axial velocity has to decrease accordingly. In other words, the effect of streamline constraint at the nozzle in combination with the adverse pressure gradient of the model generates an interference effect at the nozzle location that attenuates in the downstream direction.

BLOCKAGE EFFECT FOR THE PLENUM-METHOD

For the plenum-method the pressure difference between the settling chamber and the plenum chamber surrounding the open jet is measured, see Fig. 1b. Since the plenum chamber is vented to the ambient air, the static pressure is constant there and independent of the model arrangement inside the test section. Additionally, since the most upstream pressure tap in the settling chamber is usually far ahead of the vehicle the second pressure location is also not affected by the vehicle. Thus, although the velocity distribution in the nozzle exit plane may be disturbed by the model, the velocity at the circumferential contour of the nozzle exit will be equal to the free stream velocity due to Bernoulli's principles. Consequently, regardless of model size and model position the calibrated tunnel speed of the empty test section can not be exceeded at the boundary of the jet and remains constant. Within the jet, however, the velocity can assume any arbitrary value depending on the model geometry and its distance from the nozzle.

It is important to note that the constancy of the static pressure is fulfilled only for sufficiently large plenum chambers. For a small plenum, strong flow recirculation may take place when a model is placed into the test section, if due to the so called "Coanda-effect" the jet attaches to the walls or the roof of the plenum. This tends to lower the static pressure in the plenum hall. R. Deutenbach [4] has evaluated the Coanda effect for open jet wind tunnels experimentally. In his work criteria are derived for the minimum size of the plenum chamber needed to prevent significant recirculation.

Fig. 6 shows model scale measurements, confirming the constancy of the velocity versus model position for the plenum-method. During this experiment the model was moved towards the nozzle in steps. The wind tunnel speed V_T was determined either by the plenum-method or by the nozzle method. Simultaneously the velocity V_{vane} at the outer contour of the nozzle exit was measured by a vane anemometer. Apparently, with the nozzle-method the velocity ratio varies gradually with the model position, whereas with the plenum-method a more or less constant value is attained.

In order to assess the model/nozzle interaction for the plenum method, the situation in the wind tunnel has to be compared again with the flow in an infinite stream. Fig. 7 shows schematically the velocity distribution in front of a vehicle for the free stream case and for the wind tunnel flow. It is interesting to note that for the plenum-method the volume flux through the nozzle is now smaller with the model installed than the volume flux of the empty test section.

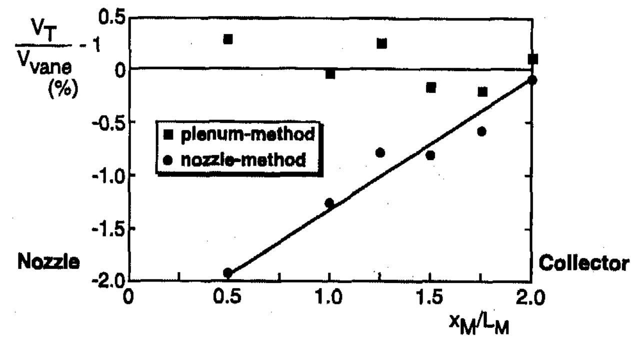


Fig. 6 Relative variation of measured V_{vane} and calibrated V_T tunnel speed versus model position x_M

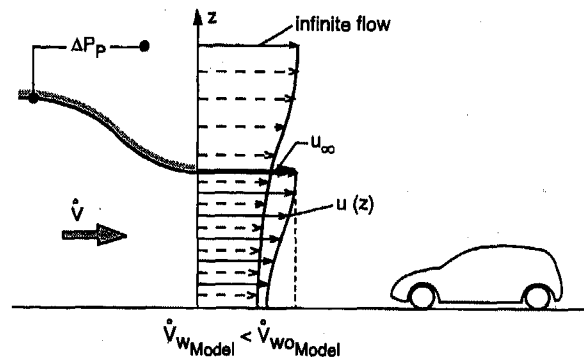


Fig. 7 Qualitative velocity distribution for plenum-method. ($\dot{V}_{W_{Model}}$ = Volume flux with model; $\dot{V}_{WO_{Model}}$ = Volume flux without model)

The magnitude is dependent on the model size and the model position inside the test section. In contrast, the volume flux for the nozzle-method is constant and independent of the model and its position, as explained above. Similarly to closed wall wind tunnels, it is the latter property which eventually enables the correction of the streamwise aerodynamic quantities.

Assuming self-similar velocity profiles at the nozzle position also for the plenum-method, a perturbation velocity ε_{QP} can be defined by the ratio of the velocity vectors between the infinite flow and the flow in a wind tunnel.

$$\varepsilon_{QP} = \Delta U / U_\infty = [\, U_\infty - (U_N)_P] / U_\infty \qquad (15)$$

where $(U_N)_P$ denotes the horizontal velocity component at the edge of a fictitious nozzle in a non-constraint flow. - Since, for the plenum method, at the outer contour of the nozzle the free stream velocity U_∞ is yielded, see Fig. 7, integration over the cross-section of the nozzle is not necessary. The perturbation velocity ε_{QP} can then be determined with the aid of the same potential flow model, also used for the nozzle-method, by calculating the horizontal velocity component at a

distance x_S and a radius R_N from the source point. With

$$(U_N)_P = U_\infty [1 - A_M/(2\pi)] [x_S/(R_N^2+x_S^2)^{3/2}] \quad (16)$$

and Equation (15) it can be written:

$$\varepsilon_{QP} = [A_M/(2\pi)] [x_S/(R_N^2+x_S^2)^{3/2}] \quad (17)$$

For distinctness it should be noted that for the nozzle-method the integration of the velocity over the fictitious nozzle is necessary to evaluate ε_{QN} because through the volume flux only the averaged velocity of the wind tunnel flow is known. In contrast, for the plenum-method merely the velocity at the edge of the fictitious nozzle has to be calculated to evaluate ε_{QP}. As shown above, in a wind tunnel U_∞ will always be attained at the outer edge of the nozzle, regardless of where the model is placed in the test section.

Furthermore, it is important to realize that for the same dynamic pressure adjusted either by the nozzle-method or by the plenum-method different aerodynamic forces will be experienced by the car. These forces are proportional to the momentum flux discharged by the nozzle (I_N = momentum flux for the nozzle-method, I_P = momentum flux for the plenum-method). Assuming self-similar velocity profiles in the nozzle exit plane, in terms of the drag coefficient it can be written:

$$(C_{D\,m})_N / I_N = (C_{D\,m})_P / I_P \quad (18)$$

Due to the fact that the momentum flux is proportional to the square of the volume flux or the dynamic pressure of the flow, it can be concluded that, compared to an infinite flow, the momentum flux I_{inf} through the cross-section of the nozzle is increased by a factor $(1 + \varepsilon_{QN})^2$ for the nozzle-method and $(1 + \varepsilon_{QP})^2$ for the plenum-method:

$$I_N = I_{inf} (1 + \varepsilon_{QN})^2 \quad \text{and} \quad I_P = I_{inf} (1 + \varepsilon_{QP})^2 \quad (19)$$

With Equation (5) the corrected drag coefficient for the plenum-method can then be expressed by:

$$C_{D\,cor} = [(C_{D\,m})_P / (1+\Sigma\,\varepsilon)^2][(1 + \varepsilon_{QN})^2/(1 + \varepsilon_{QP})^2] \quad (20)$$

Finally, by neglecting higher order terms $O(\varepsilon^2)$, Equation (20) reduces to

$$C_{D\,cor} = (C_{D\,m})_P / (1 + \varepsilon_{QP} - \varepsilon_{QN} + \varepsilon_S + \varepsilon_C + \varepsilon_N)^2 \quad (21)$$

Considering buoyancy forces as well, to a good order of approximation it can be written:

$$C_{D\,cor} = [(C_{D\,m})_P + \Delta C_{D\,HB}] / [(1+2(\varepsilon_{QP}-\varepsilon_{QN}+\varepsilon_S+\varepsilon_C+\varepsilon_N)] \quad (22)$$

JET DEFLECTION CORRECTION

Besides the confinement of the flow due to the nozzle, the streamline will be further deflected. This takes place at the nozzle due to flow displacement originating from the model. Whereas the vortex model is designed to fulfil the fluid dynamic boundary condition at the center of the jet, the flow tends to be deflected at the outer contour of the nozzle to assume larger flow angles. This effect is not taken care of by the vortex model. In the work of Mercker and Wiedemann it was shown that jet deflection also tends to widen the jet at the model position. Further it was shown that the nozzle may be replaced by one of reduced cross-section, so that its effective area is smaller than the geometrical area. The jet deflection can then be treated the same way as the jet expansion effect, which occurs due to the limited lateral extent of the jet. Details about how Mercker and Wiedemann derived the appropriate perturbation velocity can be taken from their SAE paper [1].

APPLICATION OF THE CORRECTION FORMULA

In the following a complete correction scheme will be outlined for the nozzle-method as well as for the plenum method. All geometrical and fluid dynamic properties needed in order to execute the correction procedure are listed.

However, in the first instance the various correction terms are summarised in Fig.8 in order to provide a better understanding and a convenient overview of the applied flow models used to deduce the different interference effects. Starting from the ideal situation of a body placed in an infinite stream, the open jet wind tunnel consisting of jet, nozzle and collector is attained in four steps. In the left hand column this situation is sketched for an idealised flow around a sphere. The center column shows how the wind tunnel flow situation can be replaced by superimposing different potential flow models. Finally, the last column yields the appropriate correction factors for the different flow cases.

In this overview, the effect of horizontal buoyancy is not included. As mentioned above, from a physical point of view this effect is treated differently. Furthermore, the effect of streamline deflection is included in the jet expansion correction.

The effect of jet expansion due to the finite width of the jet (ε_S) and the interference effect of the flow collector (ε_C) are not examined in detail in this paper. However, for completeness they are included in the overview of Fig. 8. More detailed information can be taken from [1]. At this stage it will only be mentioned that jet expansion can be evaluated with the

	Idealised flow situation		potential flow model		correction
step I		body in an infinite flow field		dipol in translatory flow	$\frac{q_{corr}}{q_\infty} = 1$
step II		body in a jet of infinite length but finite width		jet boundary modelled by image planes with alternating signs	$\frac{q_{corr}}{q_\infty} = (1+\varepsilon_S)^2$
step III		the nozzle at finite distance of the body (streamline constraint by solid walls of nozzle)		local flow modification by a vortex ring to simulate streamline constraint of nozzle	$\frac{q_{corr}}{q_\infty} = (1+\varepsilon_S+\varepsilon_N)^2$
step IV		the collector at finite distance behind the body (second streamline constraint)		second vortex ring behind the model for collector effect	$\frac{q_{corr}}{q_\infty} = (1+\varepsilon_S+\varepsilon_N+\varepsilon_C)^2$

Fig. 8 Potential flow model for the open jet wind tunnel

imaging technique whereas the flow collector effect can be quantified with the aid of a ring-vortex model, similar to the one for the nozzle effect.

The entire correction formula for the drag coefficient measured in open jet wind tunnels can be expressed as:

$$C_{Dcor} = (C_{Dm} + \Delta C_{DHB}) / (q_{cor} / q_\infty)$$

For nozzle-method:

$$q_{cor} / q_\infty = (1 + \varepsilon_S + \varepsilon_C + \varepsilon_N)^2$$

For plenum-method:

$$q_{cor} / q_\infty = (1 - \varepsilon_{QN} + \varepsilon_{QP} + \varepsilon_S + \varepsilon_C + \varepsilon_N)^2$$

C_{Dm} = measured drag coefficient
q_{cor} = corrected dynamic pressure
q_∞ = calibrated dynamic pressure
ε = perturbation velocity

ΔC_{DHB}= **horizontal buoyancy force** (details see Ref.[1])

$$\Delta C_{DHB} = [1.75 / A_M][0.5\ V_M\ G] \text{ with}$$

$$G = [(dc_p / dx)_N + (dc_p / dx)_C]$$

$(dc_p / dx)_N$ = static pressure gradient from the nozzle of empty test section, taken over the length of half the model volume V_M.
$(dc_p / dx)_C$ = static pressure gradient from the collector of empty test section, taken over the length of half the model volume V_M.
A_M = frontal area of vehicle or model

ε_S = **jet expansion correction factor** (details see Ref.[1]).

$$\varepsilon_S = \tau\ (V_M / L_M)^{1/2} A_M / (A^*)^{3/2}$$

τ = tunnel shape factor. Constant for each tunnel. Can be extracted from published charts, see for instance Ref.[5].
V_M = vehicle/model volume
L_M =vehicle/model length
A_M = frontal area of vehicle/model
$A^* = A_N / (1+ \varepsilon_{QN})$
ε_{QN} = q-correction factor, see below

ε_C = **collector blockage correction factor** (details see Ref.[1])

$$\varepsilon_C = [\varepsilon_W R_C^3] / [(L_{TS} - x_M)^2 + R_C^2]^{3/2}$$

$\varepsilon_W = [A_M / A_C] (C_{Dm} / 4 + 0.41)$ = wake blockage factor

$R_C = (2\,A_C/\pi)^{1/2}$ = equivalent duplex collector radius
A_C = collector cross section
L_{TS} = test section length
x_M = distance nozzle - model center

ε_N = **nozzle blockage factor**

$$\varepsilon_N = \varepsilon_{QN}\, R_N^{\,3} / (R_N^{\,2} + x_M^{\,2})^{3/2}$$

$\varepsilon_{QN} = [A_M/(2\,A_N)][1 - x_S/(x_S^{\,2} + R_N^{\,2})^{1/2}]$ = blockage factor for nozzle-method
$\varepsilon_{QP} = [A_M/(2\pi)]\,[x_S/(R_N^{\,2} + x_S^{\,2})^{3/2}]$ = nozzle blockage factor for plenum-method
$x_S = x_M - L_M/2 + [A_M/(2\pi)]^{1/2}$ = distance model-source point
x_M = distance nozzle - model center
$R_N = (2\,A_N/\pi)^{1/2}$ = equivalent duplex nozzle radius
A_N = nozzle cross section
A_M = frontal area of vehicle/model
L_M = vehicle/model length

VERIFICATION OF THE CORRECTION METHOD

In order to verify the correction method, especially with respect to the nozzle blockage effect for the plenum-method, a rather limited set of data was available. In a data set, provided by EADE (European Aerodynamic Data Exchange Committee) and some other full scale wind tunnels, only two wind tunnels used the plenum method. The data set consists of full scale measurements of 10 different passenger cars in 8 different wind tunnels. In the original data set, the uncorrected data for the two wind tunnels employing the plenum method were adjusted to the nozzle method according to experimental evidence, because at that stage a correction for the plenum-method was not available. After extending the correction method of Mercker and Wiedemann to the plenum-method, the original data set was reanalysed.

Table 1 shows the uncorrected data for the drag coefficient of the 10 cars denoted by V1 to V10. The different wind tunnels are denoted by WT1 to WT8. Details about the vehicles and the wind tunnels can be found in the paper of Mercker and Wiedemann [1]. It is noteworthy that the standard deviation of the results in the different wind tunnels varies between 8 and 1 drag-count. The mean standard deviation is therefore only about 6 drag-counts, however with a variation of ± 3 drag-counts. This means, depending on the test vehicle, that it is possible to attain fairly consistent results in different wind tunnels without any correction. However, this result is rather arbitrary and therefore unreliable.

Table 2 shows the corrected data for all data measured either with the nozzle-method or with the plenum-method.

The standard deviation of the results measured in the different wind

Vehicle	V 1	V 2	V 3	V 4	V 5	V 6	V 7	V 8	V 9	V 10
Wind Tunnel										
WT 1	.336	.341	.272	.316	--	.339	--	.285	--	--
WT 2	.345	.356	.271	.325	.315	.348	.324	.288	--	--
WT 3	--	--	--	--	--	--	--	--	--	.354
WT 4	.354	.360	.273	.330	--	.359	.335	.298	.295	.352
WT 5	.349	.350	.276	.320	.314	.338	.324	.290	.288	.345
WT 6	.333	.339	.270	.318	.315	.336	.313	.285	.292	--
WT 7	.344	--	.272	.316	--	.341	.319	.286	.288	--
WT 8	.353	--	.279	.326	--	.348	.324	.297	.294	--
Standard Deviation	.007	.008	.003	.005	.001	.008	.007	.005	.003	.004

Table 1 Uncorrected measured drag coefficient

Vehicle	V 1	V 2	V 3	V 4	V 5	V 6	V 7	V 8	V 9	V 10
Wind Tunnel										
WT 1	.347	.356	.283	.329	--	.355	--	.298	--	--
WT 2	.348	.362	.276	.330	.320	.354	.329	.293	--	--
WT 3	--	--	--	--	--	--	--	--	--	.351
WT 4	.353	.359	.277	.329	--	.357	.333	.297	.294	.350
WT 5	.353	.359	.283	.328	.321	.347	.332	.297	.295	.352
WT 6	.343	.355	.283	.332	.328	.352	.329	.299	.306	--
WT 7	.353	--	.284	.328	--	.356	.334	.298	.300	--
WT 8	.358	--	.286	.334	--	.357	.334	.305	.301	--
Standard Deviation	.004	.002	.004	.002	.004	.003	.002	.003	.004	.001
Mean C_{D_D}	.351	.358	.282	.330	.323	.354	.332	.298	.299	.351
WT_{SW} (9)	.354	.359	.283	--	.330	--	.334	--	.299	--
WT_{SW} (10)	--	--	.283	--	.330	--	--	--	.301	--

Table 2 Corrected drag coefficient

tunnels has now largely decreased. An averaged standard deviation of approximately 3 drag-counts can be calculated and the variation of the standard deviation is now only ± 1 drag count. This demonstrates how well the results measured in different wind tunnels will compare after the new correction has been applied. Table 2 also includes the corrected results from two solid wall wind tunnels. These wind tunnels are denoted by $WT_{sw}(9)$ and $WT_{sw}(10)$. The agreement between the mean values of the open jet wind tunnels and the closed wall wind tunnels is remarkably good.

SUMMARY

A correction method for open jet wind tunnels, as suggested by Mercker and Wiedemann, has been extended for use in wind tunnels in which the plenum-method is used to determine tunnel speed. For this purpose the physics of both methods, the nozzle and the plenum-method, have been reconsidered. It has been shown that, for both methods, a correction has to be applied because the adverse pressure gradient in front of a model extends into the nozzle to some extent and therefore influences the velocity distribution in the nozzle exit plane. This generates a deviation from the velocity distribution that would be measured in an infinite stream at the same axial distance from the model. As a result of the nozzle interference a perturbation velocity is generated at the model position, which has to be corrected.
Application of the method to full scale passenger cars tested in a large number of different wind tunnels showed good agreement among the corrected data.

ACKNOWLEDGEMENT

The authors wish to express their thanks to Mr. Geoffrey Carr, Mr. Mark Gleason, Mr. Jim Hackett and Dr. Jorg Vagt for reviewing the paper. Their comments and constructive criticism improved the paper greatly.

REFERENCES

[1] ***E. Mercker and J. Wiedemann***:
On the Correction of Interference Effects in Open Jet Wind Tunnels.
SAE Technical Paper 960671, 1996.

[2] ***A. Kuhn:***
Ermittlung der Anströmgeschwindigkeit und der effektiven Windgeschwindigkeit in 3/4-offenen Fahrzeugwindkanälen.
ATZ Automobiltechnische Zeitschrift 97,1995.

[3] ***R. Künstner, K.R. Deutenbach and J.D. Vagt:***
Measurement of Reference Dynamic Pressure in Open Jet Automotive Wind Tunnels.
SAE Technical Paper 920344, 1992.

[4] ***K.R. Deutenbach:***
Influence of Plenum Dimensions on Drag Measurements in 3/4-Open-Jet Automotive Wind Tunnels.
SAE Technical Paper 95100, 1995

[5] Aerodynamic Testing of Road Vehicles
SAE Surface Vehicle Information Report J 2071, 1990.

970138

Effect of Cross-Winds on Motor Car Engine Cooling

Changhua Lin
RMIT

Jeffrey W. Saunders
RMIT and Monash Univ.

Simon Watkins
RMIT

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The sensitivity of cross-winds in reducing the engine cooling ability in motor cars is highlighted. Tests on three different motor cars were conducted in the Monash University full-scale wind tunnel at different yaw angles under different wind velocities. The test results show that motor car engine cooling capability decreases with an increase in yaw angles. For a wind velocity of 14 m/s, a 13% decrease in radiator cooling capability was found at a yaw angle of 20^{o} compared to a zero yaw angle. The effect of yaw angles on the engine cooling also depends on the motor car front-end configuration, but this becomes less important with increasing wind velocity. The effect of cross-winds on car engine cooling was also evaluated by on-road engine cooling tests. A convenient experimental method to measure wind velocity and yaw angle relative to a moving car is also described.

INTRODUCTION

The fuel consumption of cars is one of the major global issues. For the average car, aerodynamic drag takes a large share of useful engine power. The reduction of aerodynamic drag is important to lower the fuel consumption and this has been achieved by cars becoming more rounded in plan view and having lower front ends. Consequently, the effect of the more streamlined plan-form of modern cars has often been to reduce the yawing moment stability. With these changes to the airflow around the vehicle front end it was desirable to investigate the sensitivity of the cooling system to yaw angle.

Though many papers have been published on the subject of radiator heat transfer and design, relatively few papers have discussed the effect of wind velocity and direction on the engine cooling. Schaub and Charles [6] in 1980 investigated the ram air effect on air side cooling system performance. They found that, under ram conditions, front-end loss could be divided into a basic front-end loss measured at zero approach air velocity, which depended on the radiator-fan cooling air flow rate, and a ram-air loss which varied with the vehicle speed. Saunders and Kolodziejczyk [5] tested the effect of tailwinds on a modified full-scale car. Their tests showed that exiting hot-air recirculating through the radiator from under the car at a 1 m/s tailwind velocity reduced radiator heat rejection by 16%. At 7 m/s, the reduction increased to 28%. Due to the complexity of engine cooling wind-tunnel simulation and the commercial sensitivity of experimental results, little information on the effect of cross-winds on the engine cooling has been published. This paper addresses the effect of cross-winds on the performance of engine cooling.

SPECIFIC DISSIPATION

In order to measure the effect of cross-winds on the motor car engine cooling, a convenient parameter is required. The parameter used is called Specific Dissipation (SD). SD is defined as the heat transfer rate of a heat exchanger divided by the maximum temperature difference across the heat exchanger [7]. SD can be expressed as:

$$SD = \frac{Q}{T_{hot,in} - T_{cold,in}} \qquad (1)$$

where $T_{hot,in}$ and $T_{cold,in}$ are the hot and cold fluid inlet temperatures of the heat exchanger, and Q is the heat exchanger heat transfer rate. The application of SD on evaluating engine cooling was given by Lin et al [4]

In essence if all experimental conditions are maintained constant except for a change in front-end configuration, a 10% improvement in SD corresponds to a 10% increase in the heat rejection by the radiator.

WIND-TUNNEL ENGINE COOLING TESTS

WIND-TUNNEL TEST INSTRUMENTATION

The Monash full-scale wind tunnel has been used as a joint RMIT and Monash aero-acoustic facility. The vehicle test section is shown in Figure 1. The maximum wind velocity is about 50 m/s. The turntable provides an opportunity of investigating the cross-wind effect on engine cooling.

The vehicles used for engine cooling tests at the Monash wind tunnel were three locally-made vehicles (Vehicle A to C). The test set-up is shown in Figure 2. During engine cooling tests, the car engine was not in operation and the electric cooling fans were disconnected. The hot coolant through the radiator was provided by a heat bench located outside of the wind tunnel. The maximum coolant temperature in the heat bench was controlled by a thermostat. The coolant flow rate was measured by a magnetic flowmeter (Fischer & Porter Mini-Mag 10D1475), which had a system accuracy of within 1%. The radiator coolant inlet and outlet temperatures and radiator approach air temperature were measured by six calibrated T-type thermocouples. Signals from the magnetic flow meter and thermocouples were recorded by a host computer through a Fluke 2620A Hydra Data Acquisition Unit.

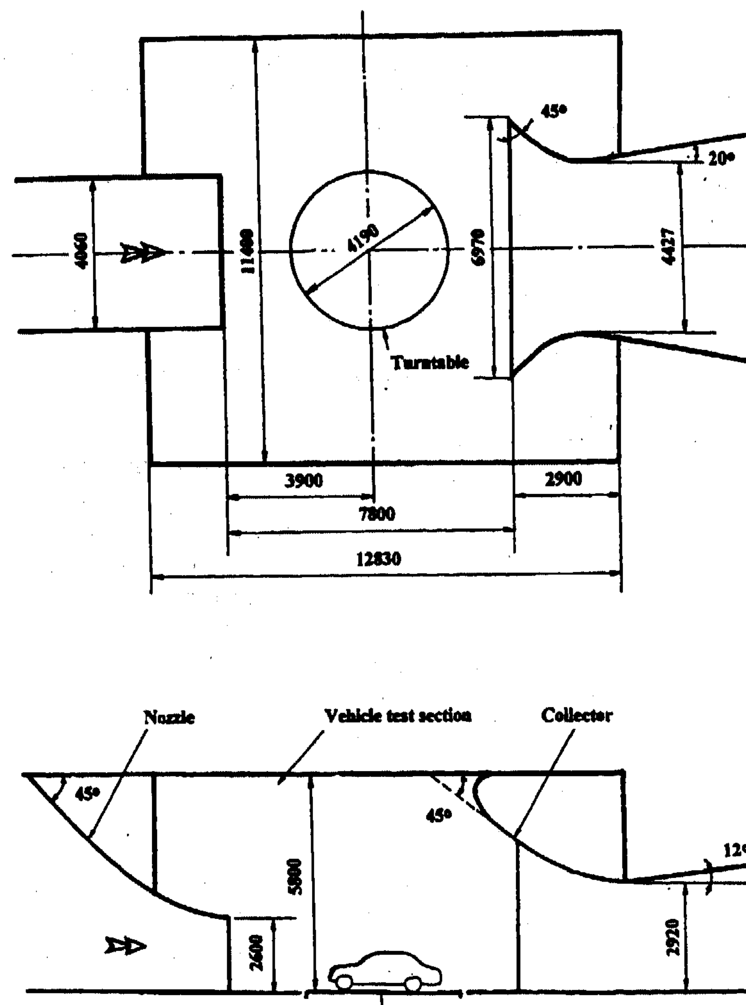


Figure 1. The vehicle test section of the Monash University wind tunnel

WIND-TUNNEL TEST RESULTS AND DISCUSSION

Tests were conducted at relative wind velocities of 14, 19, 28 and 33 m/s, and yaw angles varying from 0^{o} to 20^{o} with an interval of 5^{o}. Test results are plotted in Figures 3 to 5.

Mean SD values plotted in Figures 3(a), 4(a) and 5(a) have a standard deviation of less than 0.5%. The change of SD with yaw angle, is normalised to $[100 \times (SD_{yaw} - SD_{yaw=0}) / SD_{yaw=0}]$ and plotted in Figures 3(b), 4(b) and 5(b). Results suggest that Specific Dissipation significantly:

(a) decreases with the increase of yaw angles when other parameters (e.g., resultant wind velocity, coolant mass flowrate, etc.) are maintained constant; for example, about 13% decrease of SD is shown at a yaw angle of 20 degrees and a wind velocity of 14 m/s;
(b) depends on the car front-end configuration; and, as expected,
(c) increases with an increase in wind velocity.

The above test results indicate that tests in full-scale wind tunnels:

(a) can show the effect of cross-winds on engine cooling, and
(b) enable the automotive cooling engineers and stylists to evaluate the ranking of their designs.

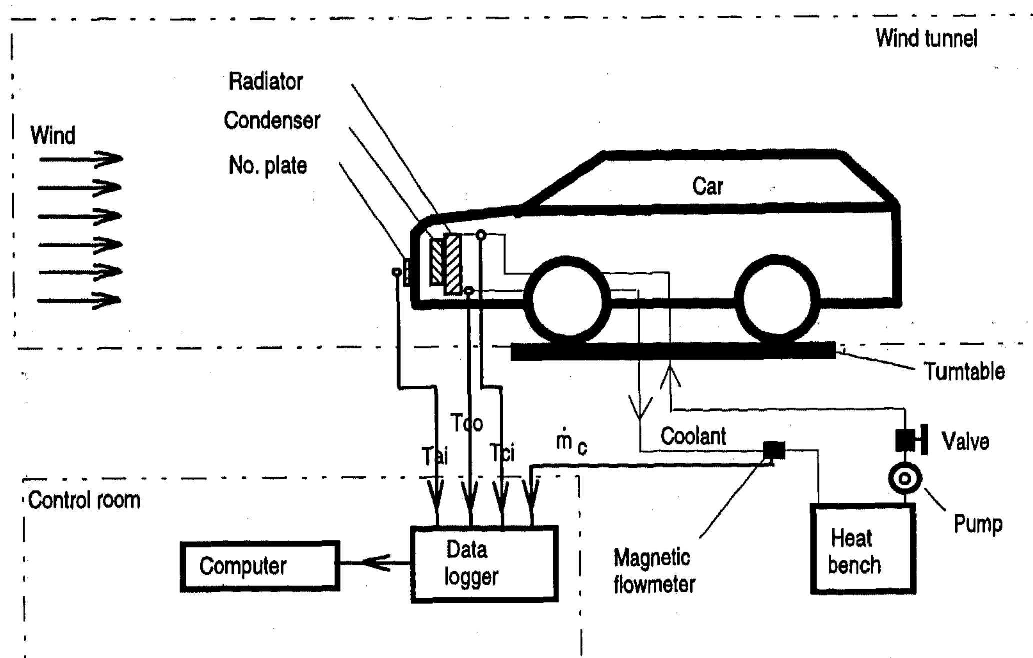


Figure 2. Car engine cooling system test layout at the Monash wind tunnel

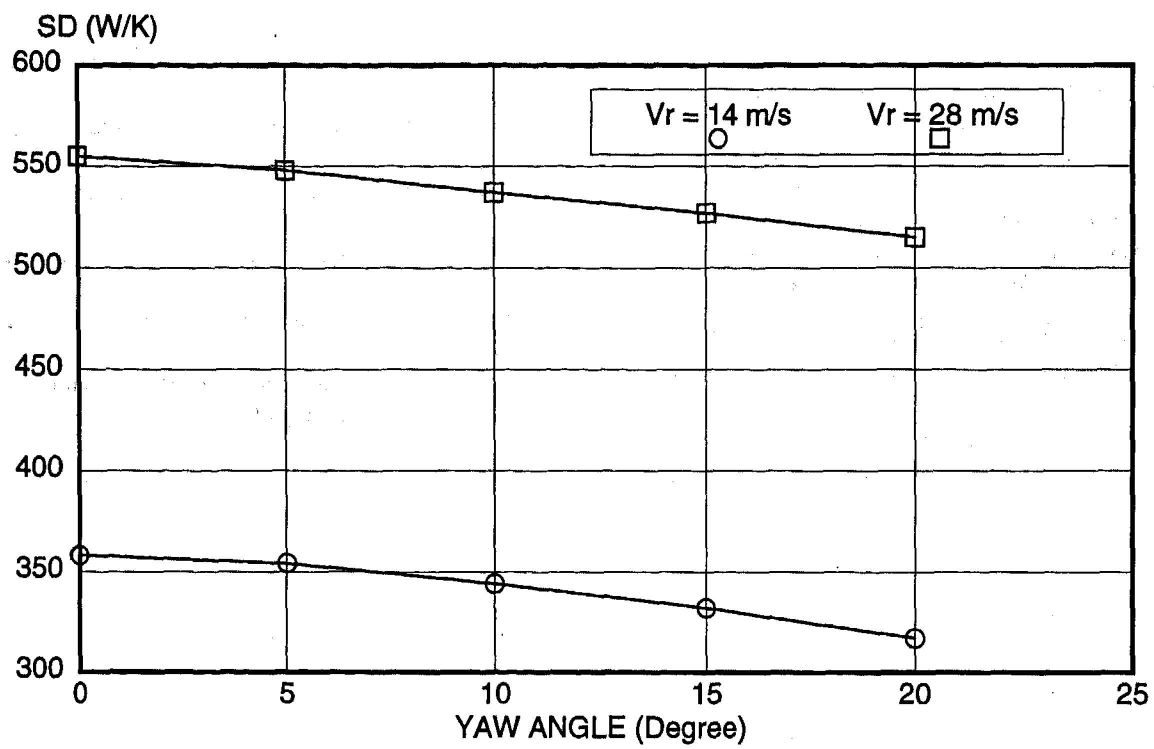


Figure 3(a). Specific Dissipation at different yaw angles

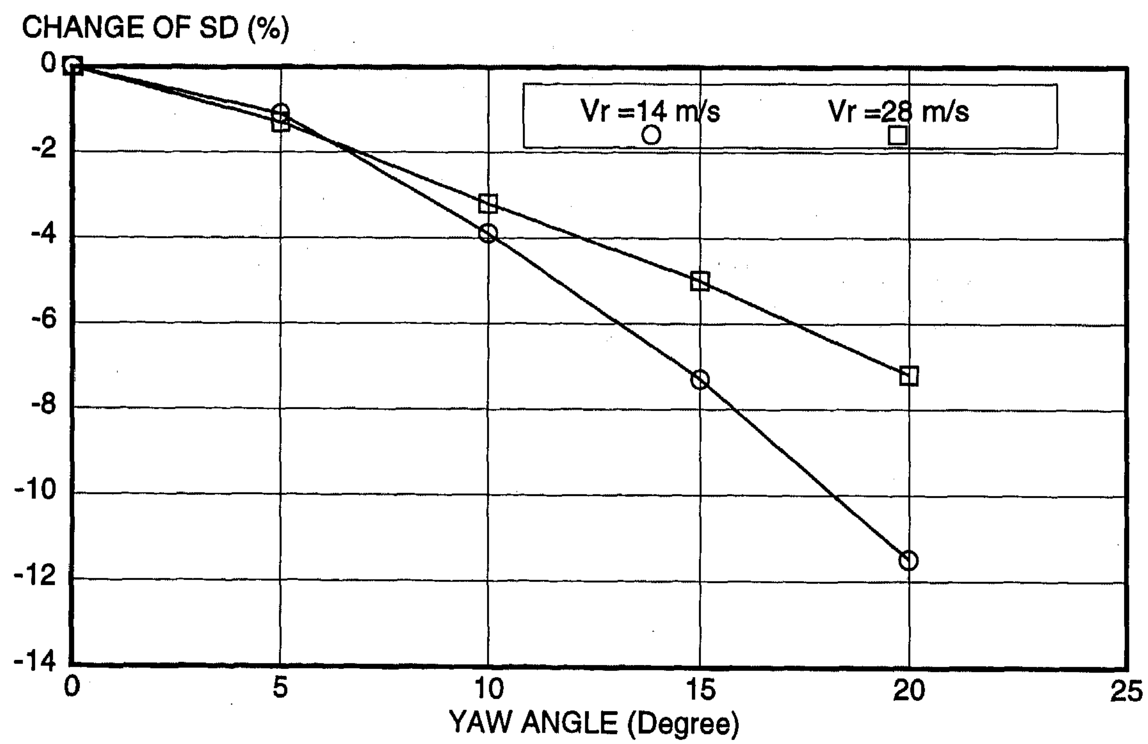


Figure 3(b). Change of Specific Dissipation with yaw angles

Figure 3. Effect of cross-winds on the motor car engine cooling for the test car Vehicle A

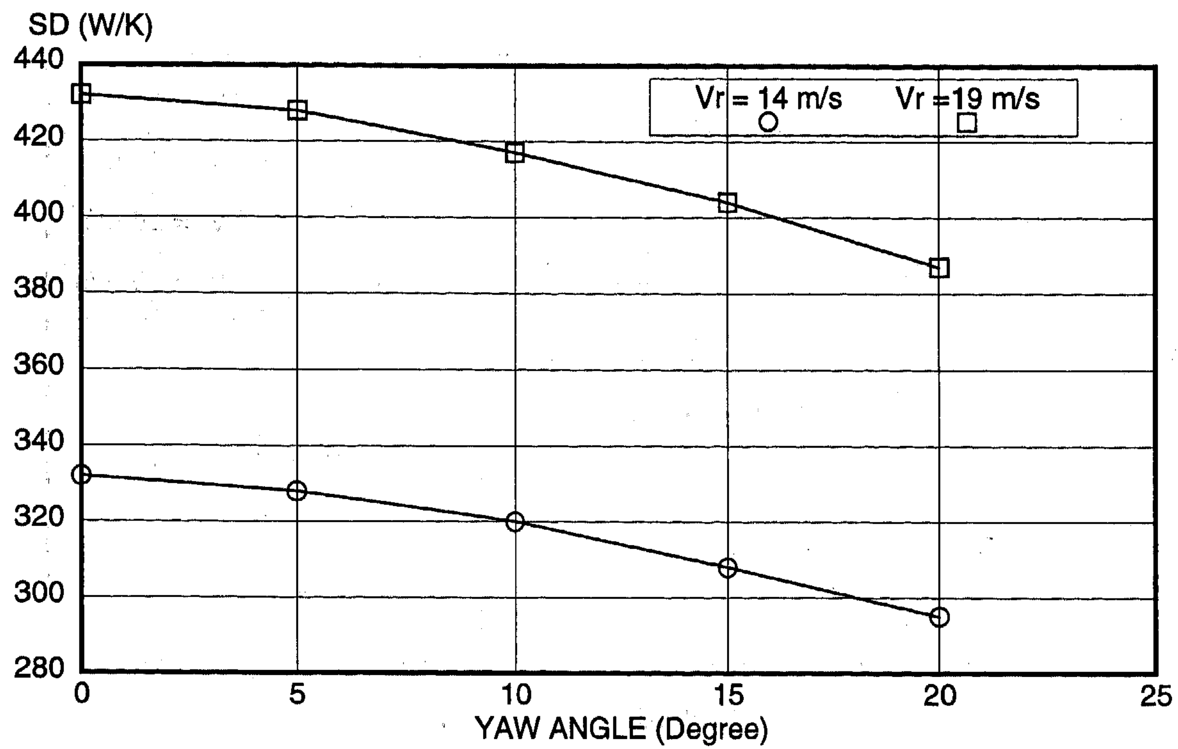


Figure 4(a). Specific Dissipation at different yaw angles

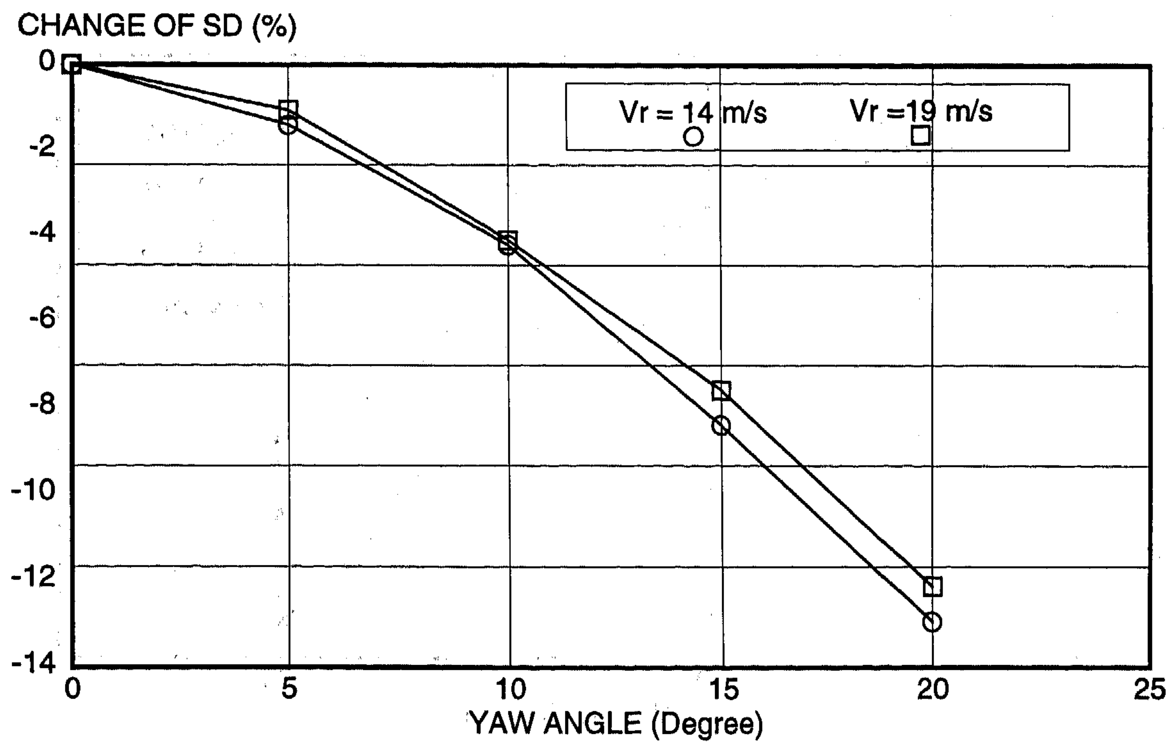


Figure 4(b). Change of Specific Dissipation with yaw angles

Figure 4. Effect of cross-winds on the motor car engine cooling for the test car Vehicle B

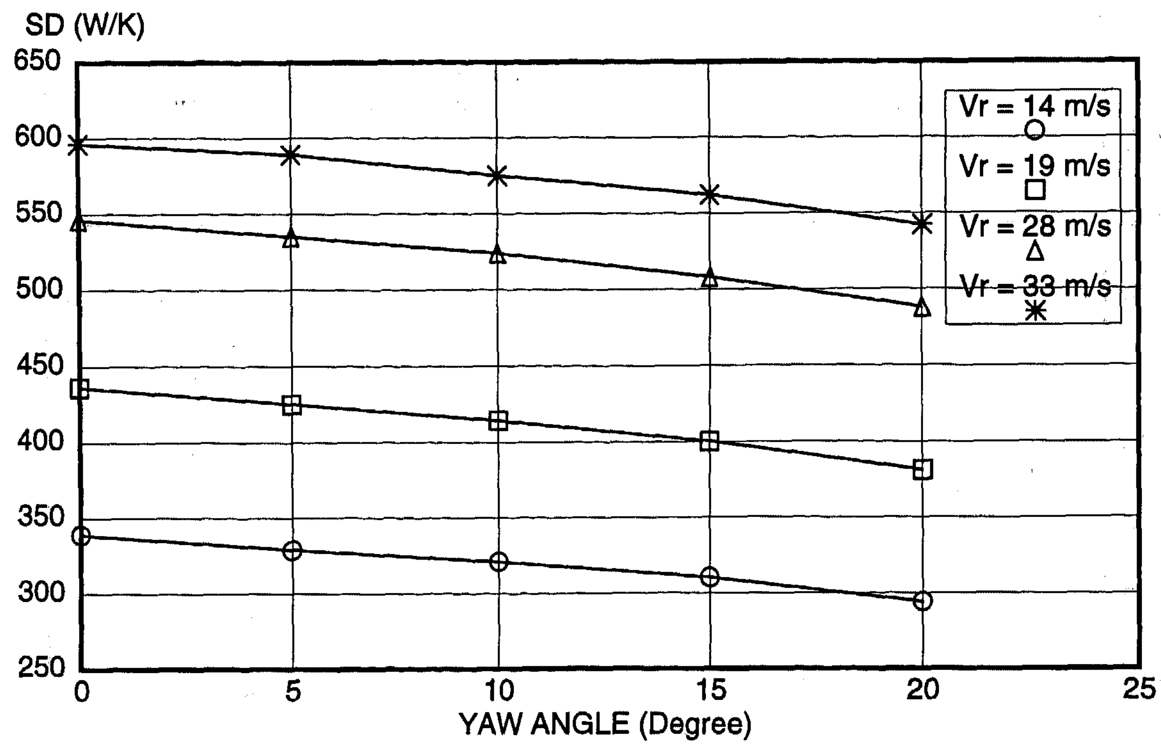


Figure 5(a). SD at different yaw angles

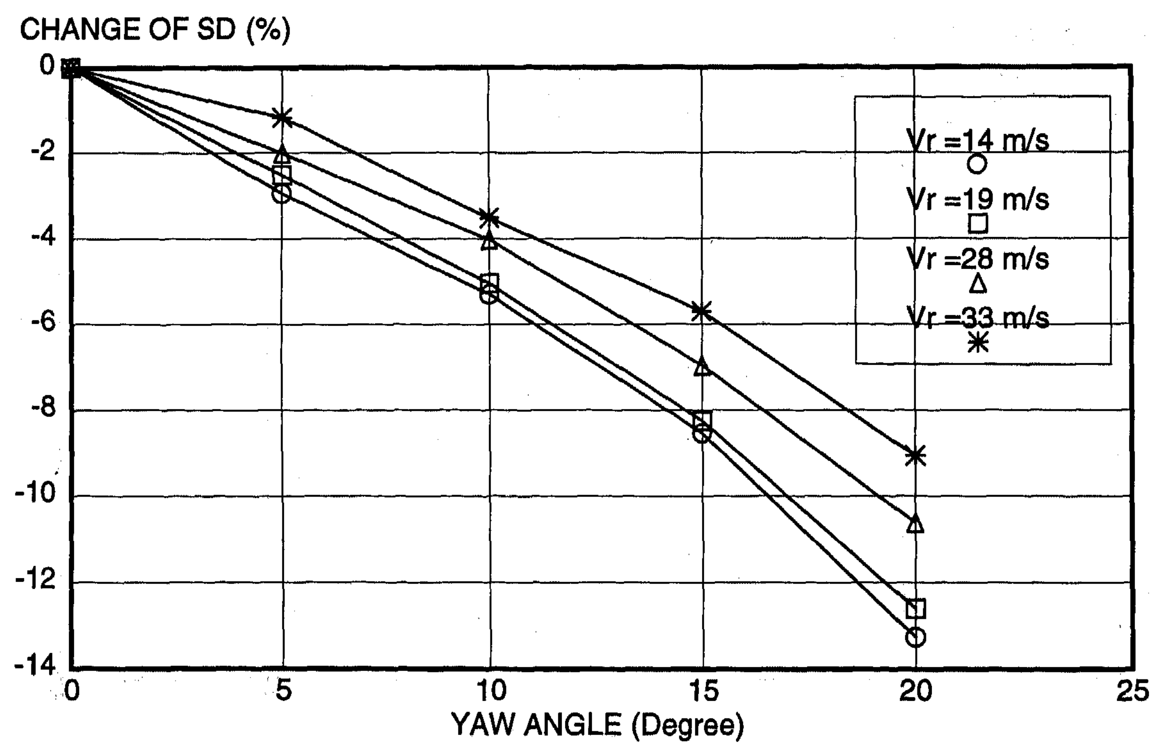


Figure 5(b). Change of SD with yaw angles

Figure 5. Effect of cross-winds on the motor car engine cooling for the test car Vehicle C

ON-ROAD ENGINE COOLING TESTS

TEST INSTRUMENTATION

The effect of cross-winds on engine cooling was also investigated by testing Vehicle A on the road. During on-road engine cooling tests, radiator approach air, and coolant radiator inlet and outlet temperatures were measured by six calibrated T-type thermocouples. Coolant flowrate was measured by a calibrated system which measured the pressure differences across a fully open thermostat [3]. Wind velocities and yaw angles were measured with a system specially developed for on-road vehicle tests. Signal outputs from all sensors were recorded by a Lap-top PC through the same Fluke data acquisition system used at the Monash wind tunnel.

Wind Velocity and Yaw Angle Measurement

Different methods to measure the relative wind characteristics (velocity and yaw angle) have been addressed by several authors. Buckley et al [1] measured wind velocity and angle during on-road tests with a Gill propeller-vane anemometer which was mounted on a mast fitted to the test car's roof rack. Watkins et al [9] also used a Gill anemometer but supplemented the measurement with cross-wire anemometers to give information at higher frequencies. In both cases the anemometers were calibrated in-situ (in-situ calibration of velocity and angle is required when anemometers are influenced by the proximity of the vehicle). Schaub and Charles [6] measured approach air velocity with a five-point Pitot-static probe rake supported one metre in front of the vehicle based on a proximity profile calibration factor established during earlier road tests. However, all those methods need to attach measuring sensors external to the vehicle. Generally this is not convenient and additionally those instruments interfere with the flow field of the car, incurring a drag penalty and generating aerodynamic noise. One of the objectives of this study was to develop a method to measure the relative wind velocity and yaw angle in a non-obtrusive manner.

Ignoring compressibility effects, then from the stagnation pressure equation for an incompressible fluid, the velocity, V, may be calculated from measurements of stagnation and static pressures [8].

$$V = \sqrt{2(P_t - P_s)/\rho} \qquad (2)$$

where P_t is the stagnation pressure, P_s is the static pressure of the undisturbed flow and ρ is local air density. However, when Equation (2) is applied to measure the air velocity near the intake cooling air slot of a car front bumper bar it is very difficult to locate the stagnation point and also static pressure positions which reflect the undisturbed flow static pressure, accurately. Consequently, if the pressure difference is measured between two points, one of which is close to the stagnation point and another close to a point which has a similar static pressure as the undisturbed flow and then calibrated against wind velocity in a full-scale wind tunnel, the pressure difference may be used to measure wind velocities during on-road tests. Crosswinds and the resultant yaw angle significantly move the stagnation point and influence the other static pressures around the vehicle. Therefore to calculate the yaw angle, an additional set of pressure taps are needed which are sensitive to yaw angle and velocity.

Experimental Set-up and Procedures

The experiments were conducted in the vehicle test section of the Monash full-scale wind tunnel. The test vehicle (an Australia-made Ford Fairmont) was placed on a turntable, as shown in Figure 6. The main instruments used for this program were two pressure transducers (Honeywell 160PC), the Fluke 2620A Hydra Data Acquisition Unit, and a host-computer. Experimental procedures consisted of the determination of pressure-tap positions on the car bumper bar and the calibration of those pressure taps, including dynamic calibration.

Determination of Pressure Tap Positions - In order to measure both wind velocity and angle it is convenient to drill four pressure tap holes through the bumper bar. For many modern cars, the bumper is a large removable one-piece plastic moulding which is easily machined and transportable between many vehicles and is frequently one of the first components to be manufactured in the production of a new car. A single instrumented bumper bar is convenient for car manufacturers as it can be used on several different test vehicles. For example, it can be used for aerodynamic studies on cooling, drag, noise, etc.

To estimate the best positions on the bumper bar without an extensive pressure tapping exercise, a static pressure tap was made from a thin washer as shown in Figure 7 which was connected to a pressure transducer via 1.5 mm diameter plastic tubing. The pressure transducer measured the difference between local pressures and the pressure inside the vehicle. The pressures at pre-marked spots on the bumper bar (as shown in Figure 8) were measured at zero yaw angle and a tunnel speed of about 19 m/s. Results are plotted in Figures 9 and 10. Figure 9 shows the variation of pressure with position around the sides of the bumper bar and Figure 10 shows that point 3' seems to be the closest point to the stagnation point on the centre of the bumper middle cross-bar. Point 7 was located opposite point 3 and these two points were chosen as they exhibited good repeatability and sensitivity to yaw angle. Point 6' was chosen on the flat surface of the slot to measure local static pressure of the flow through the slot. Finally, pressure taps were inserted in the bumper bar at those four positions and connected to two pressure transducers through 1.5 mm

diameter plastic tubing. One pressure transducer measured the differential pressure between locations 3 and 7 and the other measured the differential pressure between 3' and 6'.

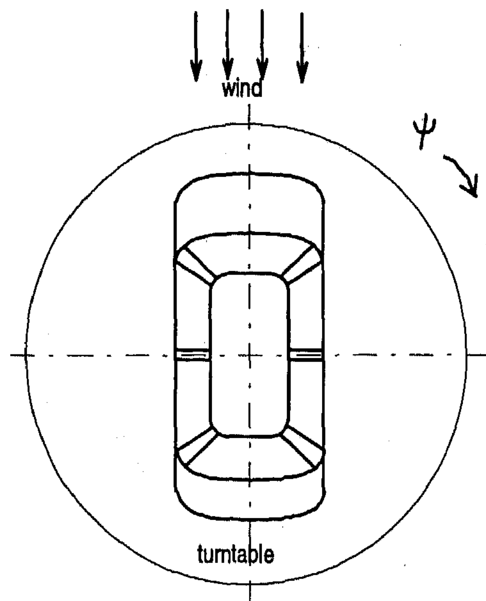


Figure 6. Test car on the turntable at the Monash wind tunnel

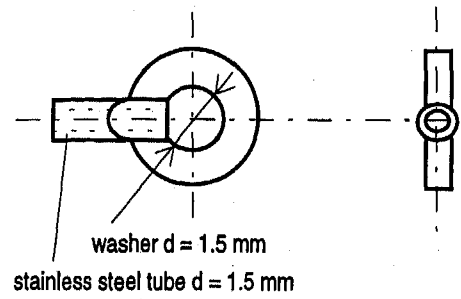


Figure 7. Washer-shaped pressure tap

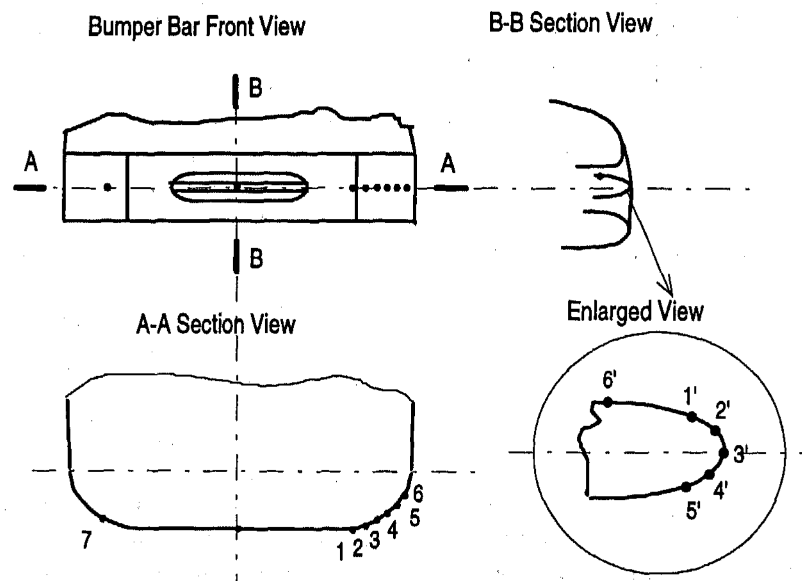


Figure 8. Partial view of a car bumper bar

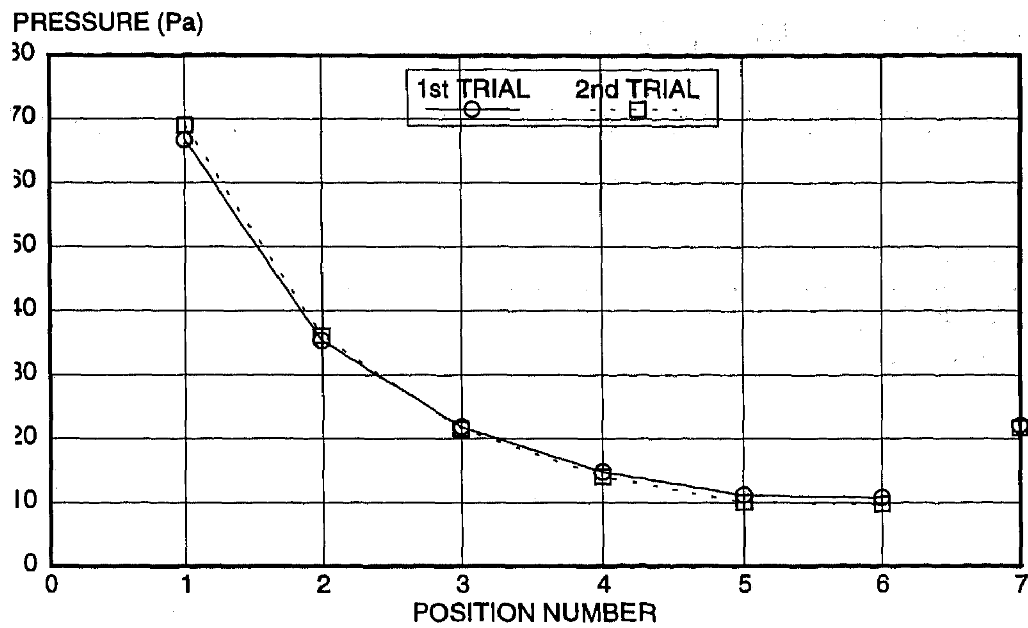


Figure 9. Pressure profile for the determination of side pressure taps

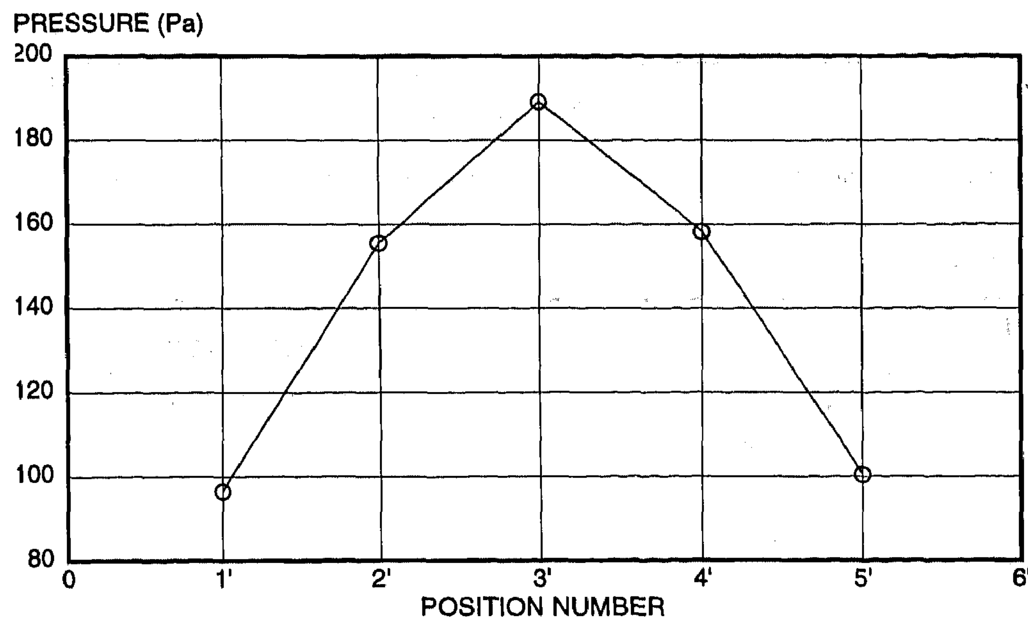


Figure 10. Pressure profile for the determination of centre pressure taps

Calibration of Pressure Taps - The calibration of pressure taps was performed at tunnel velocities of 17.8, 22.5, 28.3, and 33.9 m/s and yaw angles from -15° to +15° with an interval of 3°. The voltage outputs from those two pressure transducers were recorded by the host computer. Then correlations between the differential voltage output of v_c between 3' and 6' and v_s from 3 and 7 with velocity, V, and yaw angle, ψ, were established using polynomials with constants determined via regressions. The polynomials are seen in Equations (3) and (4). Thus in the wind tunnel or on the road, measurements of v_s and v_c can be inserted into Equations (3) and (4) and from those simultaneous equations, V and ψ can be calculated. Comparisons of results from both tests and prediction with regression equations at a wind velocity of 28.3 m/s are plotted in Figures 11 and 12. The figures indicate that both experimental and correlated results match with each other very well. The comparison of predicted results with experimental results is also listed in Table 1.

$$v_c = (a_1 + a_2 V + a_3 V^2) + (b_1 + b_2 V + b_3 V^2)\psi + (c_1 + c_2 V + c_3 V^2)\psi^2 \quad (3)$$

$$v_s = (a_1' + b_1' V + c_1' V^2) + (a_2' + b_2' V + c_2' V^2)\psi + (a_3' + b_3' V + c_3' V^2)\psi^2 + (a_4' + b_4' V + c_4' V^2)\psi^3 \quad (4)$$

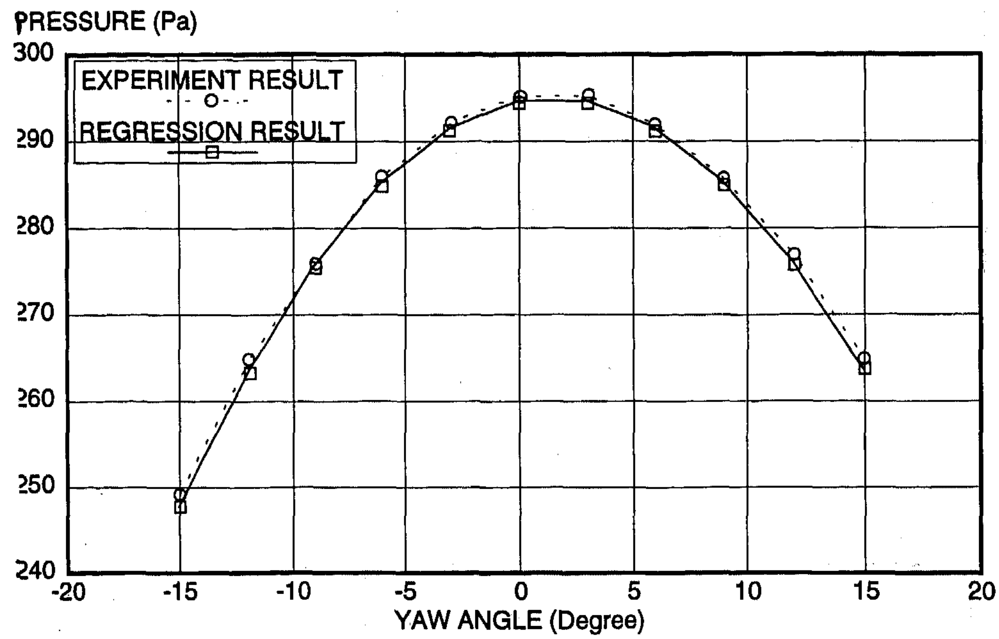


Figure 11. Centre pressure tap readings vs yaw angle at a wind velocity of 28.3 m/s

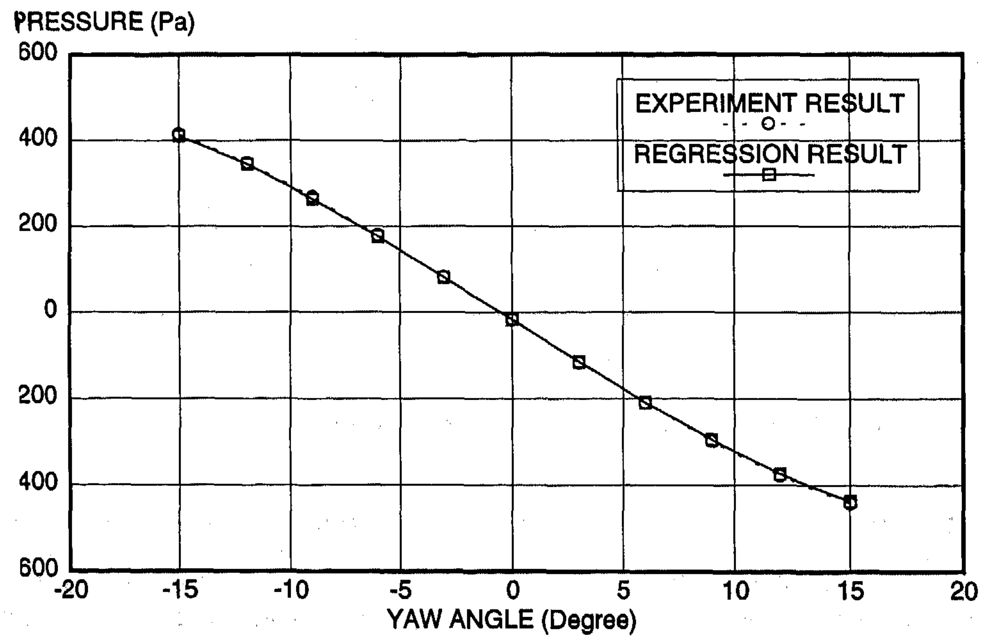


Figure 12. Side pressure tap readings vs yaw angle at a wind velocity of 28.3 m/s

Table 1. The comparison of predicted results with experiment results at a wind velocity of 28.3 m/s

Yaw angle measured	Yaw angle predicted	Error	Wind velocity predicted	Error
(degree)	(degree)	(degree)	(m/s)	(m/s)
-15.0	-14.5	-0.5	28.1	0.1
-12.0	-12.0	0.0	28.4	-0.1
-9.0	-8.0	-1.0	28.3	0
-6.0	-6.5	0.5	28.3	0
-3.0	-2.6	-0.4	28.3	0
0.0	0.0	0.0	28.3	0
3.0	3.6	-0.6	28.4	-0.1
6.0	7.0	-1.0	28.3	0
9.0	10.0	-1.0	28.4	-0.1
12.0	12.5	-0.5	28.4	-0.1
15.0	15.5	-0.5	28.4	-0.1

Dynamic Response - For certain types of testing, a rapid response of a system to relative velocity changes is needed. Examples include coastdown testing (to establish aerodynamic drag and rolling coefficients) and the estimation of effects of turbulence in relative wind on stability and noise. The limiting factor for the system described here is the response of the pressure tubing system. Because pressure transducers (Honeywell 160PC) were mounted inside the test car, about 2 meters of 1.5 mm diameter thin plastic tubing was required. In the resulting dynamic system, a strong amplification of pressure fluctuation around a frequency of 200 Hz may occur and in order to damp out the resonant peak a capillary tube restrictor of internal diameter 0.4 mm and length 5 mm developed according to the procedures of Holmes [2] was located symmetrically about the centre of each plastic tubing. Dynamic calibration of the system, involving step changes to the local airflow around the pressure tappings and monitoring the output on a storage oscilloscope, indicated that the typical time constant was about 40 ms.

ON-ROAD TEST RESULTS AND DISCUSSION

The on-road tests were conducted under following car conditions:

(a) thermostat fully open,
(b) cooling fan disconnected, and
(c) air conditioning off.

On the test day the average atmospheric wind velocity was 6.7 m/s and the direction was 68° from right side of the road (as shown in Figure 13). Tests were conducted at car speeds of 22, 28 and 30.6 m/s. The car was driven in both directions on the road. Typical road test results are listed in Table 2. The results indicate when cars are driving into the wind the engine cooling ability appears to increase and, as expected, and similarly to decrease when downwind.

The on-road test results were also compared with the results from wind-tunnel tests (from Figure 3a). The on-road test results were found to be very similar to the finding from the tests conducted at the Monash wind tunnel, see Figure 14.

Whilst the reduction in cooling is relatively small under the yaw angles tested, a more critical on-road condition is under strong tailwinds which can also generate larger yaw angles and lower relative velocities for the vehicle than tested here. Under such conditions further work is needed, which should be performed with an active cooling fan (if normally fitted to the vehicle) and this may indicate the need to yaw vehicles in cooling tests.

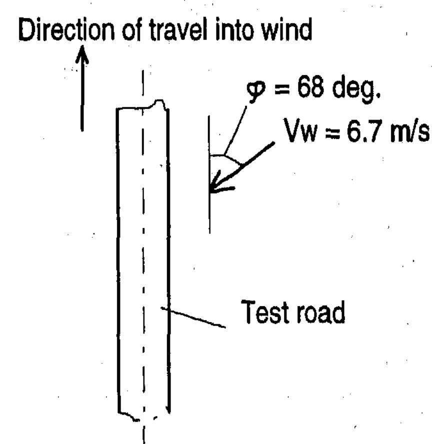


Figure 13. Wind direction during on-road tests

Table 2. On-road test results

Car speed	Headwind			Tailwind		
	Vr	Yaw angle	SD	Vr	Yaw angle	SD
(m/s)	(m/s)	(degree)	(W/K)	(m/s)	(degree)	(W/K)
22	25.3	14	510	20.5	17	401
28	31.1	12	565	26.2	14	515
30.6	33.7	11	591	28.8	12	539

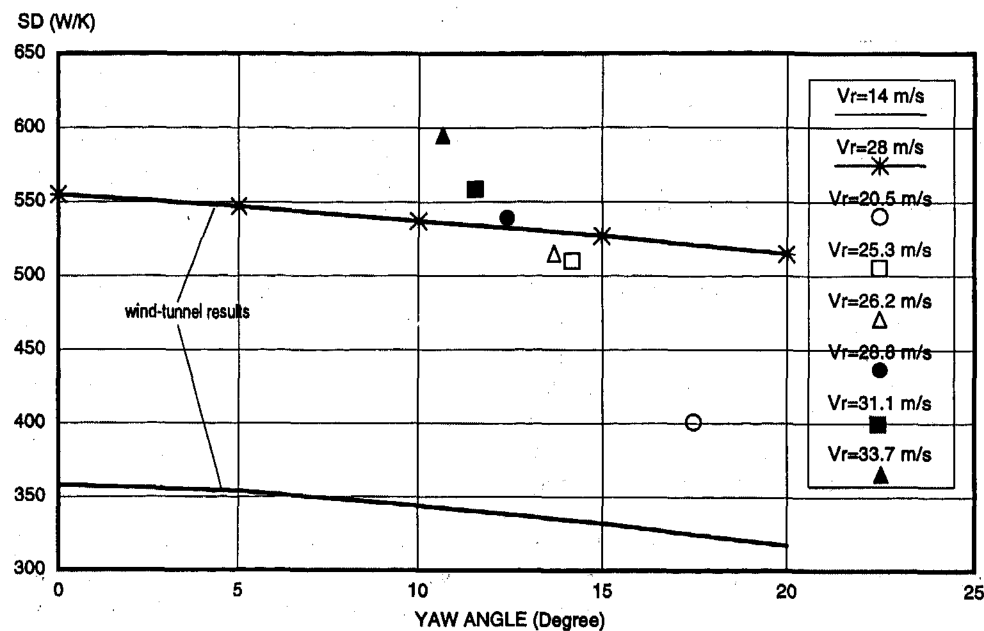


Figure 14. Effect of cross-wind on motor car engine cooling: on-road versus experimental results

CONCLUSIONS

Within the restrictions of the full-scale wind tunnel investigation on three locally-made motor cars and on-road evaluation, the following conclusions can be made:

1. Cross-winds can significantly reduce motor car engine cooling. The reduction increases with increasing yaw angle.
2. On-road tests under yawed conditions gave closely similar results to similar conditions in the wind tunnel.
3. As expected:
   (a) at the same yaw angle, the effect of cross-winds will decrease with the increase of resultant wind velocity; and
   (b) the effect of cross-winds will vary with the motor car front-end configuration.
4. A convenient experimental method to measure wind velocity and yaw angle relative to a moving car was found.

FUTURE WORK

Additional work should be undertaken under conditions that are closer to the maximum capacity of the cooling system. These should include the effects of strong atmospheric winds with significant tailwind components, giving high yaw angles and low relative wind speeds.

ACKNOWLEDGMENTS

The authors would like to thank Trevor Hird for his many detailed comments which are gratefully acknowledged. This program was generously supported by Ford of Australia and Holden. The gracious support of Monash University and RMIT Mechanical Engineering Departments is appreciated. The scholarship and research grant support of the Australian Federal Government is gratefully acknowledged.

NOMENCLATURE AND GREEK LETTERS

a, b, c	- constant, defined in Eq (3)
a', b', c'	- constant, defined in Eq (4)
Ps	- static pressure, Pa
Pt	- stagnation pressure, Pa
Q	- heat exchanger heat transfer rate, W
SD	- Specific Dissipation, W/K
$T_{cold,in}$	- cold fluid heat exchanger inlet temperature, °C
$T_{hot,in}$	- hot fluid heat exchanger inlet temperatures , °C
V	- air flow velocity, m/s
Vr	- resultant wind velocity, m/s
Vw	- wind velocity, m/s
v_c	- voltage output of centre pressure taps, v
v_s	- voltage output of side pressure taps, v
ρ	- local air density, kg/m^3
ψ	- wind direction, degrees

REFERENCES

1. Buckley, F. T., Jr., Marks, C. H., and Walston, W. H., Jr., 1978, A Study of Aerodynamic Methods for Improving Truck Fuel Economy, Final Report of National Science Foundation Grant No. SIA-74-14843, NTIS Publication No. PB. 80-101-272, 381 P.

2. Holmes J. D., 1984, Effect of Frequency Response on Peak Pressure Measurement, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 17, pp. 1-9.

3. Lin C. H., Experimental Simulation on Passenger Car Engine Cooling, Ph.D. Thesis, in preparation.

4. Lin C. H., Saunders J. W. and Watkins S, Increased Productivity - Use of Specific Dissipation to Evaluate Vehicle Engine Cooling, 1997 SAE Congress Paper.

5. Saunders J. W. and Kolodziejczyk E., 1992, Car Radiator Response - Front-end air Management - Tailwinds Sensitivity, Proc. 3nd ATA Int. Conf. on Innovation and Reliability in Automotive Design and Testing, Florence, Italy.

6. Schaub U. W. and Charles H. N., 1980, Ram Air Effects on the Air Side Cooling System Performance of a Typical North American Passenger Car, SAE Paper 800032.

7. Stratton D. G., Stringer R. E., and Taylor S. R. G., 1965, Engine Cooling System Design and Development, Proc. Instn. Mech. Engrs, Vol. 180, Pt 2A, No. 8, pp. 221-235.

8. Vennard J. K. and Street R. L., 1982, Elementary Fluid Mechanics, 6th Edition, John Wiley & Sons, New York.

9. Watkins S., Saunders J.W. and Hoffmann P.H., 1995, Turbulence Experienced by Moving Vehicles. Part I. Introduction and Turbulence Intensity, Journal. of Wind Engineering and Industrial Aerodynamics, Vol. 57, pp. 1-17.

970139

DIGITAL PHYSICS™ Analysis of the Morel Body in Ground Proximity

Andrew Anagnost, Ales Alajbegovic, Hudong Chen, David Hill, Chris Teixeira, and Kim Molvig
Exa Corporation

ABSTRACT

Presented is a study of the air flow over the Morel body [1] in ground proximity which was obtained using a discrete particle method, referred to as *DIGITAL PHYSICS*. The results were computed at several back-light angles and will be compared to experimental observations. Separation and reattachment along the angled section at a back-light angle of 30 degrees, and complete separation at 35 degrees, were both accurately predicted.

INTRODUCTION

The aerodynamic forces on vehicles can have a significant impact on overall performance. One major contributor to vehicle drag is the external shape. Janssen & Hucho [2] were the first to note the importance of the rear window slant on the overall drag when they observed that relatively low slant angles (between 25 and 35 degrees) resulted in a substantial increase in drag. This important result pointed out that improper design can substantially degrade vehicle performance by producing unnecessarily high drag.

The Morel body [1], shown in Figure 1, is one of the most fundamental test cases for determining the capability of an analysis methodology for predicting the basic fluid flow phenomena common in automotive vehicle aerodynamics. It represents a significant test case because it exhibits physical dependencies which are common in vehicle aerodynamics and allows the study of the affect of back-light angle on overall forces. Another body with characteristics similar to the Morel body, and used for the same purpose, is the Ahmed body [3]. The differences between the Ahmed and Morel bodies are minimal. Ahmed's body is slightly longer and thinner than the Morel body and Morel kept the slant section aspect ratio constant at 1.5, whereas Ahmed used a value of 1.9. However, the behavior of the flow over the two bodies is essentially the same.

For the purposes of this study a minor simplification to the Morel geometry was made. Specifically, Morel smoothed some of the edges of his geometry with a small radius round. Since the exact dimensions of this round were not reported by Morel, the geometry used in this study has squared edges like those of the Ahmed body. However, since all the rounded edges are parallel to the flow direction, or on the blunted forebody, their effect on the flow behavior is quite negligible.

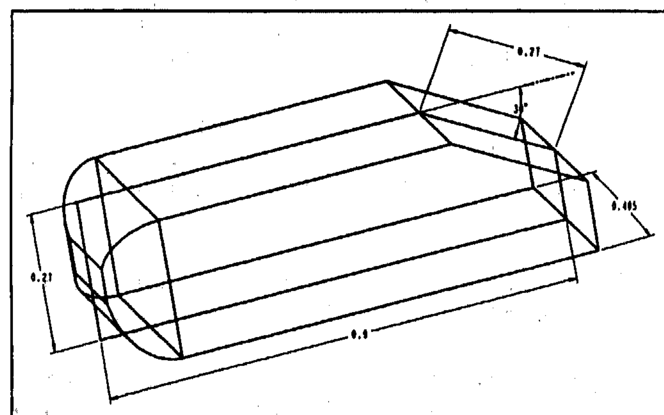


Figure 1. The Morel body (Morel, 1978); all dimensions in meters, back-light angle set to 30°.

The flow structure can be divided into three well distinguished regimes. Figure 2 shows the dependence of the drag coefficient on the back-light angle. The sub-critical regime is strongly three-dimensional. At low back-light angles drag is decreased due to a boat-tailing effect, and the minimum is reached at about 10 degrees. By increasing the back-light angle, two strong vortices at each of the side edges are formed and tend to roll up over the slanted surface. The rolled-up vortices decrease the base pressure and drag starts to increase. At about 25 degrees the flow separates at the upstream edge of the slant and reattaches downstream to form a separation bubble. A slight increase in angle above 30 degrees causes the separated region on the slant surface to join the separation bubble of the base flow. The merging of the separation regions destroys the vortical flow and increases the base pressure

resulting in a considerable decrease in drag. This behavior is then retained for all higher angles.

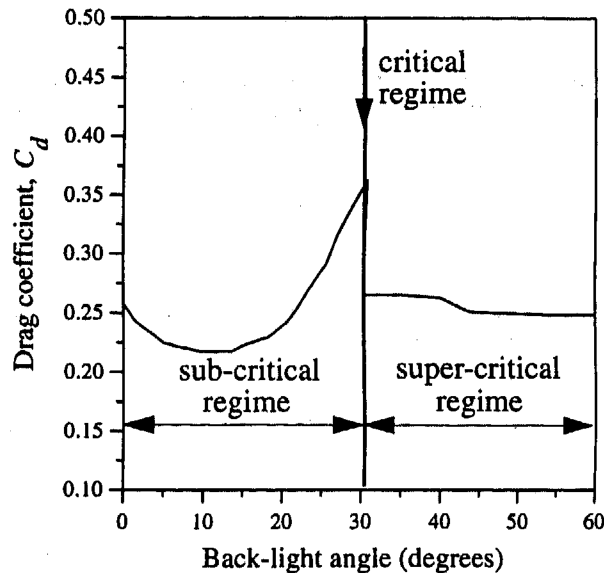


Figure 2. Measured drag coefficient of the Morel body (Morel, 1978).

The complexity of the fluid flow over the Morel and Ahmed bodies has defied many attempts to predict the experimentally observed behavior using traditional computational fluid dynamics (CFD). Han (4) used the k-ε model to compute flow over the Ahmed body. The difference between the predicted and measured drag coefficient was around 30% for slant angles between 0 and 30 degrees. However, the computational approach completely failed to predict the phenomena of vortex breakdown, and the formation of a large separation bubble, for slant angles beyond 30 degrees. There was also an extensive experimental and computational effort at MIRA to investigate the flow over the Ahmed reference model [5]. They obtained excellent experimental results, and an attempt was made to reproduce them using two commercially available CFD codes. Drag coefficient trends were qualitatively predicted, but the values were about 30% too high in the sub-critical and critical regimes. Enormous difficulties were encountered in the super-critical regime. Several mesh resolutions were used with limited success. The complete separation at 35 degrees was never predicted. In addition, the pressure coefficient at the stagnation point had values larger than one. More recently, Bickerton *et al.* [6] attempted to use different turbulence models and mesh structures around the corner at the top edge of the back-light. The results show strong dependence on the gridding technique and question the generality of the solution scheme.

This paper will present results for the Morel body in ground effect which have been obtained from a discrete particle method, a novel extension of lattice gas theory, which yields accurate fluid dynamics using a simple rectilinear lattice; this method is referred to as *DIGITAL PHYSICS*. The results for the Morel body will be described at several back-light angles. The results are remarkably accurate and include the prediction of separation and reattachment along the angled section at a back-light angle of 30 degrees. This phenomenon was observed experimentally by Ahmed *et al.* [3], but has never been computed numerically until the current results. In addition, the phenomena of vortex break down and the subsequent drop in drag was adequately predicted. Included in the discussion will be a presentation of the underlying technology of *DIGITAL PHYSICS*.

THEORETICAL FOUNDATIONS OF *DIGITAL PHYSICS*

DIGITAL PHYSICS extends lattice gas theory to allow application of the method to high Reynolds number flows over external geometries. The lattice gas automata (LGA) method can be viewed as a discrete version of kinetic theory for a dense gas. The mean behavior of the microscopic model of a digital fluid can be shown rigourously to agree with the macroscopic governing equations of fluid dynamics, the Navier-Stokes equations. In the past, lattice gases suffered from the presence of discreteness artifacts and other problems which prevented the quantitative prediction of fluid dynamic lift, drag, and flow structures. The techniques employed in *DIGITAL PHYSICS* remove these artifacts and produce accurate quantitative predictions [7], [8], [9], [10]. The motion of particles is simulated on a face-centered-hypercubic (FCHC) lattice which ensures macroscopic isotropy [11]. The method used has three-speeds with one rate coefficient which moderates the equilibrium in order to remove the lattice artifacts [8]. In addition, the algorithm is unconditionally stable, inherently parallel and computationally very efficient.

The microdynamics of the discrete particle system simulated using *DIGITAL PHYSICS* are described by the masters equation:

$$N_{ji}(\mathbf{x}+\mathbf{c}_{ji}, t+1) - N_{ji}(\mathbf{x}, t) = C_{ji}\{(N_{ji})\} \tag{1}$$

where $\mathbf{c}_{ji}$ is particle velocity, N_{ji} is the number of particles having speed j $(1/2c_j^2=\{0,1,2,...\})$ and moving in direction i; all particles have unitary mass. $C_{ji}\{(N_{ji})\}$ is a sequential series of valid binary collisions that transfer an integer number of particles from one pair of states to the other pair. A valid collision exactly conserves the mass, momentum, and energy at the node. It can be proven that the sequential multilinear collision process monotonically drives the particle distribution towards a stable equilibrium distribution [9], [10]. This equilibrium can be calculated using $C_{ji}\{\left(N_{ji}^{eq}\right)\} = 0$ and has Maxwell-Boltzmann form:

$$N_{ji}^{eq} = R_j^{-1}\exp\left[\left(-\alpha+\gamma\cdot\mathbf{c}_{ji}+\beta\frac{c_{ji}^2}{2}\right)\right] \tag{2}$$

where the coefficients α, γ and β are associated with the conservation of mass, momentum and energy. The energy dependent factors R_j^{-1} are the transition rates.

The hydrodynamics equations are obtained by using the same tools developed from continuum kinetic theory

applied to the discrete lattice gas [11], [12]. The terms in the ensemble averaged microdynamical equation are expanded about a given cell location (x, t) and the 2nd order Chapman-Enskog expansion is performed. The solutions of the resulting equations are found order by order for the deviation from the equilibrium, N_{ji}^{eq}. It can be shown that the macroscopic conservation equations simulated by *DIGITAL PHYSICS* reproduce the well known mass and momentum conservation equations obtained from continuum mechanics to the indicated order in velocity [8]:

$$\frac{\partial\rho}{\partial t} + \nabla \cdot \rho \mathbf{v} = 0 \quad (3)$$

$$\frac{\partial\rho \mathbf{v}}{\partial t} + \nabla \cdot \rho \mathbf{v}\mathbf{v} = -\nabla p + \nabla \cdot \rho\nu\left(\nabla \mathbf{v} + \nabla \mathbf{v}^T - \frac{2}{D}(\nabla \cdot \mathbf{v})\right) + O\left(v^3\right) \quad (4)$$

where,

$$\nu = \left(\frac{1}{\lambda} - \frac{1}{2}\right)T \quad (5)$$

T represents the temperature of the system and D the dimension of the flow. The quantity λ is the relaxation time scale which is used for the adjustment of the viscosity. In the simulations reported in this paper isothermal behavior is assumed. An appropriate energy equation could be derived in a manner similar to that used for equation (4), but it would be contaminated by the presence of discreteness artifacts. As a result, the current methodology is limited to flow Mach numbers below 0.3. The scaling between the lattice and physical units is done through the nondimensional Reynolds number, *Re*.

In order to capture the physics of high Reynolds number flow without resolving all of the relevant physical scales in the boundary layer, a simple wall boundary condition is employed which accounts for the effect of the unresolved boundary layer. The wall boundary conditions are specified through the wall shear stress as:

$$\tau_w = C_f''\frac{1}{2}\rho U_s^2 \quad (6)$$

where U_s is the instantaneous velocity on the surface where the local boundary conditions are applied and C_f'' is a local friction coefficient. Note that this definition of skin friction coefficient is different from standard definitions based on free stream or potential flow velocities. The friction coefficient is calculated by using the local resolution, the global Reynolds number, and the *DIGITAL PHYSICS* equivalent of the law-of-the-wall.

This law-of-the-wall is derived directly from fundamental turbulent boundary layer theory and is derived as follows. First we define a friction velocity, u_τ, such that:

$$\tau_w = \rho u_\tau^2 \quad (7)$$

For sufficiently large Reynolds number (as is the case for the external flow over automobiles), the following well known scaling law is appropriate within the region of logarithmic behavior.

$$\frac{\partial u}{\partial y} = \frac{u_\tau}{\kappa y} \quad (8)$$

where κ is the VonKarman constant($\kappa = 0.4$). Integrating this relation to some height y_s at which the velocity is U_s yields the following:

$$\frac{U_s}{u_\tau} = \frac{1}{\kappa}\left[\ln\left(\frac{u_\tau y_s}{\nu}\right) + B\kappa\right] \quad (9)$$

Where B is a constant of integration with an empirical value of approximately 5. Using equations (6) and (7) we note that

$$\frac{U_s}{u_\tau} = \sqrt{\frac{2}{C_f''}} \quad (10)$$

Hence,

$$C_f'' = \frac{2\kappa^2}{\left(\ln\left(\frac{u_\tau y_s}{\nu}\right) + B\kappa\right)^2} \quad (11)$$

We next eliminate y_s and u_τ from the equation by matching the lattice viscosity to the turbulent eddy viscosity at a height y_s in the logarithmic region of the boundary layer. That is,

$$\nu_T = \kappa u_\tau y_s = \nu_{lat} \quad (12)$$

where ν_T is the turbulent eddy viscosity and ν_{lat} is the lattice viscosity. Hence, we have assumed that there exists a y_s such that the lattice viscosity matches the turbulent viscosity at some point in the logarithmic region of the boundary layer. If this is to be strictly true, then y_s must be of the order of the boundary layer displacement thickness. For Digital Physics, this implies that the near wall voxel size is of the same order. In practice, excellent results are obtained for simulations where the near wall voxels size is much greater than the appropriate boundary layer displacement thickness.

Using the results of (11) and (12) we obtain the basic form of the friction law:

$$C_f'' = \frac{2\kappa^2}{\left(\ln\left(\frac{Re}{\kappa Re_{lat}}\right) + B\kappa\right)^2} \quad (13)$$

where Re is the flow Reynolds number, and Re_{lat} is Reynolds number of the lattice based on the freestream lattice velocity, viscosity, and number of voxels along the characteristic length. This form is strictly valid as long as the near wall voxel size is within some practical limit of the boundary layer displacement thickness.

However, it is well known that adverse pressure gradients decrease the extent of logarithmic behavior within the

boundary layer and that the thickness of the logarithmic region goes to zero at separation. As a result, the value of y_s should decrease in the presence of an adverse pressure gradient. This decrease in y_s can be modeled by altering the viscosity matching of equation (12). Specifically,

$$\nu_T = \kappa u_\tau y_s = \nu_{lat}\left(1 + f\left(\frac{\partial p}{\partial x}\right)\right) \quad (14)$$

where the pressure gradient function is negative in an adverse pressure gradient and zero otherwise. Utilizing (14) yields the final form of the friction coefficient:

$$C_f'' = \frac{2\kappa^2}{\left(\ln\left(\left(\frac{Re}{\kappa Re_{lat}}\right)\times\left(1 + f\left(\frac{\partial p}{\partial x}\right)\right)\right) + B\kappa\right)^2} \quad (15)$$

The results presented in the next section will demonstrate the utility of this simple model for predicting the behavior of high Reynolds number, turbulent, external fluid flow. The accuracy of the results can be directly attributed to the high temporal and spatial resolution which is employed across the entire flow domain.

SIMULATION

The simulation of the flow over the Morel body was performed in a digital wind tunnel similar to that shown in Figure 3. Simulations were run in both fully three dimensional and with a symmetry plane splitting the body in half. There were no significant differences between the two types of simulations. The parameters used in the simulation were the same as the ones reported by Morel [1]. The ground clearance between the Morel body and the floor was 0.0445 m. The inlet (source) was set to a uniform air speed of 58 m/sec. The size of the wind tunnel was set to produce an area weighted blockage of 1.7%; identical to the blockage of the experimental set up. The total stream wise extent of the tunnel was twelve body lengths and the body was placed four lengths from the inlet.

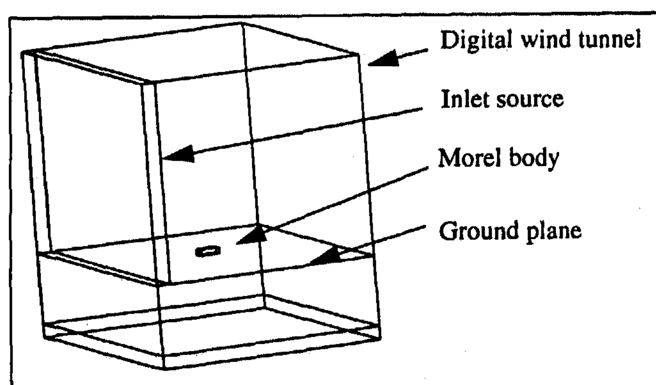


Figure 3. The Morel body in the digital wind tunnel.

The inlet flow was uniform and a modest boundary layer was allowed to grow along the tunnel floor as a result of an imposed friction force on the ground plane.

Since Morel's experimental data had been corrected for the presence of struts, no mounting struts were modeled for the simulation.

The computation advanced time-accurately to a final state which is steady in the mean. Although thousands of time steps are required to reach a quasi-steady state, the computational efficiency of the method ensures that the simulations are completed in relatively short periods of time (a few days) with enough samples to minimize sampling errors. It is interesting to note that the method of collecting and evaluating data from the digital wind tunnel is quite analogous to the procedures and methods employed in sampling data from a real wind tunnel.

The simulation required approximately 8 million voxels (or volume elements) to represent the wind tunnel, and 800 000 surfels (or surface elements) to represent the surface of the Morel body. Figure 4 provides an illustration of the computational lattice. The same lattice configuration was used for each back angle; the only difference being the additional cells required to fill the volume as the back angle increases. Less than one hour was required to prepare the geometry and prescribe boundary conditions. The simulations were run on a four processor SUN UltraSparc server and required approximately 4 days for each back-light angle.

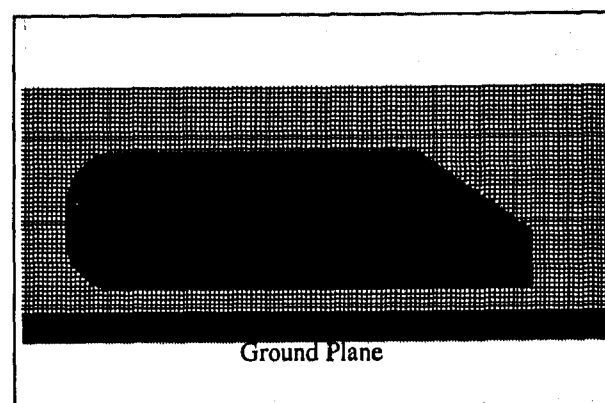


Figure 4. Computational lattice.

RESULTS AND DISCUSSION

LIFT AND DRAG COMPARISON - Drag and lift are major parameters of interest for any flow study over bluff bodies. Simulations were performed for the Morel bodies with back-light angles at 5, 12.5, 25, 30, and 35 degrees. The corresponding drag and lift coefficients as a function of back-light angles are shown in comparison to experiments in Figures 5 and 6.

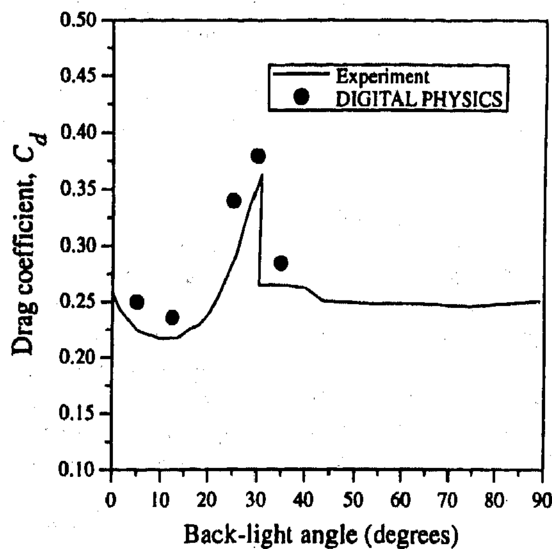


Figure 5. Comparison between the measured and predicted drag coefficient as a function of the back-light angle.

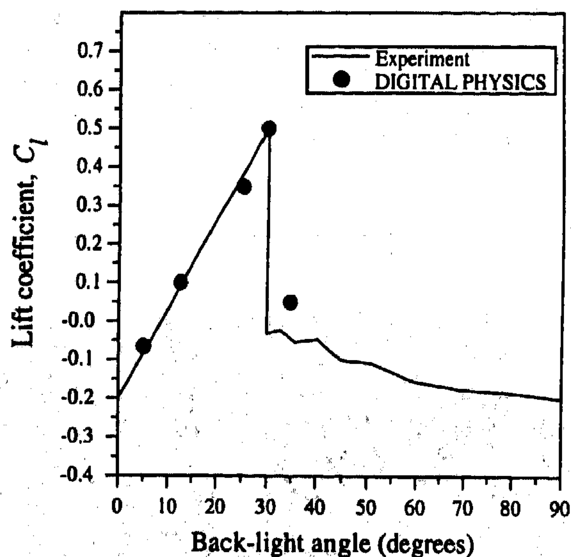


Figure 6. Comparison between the measured and predicted lift coefficient as a function of back-light angle.

The simulation predicted a sharp drop in both drag and lift coefficients for angles between 30 and 35 degrees. This indicates that *DIGITAL PHYSICS* captures the important physics of the flow and is able to duplicate the observed flow structures in the critical and super-critical regimes. There is also good agreement between the measured and predicted values for drag and lift coefficients. The difference varies between 10% (5 degrees) and 5% (30 degrees) for drag. In addition, the change in drag coefficient from one angle to the next is well predicted.

At 30 degrees the simulation exhibited the bi-stable behavior indicated by Ahmed [3]. Ahmed noted that the drag at 30 degrees could be either low or high depending on the initial state of the flowfield. The high drag result presented above persisted for several thousand time steps before suddenly transitioning to the low drag result indicative of high back light angles. This is consistent with Ahmed's observations of the dual nature of the critical angle.

Figures 7 to 8 show the spatial distribution of the pressure coefficient, C_p, on the upper surface centerline of the Morel body at three different back-light angles. Pressure distribution is a convenient indicator of the flow structure. Note that for all slant angles the stagnation pressure coefficient is always unity. At 5 degrees back-light angle there is no separation (Figure 7). However, at 30 degrees the C_p distribution suggests the existence of a separation bubble over the back-light (Figure 8). Previous computational efforts by other authors failed to predict this behavior. Further increase in back-light angle (35 degrees) results in a complete separation at the back (Figure 9). The back pressure recovers considerably and the drag is reduced.

Dips in the C_p distributions at the top edge of the slanted section for angles of 30 and 35 degrees indicates a slight delay in separation which is indicative of a finite resolution effect. Increasing resolution near the start of the back-light angle shifts the separation point further toward the sharp edge and decreases the magnitude of the pressure dip.

Additional finite resolution effects are visible at the pressure minimum over the nose of the body. In this region the surface pressures are "noisy" near the point of maximum velocity. This is due to inadequate surface resolution in a region of large tangential velocity. As resolution is increased along the rounded edge, the spatial oscillation of the surface pressures disappear.

Figure 10 shows the pressure distribution in two regimes in comparison with the experimental data. Regime I is the high drag 30 degree result and regime II is the low drag 35 degree result. The agreement is qualitatively good. It is clearly seen that the flow structure has been captured for the two regimes. The major discrepancy is the absence of the bowing of the pressure coefficient across the slanted section which was observed experimentally for regime I.

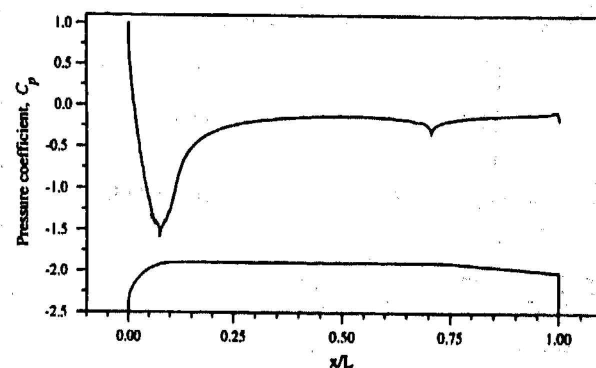


Figure 7. Pressure coefficient at 5 degrees back-light angle (upper surface centerline).

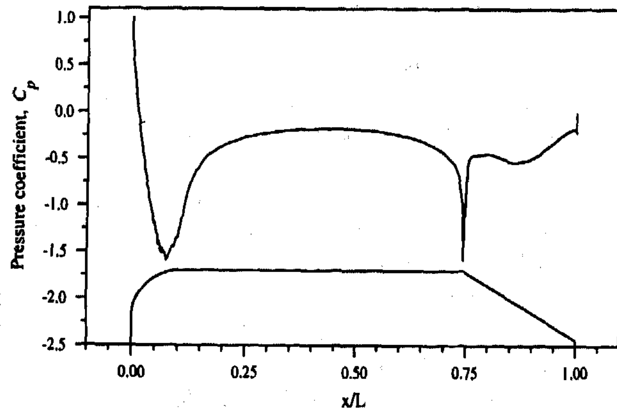


Figure 8. Pressure coefficient at 30 degrees back-light angle (upper surface centerline).

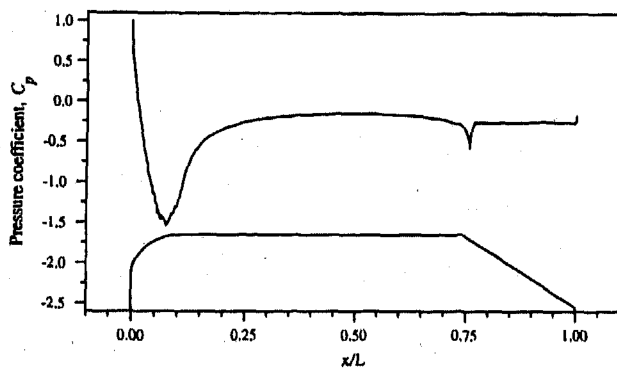


Figure 9. Pressure coefficient at 35 degrees back-light angle (upper surface centerline).

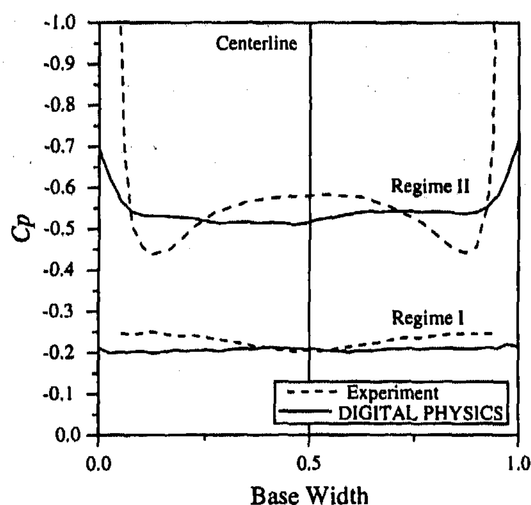


Figure 10. Pressure distribution on the middle of the back light in comparison with the Morel's data.

FLOW STRUCTURE - Typical flow structure in the critical regime can be inferred from Figure 11. The separation bubble which is formed on the back-light surface can be clearly seen. The flow separates right after the top edge and reattaches at the bottom edge of the back-light. The separation and reattachment lines are shown in Figure 12. Beyond 30 degrees the drag drops dramatically and maintains a relatively constant value for higher angles. At 35 degrees there is complete separation at the back (Figure 13) and the final state is free of vortical flow; the resulting rise in base pressure produces the abrupt decrease in drag.

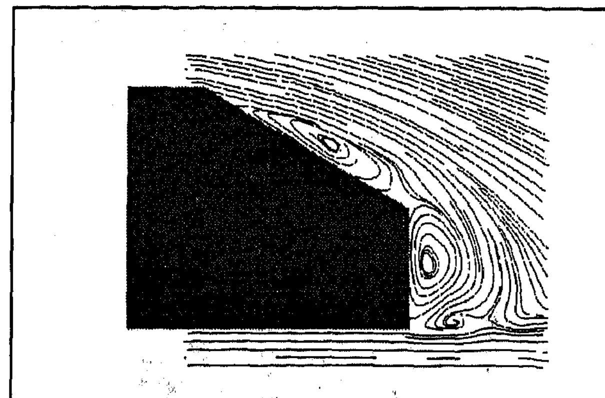


Figure 11. Centerplane streamlines showing separated region; back-light angle = 30°.

Figure 14 shows the comparison between the simulated back flow structures for 30 and 35 degree back-light angles. A clear distinction can be observed between the two cases. At 30 degrees can be observed well formed side vortex and separation babble. The case at 35 degrees clearly shows the complete separation.

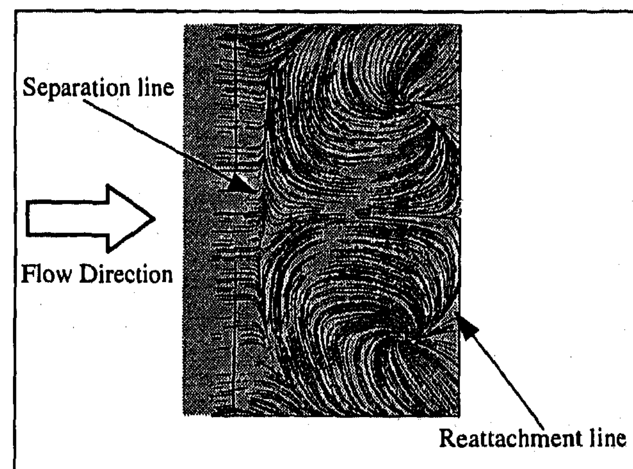


Figure 12. Simulated surface oil flow patterns showing reattachment line; back-light angle = 30°.

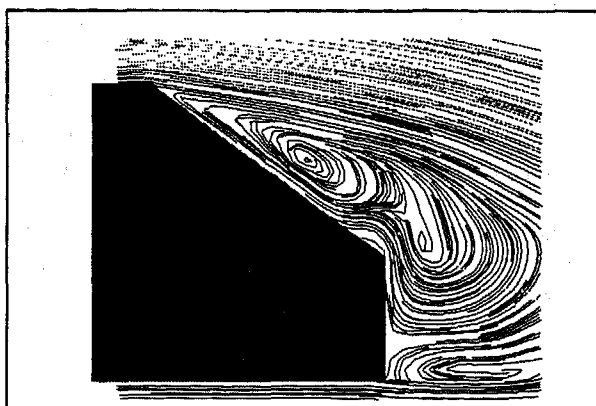


Figure 13. Centerplane streamlines showing separated region; back-light angle = 35°.

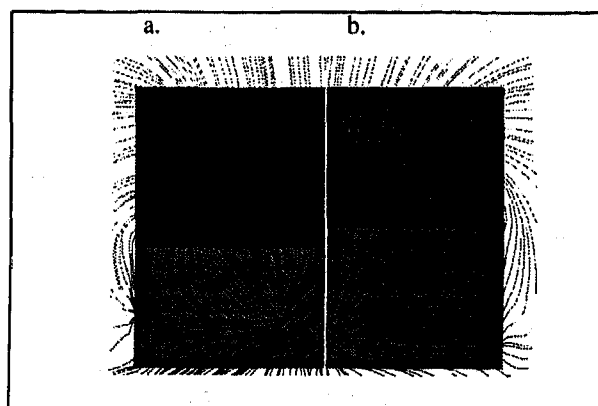


Figure 14. Backplane streamlines; a.) back-light angle = 35°, b.) back-light angle = 30°.

CONCLUSION

The present study illustrated the simulation of fluid flow over the Morel body at several different back-light angles. The computed drag and lift coefficients are in very good agreement with experimental data. Vortex breakdown phenomena were accurately predicted. Furthermore, the obtained flow structures clearly agree with the experimental observations by Morel [1] and Ahmed *et al.* [3]. The agreement between the predictions and measured values clearly demonstrate the capability of *DIGITAL PHYSICS* to simulate high Reynolds number external fluid flows.

ACKNOWLEDGMENTS

The authors would like to thank Jonathan Rothstein and Mark Casella for their considerable effort in the completion of the figures presented in this paper. The authors would especially like to thank Steve Remondi for his contribution to this project.

REFERENCES

[1] Morel, T., Aerodynamics Drag of Bluff Body Shapes Characteristic of Hatch-Back Cars, SAE Paper 780267, 1978.

[2] Janssen, L.J., Hucho, W.H., Aerodynamische Formoptimierung der Type VW-Golf und VW-Scirocco, Kolloquium uber Industrie-aerodynamik, Achen, 1974, Part 3, 46-49.

[3] Ahmed, S.R., Ramm, G., Faltin, G., Some Salient Features of The Time-Averaged Ground Vehicle Wake, SAE Paper 840300, 1984.

[4] Han, T., Computational Analysis of Three-Dimensional Turbulent Flow Around a Bluff Body in Ground Proximity, AIAA Journal, 1989, **27**, 9, 1213-1219.

[5] Baxendale, A.J., Graysmith, J.L., Howell, J.P., Haynes, T., Comparisons Between CFD and Experimental Results for the Ahmed reference model, *RAeS Vehicle Aerodynamics Conference*, Loughborough University, July 1994.

[6] Bickerton, J.E., Graysmith, J.L., Howell, J.P., Haynes, T.L., The CFD Investigation of Flow Separation over a Simple Vehicle Model, ISATA, Italy, 1996.

[7] Molvig, K., Donis, P., Myczkowski, Vichniac, G., Removing the discreteness Artifacts in 3D Lattice-Gas Fluids, in Discrete Kinetic Theory, Lattice Gas Dynamics and Foundations of Hydrodynamics, edited by R.Monaco, World Scientific, 1988.

[8] Teixeira, C., Continuum Limit of Lattice Gas Fluid Dynamics, Ph.D. Thesis, MIT, Boston, Massachusetts, 1992.

[9] Teixeira, C., Theoretical Foundations and Simulation Validations of Digital Physics, NIST Tutorial - Workshop on Lattice Gas - Boltzman Methods, Gettysburg, Maryland, 1995.

[10] Chen, H., Lattice Gas Theory and its Applications, '96 Symposium on Computational Physics, Beijing, China, 1996.

[11] Frisch, U., D'Humieres, D., Hasslacher, B., Lallemand, P., Pomeau, Y., Rivet, J.P., Lattice Gas Hydrodynamics in Two and Three Dimensions, Complex Systems, 1987, 1, 649-707.

[12] Wolfram, S., Cellular Automaton Fluids 1: Basic Theory, Journal of Statistical Physics, 1995, **45**, 3/4, 471-526.

970170

Driver Car Following Behavior Under Test Track and Open Road Driving Condition

R. Wade Allen, Raymond E. Magdaleno
System Technology, Inc.

Colleen Serafin, Steven Eckert, Tom Sieja
Advanced Vehicle Technology, Ford Motor Co.

ABSTRACT

This paper describes the results of an experiment concerning driver behavior in car following tasks. The motivation for this experiment was a desire to understand typical driver car following behavior as a guide for setting the automatic control characteristics of an ACC (Adaptive Cruise Control) system. Testing was conducted under both test track and open road driving conditions. The results indicate that car following is carried out under much lower bandwidth conditions than typical steering processes. Dynamic analysis shows driver time delay in response to lead vehicle velocity change on the order of several seconds. Typical longitudinal acceleration distributions show standard deviations of less than 0.05 g (acceleration due to gravity). Distributions of time headway (following distance divided by following vehicle speed) show conservative behavior by some subjects while others had median values on the order of 1.0 second, and significant occurrences of following between 0.5 and 1.0 seconds.

INTRODUCTION

The driving task can be broadly categorized according to control, guidance and navigation functions that determine vehicle path and speed profile. Control covers psychomotor functions that stabilize vehicle path and speed against various aerodynamic and road disturbances and assist in traffic interactions such as car following. Guidance involves perceptual and psychomotor functions coordinated to follow delineated pathways, adhere to implied speed profiles, interact with traffic and avoid hazards. Traffic control devices, including road markings and signs provide a significant input to the guidance level in path following and speed selection. Navigation involves higher level cognitive functions applied to path and route selection and decisions regarding higher level traffic interactions (e.g., avoiding congestion).

A considerable range of past research has focused on driver steering control and following delineated pathways with more limited attention to speed selection and control. The research discussed in this paper concerns driver headway and speed control during car following. The speed control aspects of car following can relate generally to driver closed loop response to aerodynamic disturbances and roadway geometry. The guidance aspects of car following include perception of lead vehicle headway and relative velocity and closed loop control of following distance. This research was motivated by a desire to understand the dynamics of driver headway control as a guide for setting the automatic control characteristics of Adaptive Cruise Control (ACC) Systems.

BACKGROUND

Past research on car following has dealt with driver headway control models and measurements of driver behavior. Pipes [1] carried out fundamental work in this area, and Bekey, et al. [2] provided a useful summary of past work and further analysis on the general problem of driver lead car following. These early control models assumed that during car following the driver attempts to minimize velocity differences with a lead vehicle (i.e., a well defined stimulus). Recent work on ACC (Adaptive Cruise Control) systems by Fancher, Bareket, et al., (3) and Sayer, Fancher, et al., [4] has produced some data on velocity, range and range rate distributions of drivers during car following.

Data collected by Chandler, et al. [5] has shown that the driver can be characterized according to a closed loop bandwidth (gain or crossover frequency) of 0.37 rad/sec and a time delay of 1.5 seconds in responding to changes in lead vehicle velocity. Note that this bandwidth is a factor of 10 slower than compensatory steering bandwidth, e.g. Allen [6]. Torf and Duckstein [7] also have collected data on driver detection times for several levels of lead car acceleration. They found detection times of 1.9 seconds for a lead car acceleration of 2.5 ft/sec^2, and 2.5 seconds at an acceleration of 1.6 ft/sec^2. By regression analysis they also determined that response time decreased by 0.8 seconds for each increase of 1 ft/sec^2 in lead car acceleration. This may relate to the amount of time required to sense a velocity change (i.e., higher accelerations give more rapid change in lead car range and range rate).

SIMPLE CROSSOVER MODEL — The early speed control models cited above [1,2,5] can be summarized with the classical Crossover Model of McRuer and Krendal [8] as illustrated in Figure 1. Here, the driver responds to velocity errors with throttle commands such that vehicle longitudinal acceleration is proportional to the error. In this model the driver/vehicle response is simplified to a gain (the proportionality constant between velocity errors and longitudinal acceleration), time delay between velocity error and acceleration, and a pure integration to derive vehicle speed from longitudinal acceleration.

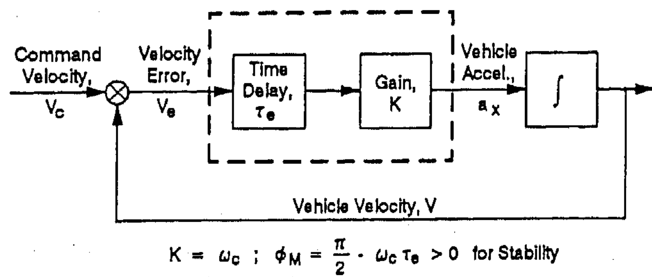


Figure 1. Simple "Crossover" Control Model for Speed Control

The Crossover Model for human control systems is efficient for experimental measurement as it can be compactly characterized by two parameters, the gain or crossover frequency (ω_C) and the time delay (τ_e)[8]. If we take the Fourier transform, e.g. Truxal [9], of the equations describing the Figure 1 system we have the following transfer function:

$$F\left(\frac{V}{V_e}\right)_{j\omega} = \omega_c e^{-j\omega\tau_e}$$

where (1)

$$\omega_c = K \text{ and } j = \sqrt{-1}$$

In the crossover model characterization of manual control system dynamic behavior, crossover frequency is a measure of the system (driver/vehicle) bandwidth, and the product of crossover frequency and time delay is a measure of system damping:

Crossover Frequency = $\omega_c \approx$ Bandwidth

Phase Margin (system damping) = $\phi_M = \pi/2 - \omega_C \, x \, \tau_e$ (2)

In the above equation ω_C is given in radians/second, τ_e is expressed in seconds and ϕ_M is given in radians. When Phase Margin goes much below unity (1.0 radian or about 60°), the system response is oscillatory. Phase Margin goes to zero, the system becomes unstable.

The Crossover Model is a convenient means for simply characterizing manual control system behavior that can easily be derived from FFT transfer function data. Even though we may postulate more complicated engineering control system models, the crossover model provides a simple basis for summarizing the dynamic behavior of a compensatory system. These two key parameters, crossover frequency (ω_C) and phase margin (ϕ_M), provide a means for characterizing the bandwidth and damping of the closed loop manual control system that can be conveniently statistically analyzed to provide a summary of typical behavior and variation across a group of subjects.

In this research it is of interest to simply characterize the dynamic coupling between the following and lead vehicles. The above crossover model parameters can also be interpreted in terms of a closed loop natural frequency by using the Pade approximation for a time delay, e.g. Truxel [9]:

$$e^{-j\omega\tau_e} = \left(j\omega\tau_e +\right)^{-1}$$

where (3)

$$j = \sqrt{-1}$$

Given the above approximation, we can then derive the closed loop transfer function which is given by:

$$\frac{\omega_c / \tau_e}{(j\omega)^2 + j\omega / \tau_e + \omega_c / \tau_e} \qquad (4)$$

where the closed loop natural frequency (ω_n) and damping ration can be expressed as follows:

$$\omega_n = \sqrt{\omega_c / \tau_e}; \zeta_n = 1/2\sqrt{\omega_c \tau_e} \qquad (5)$$

EXTENDED CROSSOVER MODEL — The above simple crossover model can be extended to include maintenance of a desired headway as illustrated in Figure 2. Here, a command feed-forward for headway has been added with the additional parameter a which is low frequency (i.e., long time constant) effect. The extended crossover model controls to minimize velocity errors (V_e) just as with the simple crossover model. In addition, the extended crossover model also develops an additional headway error term (R_e) which is also minimized, but at a much slower rate than the velocity errors.

The extended crossover model has a bandwidth that is approximately equal to the simple crossover model. Because of the higher frequency dynamics of the low frequency a term the phase margin now has an extra component:

Phase Margin = $\phi_M = \pi / 2 - \omega_c x \tau_e - \alpha / \omega_c$ (6)

In the above formula, α is expressed in the units 1/second and can be interpreted as the low frequency bandwidth for headway distance control. This extra phase component allows a simple means for identifying the α parameter in transfer function phase data.

APPROACH

The data processing and analysis used in this paper involves frequency domain procedures designed to identify transfer functions and parameters for crossover models desribed above. These procedures start with time histories of key

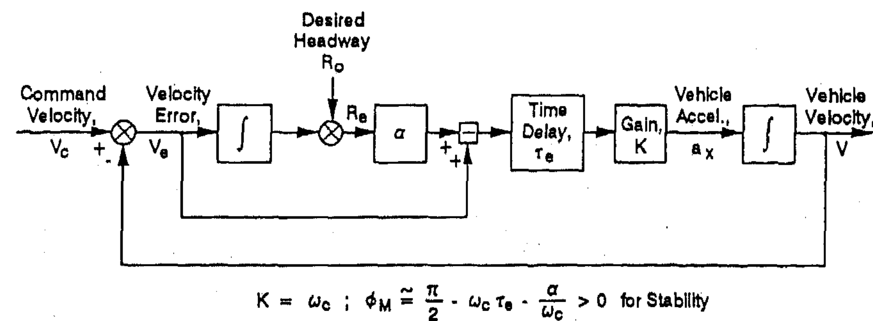


Figure 2. Extended Crossover Model for Speed and Headway Control

variables (e.g., throttle, speed, range) that define the basic human/machine system dynamic response. These time histories are then transformed using FFT (Fast Fourier Transform) procedures to give frequency domain transfer functions between key variables as discussed in Bendat and Piersol [10] and Randall [11]. The FFT procedures are analogous to linear regression analysis procedures that define the relationship between a dependent variable and an independent variable in terms of correlated and uncorrelated components.

With FFT analysis the correlated response between independent and dependent time histories then defines a transfer function that expresses an input/output relationship (e.g. driver throttle response to changes in lead vehicle headway) between key system variables. Given a transfer function, parameter identification procedures can then be used to identify key system characteristics such as driver/vehicle time delay and gain in response to lead vehicle velocity changes.

Measured time histories of key variables were recorded during experimental runs and converted to binary form including scaling constants to allow for maximum efficiency in processing and storage. The time histories were initially manually screened to locate regions of artifact free response, and to screen out segments containing drop outs in sensor response. This was particularly important regarding the range sensor data in both the test track and open highway experiments. The range signal would drop out if range became too large or a combination of lateral position and road curvature caused the lead target vehicle to exceed the limits of the sensor azimuth field of view.

Good sections of time history response were identified for further FFT processing. Additional preprocessing was carried out by applying taper windows to minimize transient artifacts associated with selecting data time windows where the time histories are abruptly started and stopped. The data were then processed by an advanced FFT routine that provided power spectra and transfer functions and coherence functions between selected variables. The coherence function is the equivalent of a linear correlation function, and defines what percentage of an output signal in a given frequency region is linearly correlated with an input signal. Typically the coherence values are used to select transfer function points that are reliable (high coherence) and reject unreliable measurements (low coherence) as indicated in the example transfer function plot in Figure 3.

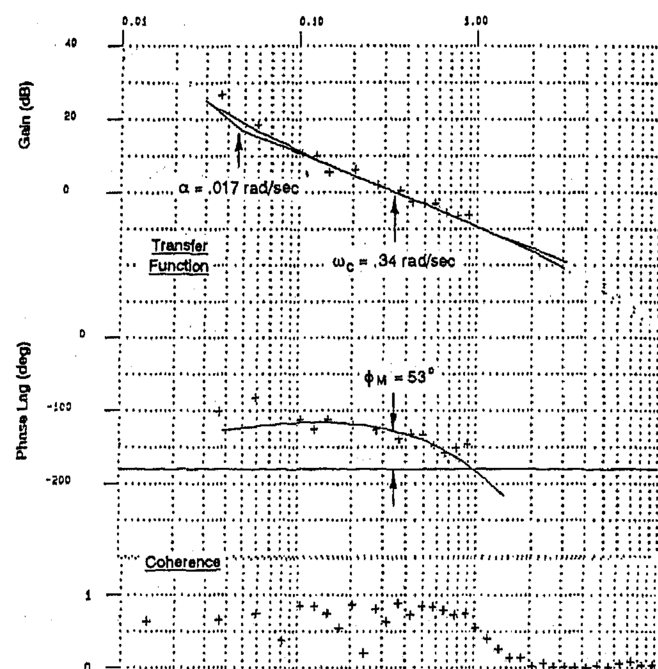


Figure 3. Example Plot of FFT Results and Crossover Model Parameter Identification

Transfer function and coherence data were plotted as indicated further on to highlight reliable transfer function data and to allow for simplified parameter identification. If it is assumed that the driver/vehicle system can be described by the simple gain and time delay model used by previous investigators, then these parameters can be easily identified from transfer function plots such as Figure 3. The gain is identified from the amplitude plot at the zero dB crossover point, and the time delay is obtained from the phase plot in the region of crossover. The α parameter in the extended crossover model is identified in the low frequency region well below crossover based on increased phase shift with decreasing frequency. The Figure 3 transfer function was derived by taking the ratio between the range signal FFT (driver input) and following vehicle velocity (driver/vehicle response) and analytically accounting for the kinematic integration between velocity and range

An example of the vehicle dynamics for longitudinal speed control are illustrated by the transfer function plot in Figure 4. This vehicle transfer function was derived by taking

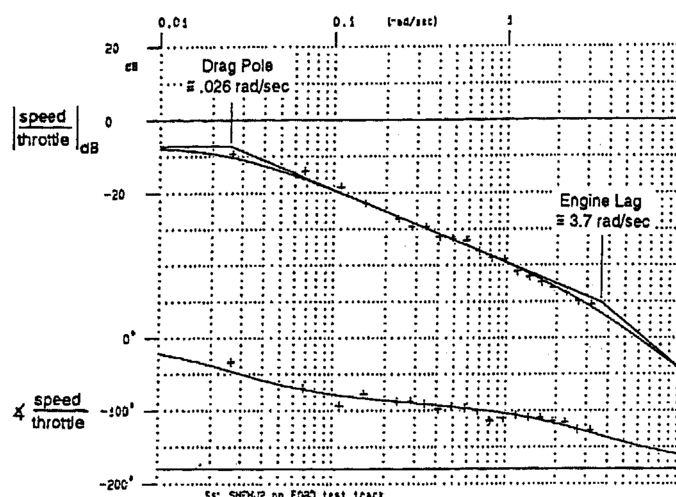


Figure 4. FFT Identification of Longitudinal Vehicle Dynamics Transfer Function

the ratio between the vehicle velocity FFT (output) and the throttle signal (input). The transfer function model fit to the FFT data involves a low frequency pole associated with vehicle aerodynamic drag and a high frequency pole associated with engine throttle response lag. The Figure 4 data show that longitudinal vehicle dynamics can be approximated over a wide frequency range by a pure integration as assumed in the crossover models of Figures 1 and 2.

Distributions of longitudinal acceleration and time headway (lead vehicle range divided by following vehicle headway) were derived directly from any artifact free time history data. In long test track or open road runs, the time series were screened for artifacts (mainly range dropouts), and the distributions were based on the remaining data.

METHODS AND PROCEDURES

TEST TRACK EXPERIMENT — Twelve Ford employees (seven men and five women) took part in the study. The human use research protocol for this study required the participants to be company employees. A reasonable distribution of age and gender was achieved as indicated in Table 1a.

Two test vehicles with automatic transmissions were utilized in the study; the lead vehicle was a midrange four door sedan and the follow vehicle was a midrange sport coupe. The follow vehicle was instrumented with a radar sensor mounted on the center of the front bumper to measure the distance of the lead vehicle with respect to the follow vehicle. Throttle angle (degrees), brake pressure (psi), and velocity (mph) of the follow vehicle were also recorded. Custom-written data acquisition software was used to collect and record data at a frequency of 20 Hz on an IBM PC-compatible computer located in the trunk of the follow vehicle. During testing, an IBM Thinkpad was used to view the data as it was collected. Interference from other vehicles on the test track as well as lap completion times and the time at which climate control tasks were performed (as explained below) were recorded and time-synched with the data file.

The low speed test track at the Ford Dearborn Proving Grounds, shown in Figure 5, was used in the study. The track is 4.3 kilometers long and has six curves varying in radii from 606 to 241 meters. Participants were greeted by the experimenter at their workplace and driven to the test track in the follow vehicle. They were first given a brief overview of the study and then filled out a consent form indicating their willingness to participate in the study and a biographical form which provided the experimenter with information regarding personal characteristics (e.g., date of birth, occupation, etc.) and driving habits (e.g., vehicle driven, annual mileage, etc.).

The test participants were instructed to follow the lead vehicle as they normally would as if they were driving in traffic. They were told that the objective of the study was to evaluate the usability of a climate control system in order to encourage natural car following behavior. The driver of the lead vehicle produced what appeared to be a random speed profile during the study. A majority of the time the lead vehicle driver used cruise control inputs to maintain uniform acceleration and deceleration rates; brake pedal and accelerator pedal inputs were used three times. The speed of the lead vehicle ranged from 30 to 60 mph for all of the laps. The number of speed changes per lap ranged from two to five. An example lead vehicle speed profile for a lap is indicated in Figure 5.

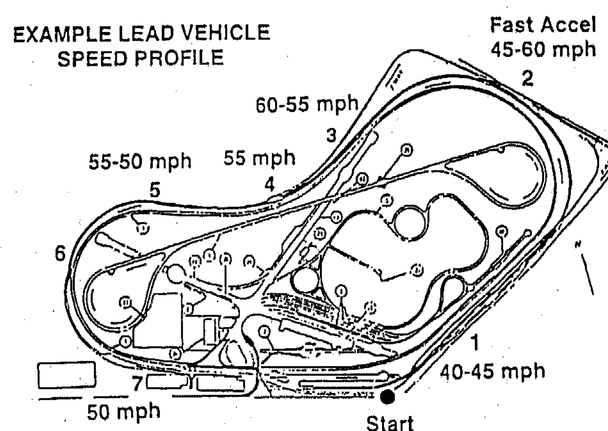


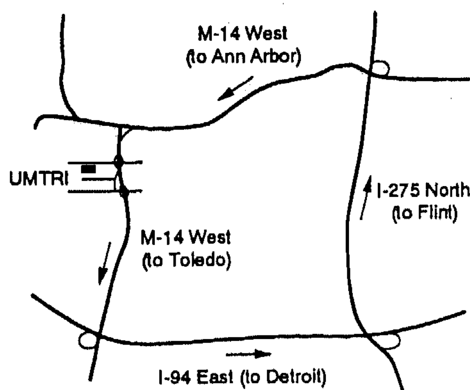
Figure 5. Ford Low Speed Test Track

Each participant drove six laps around the test track. On the final three laps, each participant performed a total of nine tasks using the climate control system as instructed by the experimenter. Examples of some of the tasks included increasing the temperature three degrees, checking the outside temperature, and decreasing the fan setting to the minimum output. At the end of the session, participants were driven back to their workplace. Each test session lasted approximately one hour.

OPEN ROAD EXPERIMENT — Researchers at the University of Michigan Transportation Research Institute (UMTRI) collected manual control data under open highway conditions. The manual control condition was one of three that was investigated; the other two conditions involved driving using conventional cruise control and driving using adaptive

cruise control where vehicle speed and headway is controlled relative to a forward vehicle in the same lane.

Thirty-six drivers (18 men and 18 women equally divided into three age groups (20-30 years, 40-50 years, and 60-70 years), were recruited from a local driver and vehicle licensing office to serve as participants. The subjects drove a 1993 compact imported sedan, instrumented with an infrared headway sensor, around a 88.5 kilometer (55 mile) route in Southeast Michigan that took approximately 50-60 minutes to complete. The highway route and average traffic density conditions are summarized in Figure 6. The study was run



Segment	Average Volume	Lanes
US 23 (South)	44,000 - 56,000	2
I-94 (East)	60,000 - 91,000	2-3
I-275 (North)	45,000 - 112,000	3-4
M14 (West)	43,000 - 70,000	2-3

Source: Michigan Department of Transportation (1993)

Figure 6. Open Road Test Route

during non-peak travel times, and due to the open road conditions was less controlled than the Ford test track study. While a variety of data were recorded, the following six measures were supplied for analysis herein: acceleration, engine throttle position, brake pedal depression and velocity of the instrumented test vehicle, and range and range rate to lead vehicles

RESULTS

DATA ANALYSIS — Data on vehicle speed, longitudinal acceleration and lead vehicle range and range rate were collected on a lap top computer and stored in data files. This data was subsequently analyzed using several time series analysis programs. Figure 7 shows a time series for a good test track data set. This time series was analyzed with FFT procedures which resulted in the transfer function data shown previously in Figure 3. The transfer function estimates represent frequency points that reached a coherence value of .65 or greater in the analysis procedure. The transfer function estimates span the crossover frequency region (.34 rad/sec). The model fit gives reasonable representation of the FFT data, and is consistent with both simple and extended crossover model interpretations.

Figure 8 a) shows a typical time trace of one subject's data during the open road experiment. Range dropouts are apparent in the data which limit FFT analysis. However, time windows were identified as indicated where reasonable FFT analyses could be obtained. Figure 8 b) shows a time expansion of data selected for FFT analysis. Data quantization can be noted here, which illustrates digital resolution in range, range rate and speed. The range drop outs plus quantization place some limitation on the quality of subsequent FFT analysis of the open road data.

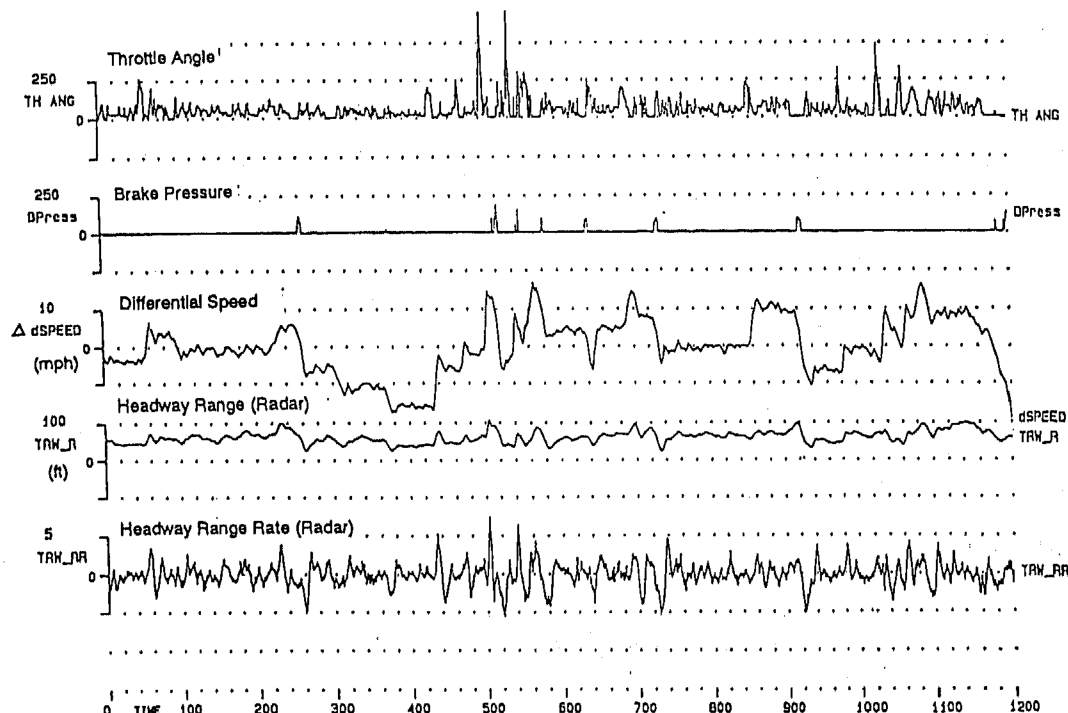
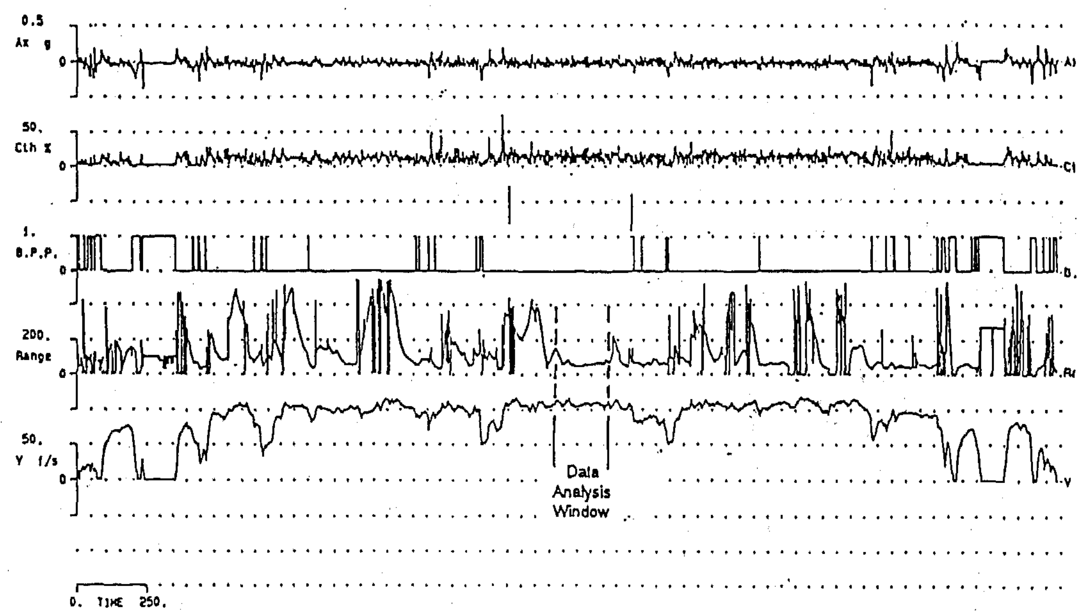
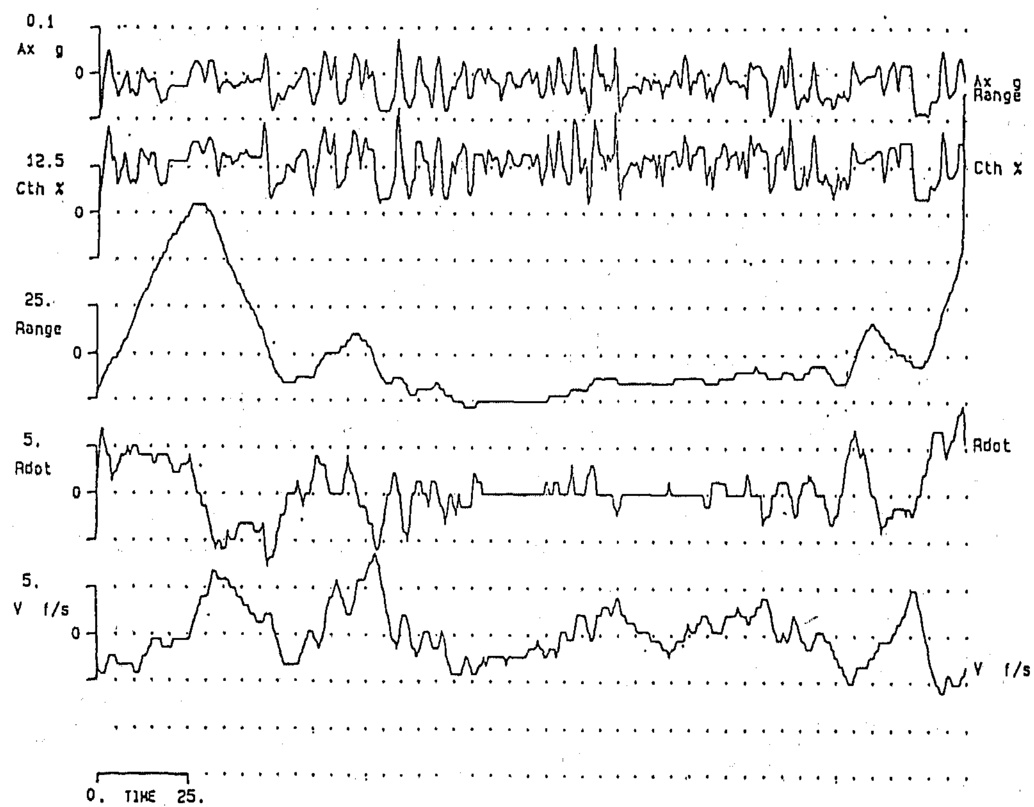


Figure 7. Example Test Track Time History Data



a) Complete Run



b) FFT Segment

Ffigure 8. Example Open Road Time History Data

DYNAMIC COUPLING — Simple crossover model (crossover frequency, phase margin and time delay) data is summarized for both the test track and open road experiments in Table 1. Note that the average crossover frequency is .21 rad/sec and phase margin is 0.7 radians for the test track data and 0.76 radians for the open road data. Equivalent time delays, as derived from the crossover frequency and phase margin, are 3.28 and 4.96 seconds respectively for the track and road data. If the open road time delay data outlier at 11.47 seconds is ignored, the track and road data sets are quite comparable. The phase margins indicate loop closures that range from critically damped to slightly under damped. The closed loop bandwidth and damping defined by Equation 5 were calculated for the track and road data and are also summarized in Table 1. Here we see that the closed loop damping is quite stable across subjects for both sets of data and indicates a slightly under damped condition.

TABLE 1. SIMPLE CROSSOVER MODEL DATA

a) Test Track

Subject		Headway			Simple Crossover Model			Closed Loop Coupling	
Gender	Age (yrs)	R_o (ft)	U_o(mph)	T_h(sec)	ω_c	ϕ_m((rad)	τ_e	ω_n	ζ_m
M	34	76.80	36.80	1.42	0.25	0.87	2.79	0.30	0.60
F	34	71.50	41.60	1.17	0.16	1.05	3.27	0.22	0.69
F	26	100.20	40.70	1.67	0.20	0.87	3.49	0.24	0.60
F	32	70.50	36.00	1.33	0.21	0.65	4.41	0.22	0.52
F	37	72.20	39.90	1.23	0.36	0.73	2.33	0.39	0.55
M	31	91.70	39.60	1.58	0.20	1.05	2.62	0.28	0.69
M	37	105.00	39.00	1.83	0.20	0.44	5.67	0.19	0.47
M	38	100.00	36.00	1.89	0.12	1.05	4.36	0.17	0.70
M	45	80.00	40.30	1.35	0.18	0.87	3.88	0.22	0.60
M	32	61.60	45.20	0.93	0.34	0.92	1.90	0.42	0.62
F	45	68.00	40.60	1.14	0.29	0.61	3.31	0.30	0.51
F	50	117.50	40.60	1.97	0.18	0.73	4.65	0.20	0.55
$\bar{X}$	33.92	84.58	36.64	1.35	0.21	0.76	3.28	0.27	0.59
σX	6.89	17.68	2.58	0.33	0.07	0.19	1.09	0.076	0.075

b) Open Road

Subject		Headway			Simple Crossover Model			Closed Loop Coupling	
Age(yrs)	Gender	R_o (ft)	U_o(mph)	T_h(sec)	ω_c	ϕ_m((rad)	τ_e	ω_n	ζ_m
20-30	F	110.00	54.29	1.38	0.35	0.79	2.24	.395	.565
20-30	F	62.10	73.47	0.58	0.24	1.05	2.18	.332	.691
20-30	M	118.00	62.52	1.28	0.22	0.38	5.39	.202	.459
40-50	F	185.10	66.12	1.90	0.24	0.52	4.36	.235	.489
40-50	F	128.00	71.29	1.22	0.15	1.05	3.49	.207	.691
40-50	F	72.90	60.95	0.81	0.12	0.79	6.55	.135	.564
40-50	M	97.20	59.93	1.10	0.13	0.87	5.37	.153	.598
60-70	F	101.20	53.61	1.28	0.29	0.52	3.61	.283	.489
60-70	M	134.20	63.81	1.43	0.11	0.37	11.47	.098	.445
	$\bar{X}$	112.08	62.89	1.22	0.21	0.70	4.96	.227	.554
	σX	36.21	6.77	0.38	0.08	0.26	2.84	.096	.093

For the test track data there is some tradeoff between driver dynamic behavior and performance. Figure 9 illustrates that driver time delay decreases and crossover frequency increases with decreasing headway time. This result is consistent with more aggressive control for shorter headway times. Multiple regression analysis was performed with headway (T_H) as the dependent variable and crossover frequency (ω_c), phase margin (ϕ_M) and effective time delay (τ_e) as independent variables. The results are summarized in Table 2, which shows that the slopes of crossover frequency and time delay are a significant function of headway time, while phase margin (a measure of closed loop control damping) is not dependent on headway time. These results indicate that driver dynamic behavior is generally related to headway time, with shorter headways leading to more aggressive driver control.

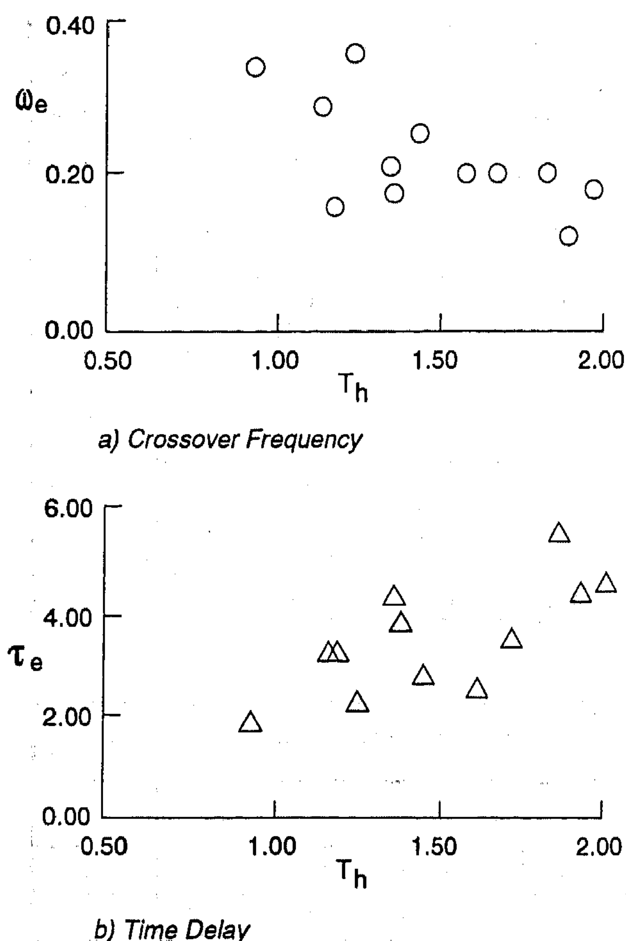


Figure 9. Test Track Crossover Frequency and Time Delay Data vs. Headway Time

The test track FFT data was also used to fit the low frequency parameter α (Figure 2) associated with outer loop hadway control (low frequency data quality was too variable in the road set to permit reliable α identification). Since a accounts for some of the open loop phase lag, the associated driver time delays are somewhat reduced as summarized in Table 3. Here we see that α, which is the equivalent of the bandwidth of the headway or range control loop, is an order of magnitude lower than the inner velocity control loop bandwidth given by the crossover frequency ω_c.

CONTROL AND HEADWAY DISTRIBUTIONS — Distributions of longitudinal acceleration for both the track and road experiments are summarized in Figure 10. Here we see that the acceleration distributions during car following are remarkably similar between subjects and between test track and open road. Maximum values are less than 0.3 g, and standard deviations are less than 0.05 g (1.0 g equals the acceleration due to gravity or 32.2 ft/sec^2). The variation is greater on the open road probably due to the more varied driving conditions. The closed throttle deceleration of modern aerodynamically designed passenger cars is on the order of 0.1 g which probable explains the lower distribution limit. The distributions appear to be reasonably symmetrical, and the upper distribution limit indicates conservative use of engine throttle to match lead vehicle velocity and maintain headway.

Given the longitudinal acceleration described above, the driver can still choose to follow a lead vehicle at any desired distance. Typical time headway distributions are shown in Figure 11. Here we see more variability between drivers in headway time than in their longitudinal acceleration distributions. We also note that the open road drivers exhibit more variability than the test track drivers. Some subjects spend a significant portion of their car following exposure at headway times of less than 1.0 seconds which might be considered fairly risky behavior. The oft stated rule of thumb of a minimum of one car length (approximately 15 feet) for each ten miles per hour of speed (approximately 15 feet/sec) gives an equivalent headway time of 1.0 seconds.

Range and velocity limits were applied to the open road headway data in Figure 11b in order to make the following conditions more consistent with the test track conditions. In Figure 12 we show open road distributions without any range or velocity limits. Here we see that occurrence of large time headways increases dramatically. In Figure 12 we also compare our instrumented vehicle headway distributions with data collected on the highway with loop detectors as presented by Farber [12]. Here we see that the open road instrumented vehicle data is generally consistent with the ensemble data collected at a single location. The instrumented data presented here does show significant variations between drivers, which may be an important consideration when analyzing the effectiveness of crash avoidance systems.

CONCLUSION

We have successfully analyzed open road car following data and validated the general findings of a test track experiment of driver car following behavior. The overall objective of the manual headway control experiment was to measure the dynamic coupling drivers achieve with respect to a lead vehicle. We have a consistent measurement of this coupling under both test track and open highway driving conditions. The open road data appears to be less constrained which may be due to the influence of more uncontrolled variables such as the occurrence of interacting traffic and the drivers' general task demands.

TABLE 2. TEST TRACK REGRESSION ANALYSIS DATA

a) Using 2 independent varialbes 1) τ_e. 2) ϕ_m

Regression Statistics	
Multiple R	0.7679
R Square	0.5897
Adjusted R Square	0.4985
Standard Error	0.2333
Observations	12

Parameter Fit Statistics			
Parameter	Standard Error	t Stat	P-value
Intercept	0.5426	0.0127	0.9900
Variable 1	0.7545	3.5639	0.0060
Variable 2	0.4249	1.4217	0.1888

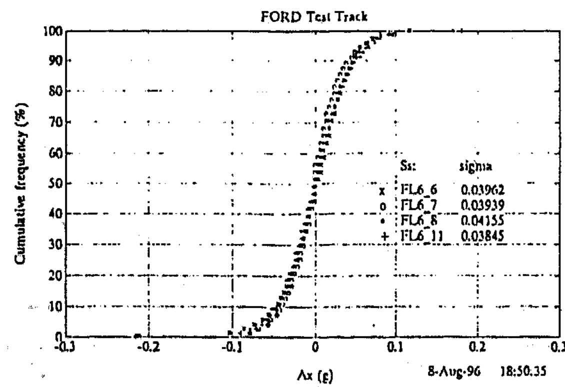
b) Using 2 independent variables 1) ϕ_m 2) ω_c

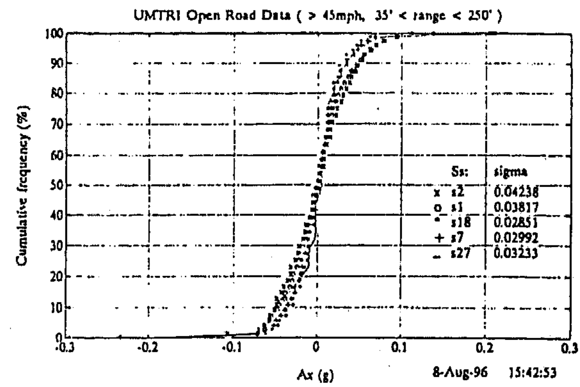
Regression Statistics	
Multiple R	0.7288
R Square	0.5312
Adjusted R Square	0.4270
Standard Error	0.2494
Observations	12

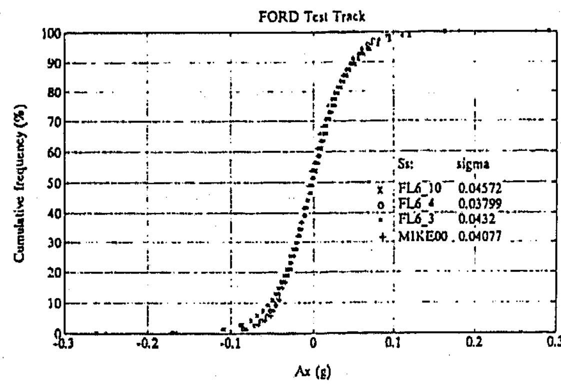
Parameter Fit Statistics			
Parameter	Standard Error	t Stat	P-value
Intercept	0.4702	5.6796	0.0003
Variable 1	0.4061	-1.3336	0.2150
Variable 2	1.0833	-3.1613	0.0115

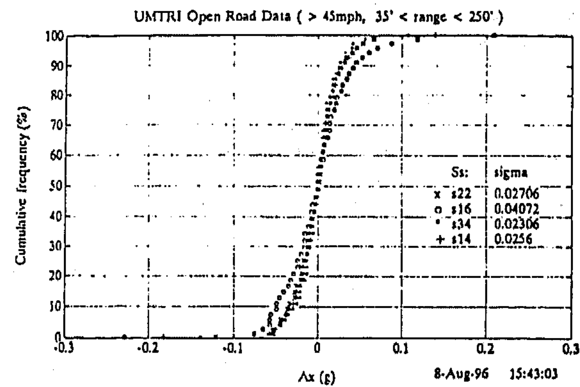
TABLE 3. EXTENDED CROSSOVER MODEL DATA
FROM TEST TRACK EXPERIMENT

Subject		Extended Crossover Model	
Gender	Age (yrs)	α	τ
M	34	0.00604	1.82491
F	34		
F	26	0.00594	1.76988
F	32	0.02459	2.67972
F	37	0.03268	1.55385
M	31		
M	37	0.12298	1.91245
M	38	0.02238	2.10306
M	45	0.00616	1.74523
M	32	0.00793	1.40859
F	45	0.00896	1.97839
F	50	0.00662	2.71962
$\bar{X}$	33.92	0.02	1.52
σ_X	6.89	0.04	0.43



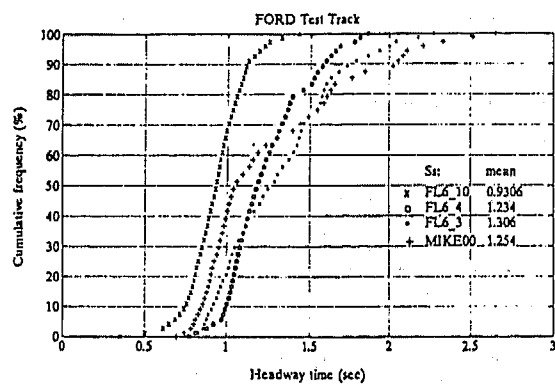


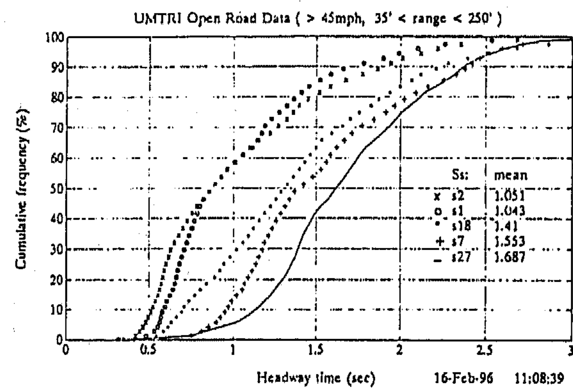


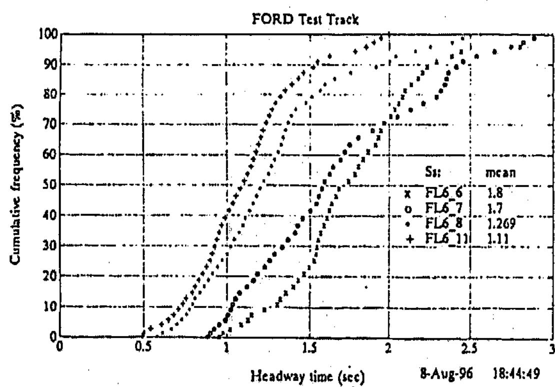


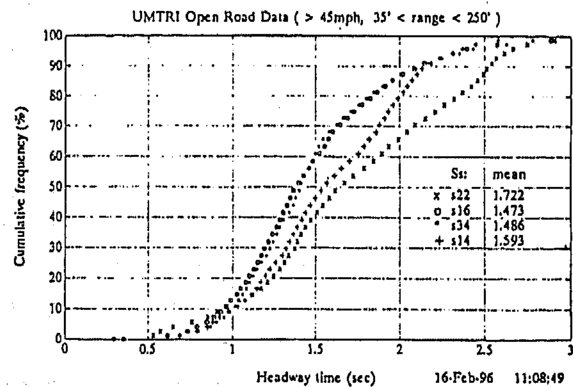
a) Test Track *b) Open Road*

Figure 10. Longitudinal Acceleration Distributions Under Test Track and Open Road Conditions for a Number of Subjects









a) Test Track *b) Open Road*

Figure 11. Headway Time Distributions Under Test Track and Road Conditions.
Open Road Data Limited to Greater than 72 kph (45mph) and Range Interval of 10-80 m (35-250 ft)

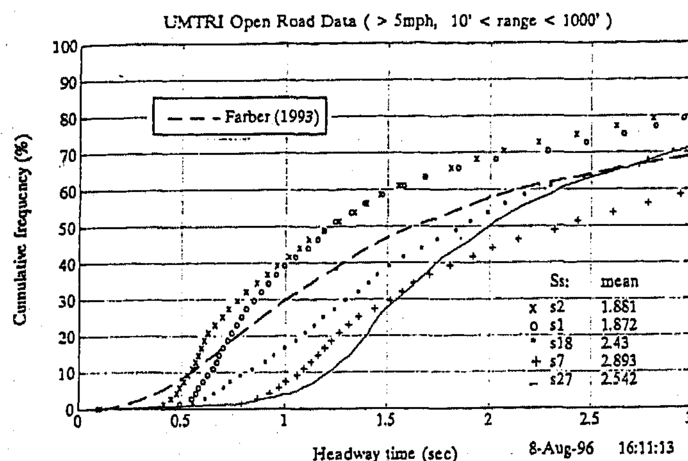


Figure 12. Open Road Headway Time Distributions Under Near Full Range Conditions: Speed Greater than 8 kph (5mph) and Range Interval 3-300 m (10-100 ft)

This study has identified driver headway control characteristics that are quite slow compared to other control processes such as steering for path following and maintaining lane position. The driver exhibits time constants and time delays on the order of several seconds to tens of seconds that limit the stable control bandwidth that can be achieved in maintaining headway. These continuous control time constants and time delays are probably due to processes associated with range and range rate perception, and should not be confused with discrete reaction times associated with braking which are considerably shorter.

ACKNOWLEDGMENTS

This work was sponsored by the National Highway Traffic Safety Administration. Mr. Michael Perel of the Office of Crash Avoidance Research served as the Contracting Officer's Technical Representative. We gratefully acknowledge the open road car following data provided by Dr. James Sayer of the Human Factors Division of the University of Michigan Transportation Research Institute.

REFERENCES

1. Pipes, L.A. (1953), "An Operational Analysis of Traffic Dynamics," *Journal of Applied Physics*, vol. 24, pp. 271-281.

2. Bekey, G.A., Burnham, G.O. and Seo, J. (1977), "Control Theoretic Models of Human Drivers in Car Following," *Human Factors*, vol. 19, no. 4, pp. 399-413.

3. Fancher, P.S., Bareket, Z. (1995), "Evaluation of Human Factors and Safety Performance in the Longitudinal Control of Headway," Proceedings of the Second World Congress on Intelligent Transport Systems, Intelligent Transport Systems World Congress, Yokohama, Japan.

4. Sayer, J.R., Fancher, P.S., et al. (1995), "Automatic Target Acquisition Autonomous Intelligent Cruise Control (AICC): Driver Comfort, Acceptance, and Performance in Highway Traffic, SAE Paper 950970, Society of Automotive Engineers, Warrendale, PA.

5. Chandler, F.E., Herman, R., and Montroll, E.W. (1958), "Traffic Dynamics: Studies in Car Following," *Operations Research*, 6, pp. 165-184.

6. Allen, R.W. (1982), "Stability and Performance Analysis of Automobile Driver Steering Control," SAE Paper 820303, International Congress & Exposition, Detroit, MI.

7. Torf, A.S. and Duckstein, L. (1966), "A Methodology for the Determination of Driver Perceptual Latency in Car Following," *Human Factors*, vol. 8, no. 5, pp. 441-447.

8. McRuer, D.T. and Krendal, E.S. (1974), "Mathematical Models of Human Pilot Behavior," AGARD-AG-188.

9. Truxal, J.G. (1955), *Automatic Feedback Control System Synthesis*, McGraw-Hill, New York.

10. Bendat, J.S. and Piersol, A.G. (1971), *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York.

11. Randall, R.B. (1987), *Frequency Analysis*, Bruel & Kjaer, Denmark.

12. Farber, E. and Paley, M. (1993), "Using Freeway Traffic Data to Estimate the Effectiveness of Rear-End Collision Countermeasures," *The Proceedings of the 1993 Annual Meeting of IVHS America*, 14-17 April, Washington, DC.

970174

CALVIN: Winner of the Fourth Annual Unmanned Ground Vehicle Design Competition

Matthew Caprio
University of Texas

Susan Larkin
Lucent Technologies

John Bay, Paul Johnson, Scott Wenger, Christopher Johnson, and Charles Reinholtz
Virginia Tech

ABSTRACT

The Unmanned Ground Vehicle Competition is jointly sponsored by the SAE, the Association for Unmanned Vehicle Systems (AUVS), and Oakland University. College teams, composed of both undergraduate and graduate students, build autonomous vehicles that compete by navigating a 139 meter outdoor obstacle course. The course, which includes a sand pit and a ramp, is defined by painted continuous or dashed boundary lines on grass and pavement. The obstacles are arbitrarily placed, multi-colored plastic-wrapped hay bales. The vehicles must be between 0.9 and 2.7 meters long and less than 1.5 meters wide. They must be either electric-motor or combustion-engine driven and must carry a 9 kilogram payload. All computational power, sensing and control equipment must be carried on board the vehicle. The technologies employed are applicable in Intelligent Transportation Systems (ITS).

A written design report and an oral presentation are required from each team, and expert judges evaluate these along with inspecting the actual vehicles. Design judging focuses primarily on the design process rather than the implementation of that design in the actual vehicle. The latter feature is evaluated by performance on the obstacle course. The team winning the design contest receives a $1000 award from SAE and is offered the opportunity to present their design paper at the SAE World Congress. The 1996 competition was held at Walt Disney World in Orlando on July 13- 15.

This paper presents the conceptual design of the vehicle and its components. Innovative aspects of the design are highlighted, along with descriptions of the electronics, software, computers, actuators, sensors, and the means of system integration. The steps followed in the design process are described along with the use of computer-aided design. Considerations of safety, reliability, and durability are included. The analyses leading to the predicted performance of the vehicle (speed, ramp climbing, reaction times, etc.) are also documented. Although not a factor in judging, the paper also includes a cost estimate (not counting student labor) for the final product if it were to be duplicated.

INTRODUCTION

CALVIN (Computerized Autonomous Land Vehicle with Intelligent Navigation), shown in Figure 1, is a battery powered three-wheeled vehicle with a computer vision system for line following and ultrasonic sensors for obstacle avoidance. CALVIN was one of two vehicles entered by the Virginia Tech Team in the 4th Annual International Autonomous Ground Vehicle Competition.

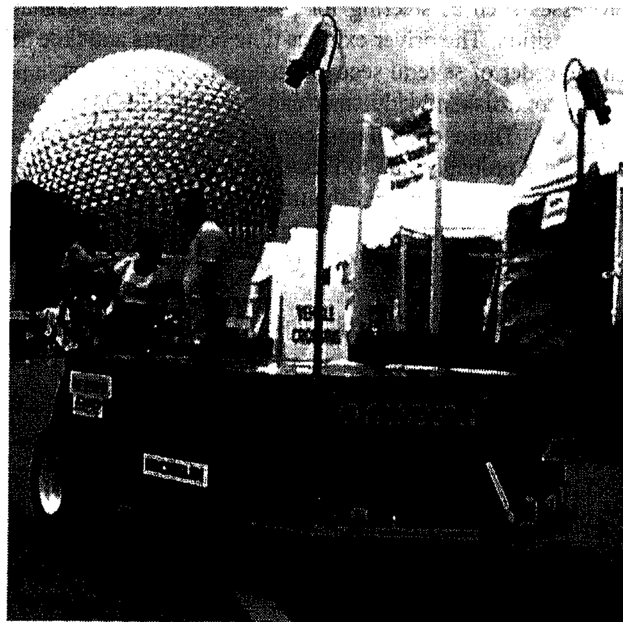


Figure 1 : Overview photo of CALVIN

This is the first year that Virginia Tech has participated in the competition. As part of our preparation, team members reviewed technical papers written by recent competitors and video tapes from the past two competitions. It became clear that the contest presents a formidable task and that a robust, well-developed and carefully tested design would be a basic requirement for success. With this in mind, the team made a concerted effort to pursue simple, reliable, cost-effective designs for the base components and subsystems. This is also the approach recommended by Gifford, et. al., [1995] as part of their winning philosophy in the 1995 competition. Keeping this approach in mind, CALVIN's design incorporates many of the features that have been used successfully in recent competitions. For example, like several past entrants, CALVIN is based on the chassis of a three-wheel golf cart.

Although this platform is larger than the ideal vehicle, it does provide a rugged, stable base with ample space for equipment and payload. Also following past successes, CALVIN uses a computer vision system for line following and ultrasonic sensors for obstacle avoidance. The innovative elements of the design can be found in those areas that caused failure or were recognized as problems in otherwise successful vehicles from past competitions. These innovative features are discussed in detail later in this paper.

PROJECT ORGANIZATION

PROJECT TEAMS - As one of the largest technical universities in the country, Virginia Tech has a talented and diverse student population from which to draw. Although many large projects (such as the SAE Mini-Baja and Solar Car) have been run in the past, this autonomous robotic project generated unparalleled excitement among students and faculty. A small team of volunteers began working to organize the project in the summer of 1995. By the fall semester, the combined team working together on the two vehicles had grown to almost 50 students from across the Colleges of Engineering and Arts and Science. To direct the efforts to build CALVIN, four separate groups were formed to cover Mechanical Design, Sensing & Control, Computer Vision, and Project Organization with each group subdivided to distribute the work. To facilitate group communication, meetings were held twice a week with monthly presentations given by each group.

PROJECT TIMELINE - Organization and scheduling is a key ingredient in the success of a project of this scope and duration. Although many intermediate objectives have been achieved along the way, major project milestones are shown on the general project timeline at the end of the paper.

DESIGN PROCESS OVERVIEW

GENERAL DESIGN DESCRIPTION - The general overall design of CALVIN will be described with reference to the photographs shown as Figures 2 and 3. Figure 2 shows a top view of CALVIN with the weather-protective top cover removed. The general layout of the base vehicle, along with a number of the major components, are visible in this picture. These components include the 24 volt drive motor, timing-belt drive and differential; the 12 volt linear steering actuator with integral feedback potentiometer; the microcontroller; the on-board emergency stop (e-stops); the three 12-volt batteries and the antenna for emergency stop and radio control when not in competition. The microcontroller and electronics mount to a hinged shelf that opens to give access to the pentium-based PC and power supplies. Figure 3 is a front view of CALVIN that shows the five bumper-mounted ultrasonic sensors and the two side-viewing cameras, along with some of the weatherproof Lexan used to encase the vehicle.

DESIGN TOOLS - As with all aspects of our design, the tools used in the creative process were dictated by the overall objective of producing a safe, competitive vehicle. All Virginia Tech engineering students are trained in the use of either AutoCad or Cadkey, and both were used extensively in the design process. The uses of computer-aided design ranged from preparing a test course layout with a shape similar to the sample course shown in the contest rules, to detailed layout of the vehicle systems and subsystems. As an example, AutoCad was used to produce a layout (Figure 4) of the existing base frame of the golf cart. Attachments and modifications were then based on these drawings. Other software used during the course of the project included

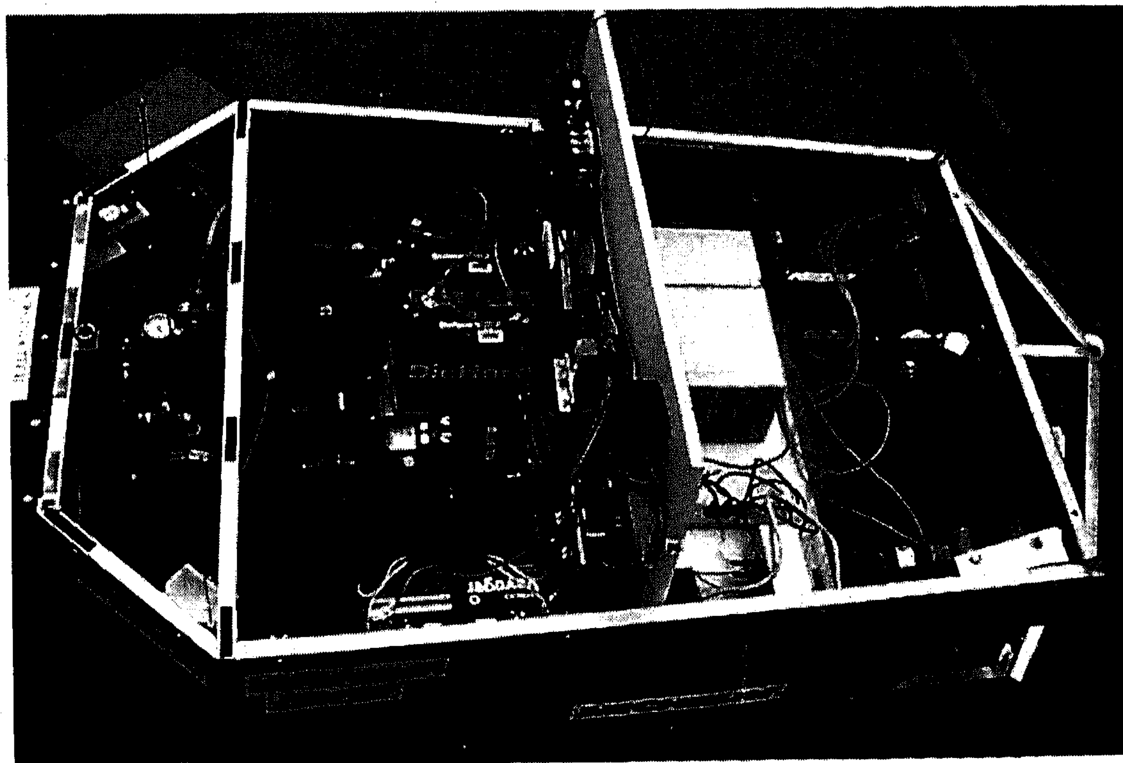


Figure 2: Top view photograph

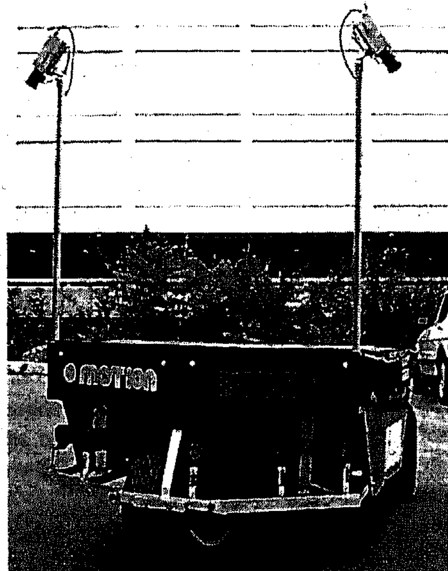


Figure 3: Front view photograph

spreadsheets, TKSolver!, MathCad and Mathematica for general mathematical modeling and MatLab for control and vibration analysis. Specialized software was also written to aid with steering kinematics and vision system modeling.

DESIGN CRITERIA (SAFETY, RELIABILITY AND DURABILITY) - The most critical design criteria were those that involved safety and those necessary to meet contest rules. Beyond these fundamental requirements, a continuing effort has been made to incorporate reliability and durability into the final design. This goal has been pursued both through the careful specification of proven, reliable components and through extensive field testing during every phase of the design. In subsequent sections, brief discussions will help to highlight the effort to produce an inherently safe, reliable vehicle. Safety was also a top priority in all shop work and during field testing.

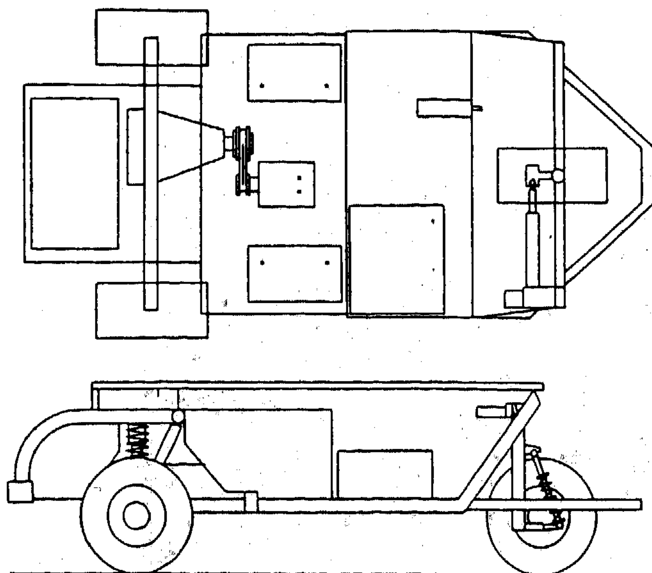


Figure 4: AutoCad Layout

FAIL-SAFE BRAKES - Acceptable braking performance was assured by using the primary components from the original golf-cart braking system. The only difference is that the actuating force normally applied by the user through the foot pedal has been replaced by a fail-safe input force from a pneumatic actuator. An electro-mechanical valve and relay system control the operation of the pneumatic actuator. While the vehicle is in operation, the main power bus closes the relay, powering the valve, thus keeping the brakes disengaged. When the circuit loses power, valve power is lost and the brakes are engaged. This condition occurs when either the onboard E-stop or the remote E-stop is activated.

STEERING ACTUATOR AND CONTROLLER - The rack-and-pinion steering system on the original golf cart was replaced with a 12-volt DC linear actuator with integrated position feedback. The motor in this system is controlled by a Pulse Width Modulation (PWM) controller. The input signal to this controller can be generated by either the microcontroller, in autonomous mode, or a radio control receiver in manual override mode. Manual override mode is used for transportation of the vehicle.

DRIVE MOTOR AND CONTROLLER - CALVIN's primary drive system uses a 0.746 kW, 24V DC motor with tachometer feedback for precise velocity control. Power is delivered to the rear wheels through a 1.8 to 1 timing belt reducer and the existing 6 to 1 golf cart differential. The speed of this motor is controlled using a PWM H-Bridge controller. The controller is capable of adjusting its PWM duty cycle so as to produce an effective continuous output voltage from -24 to +24 volts and a current of 160 amps continuous and 320 amps peak. Like the steering system, the input signal to the controller can be generated by either the microcontroller with active velocity control, in autonomous mode, or a radio control receiver in manual override mode.

ULTRASONIC AND TACTILE OBSTACLE SENSORS - CALVIN is equipped with two types of sensors in order to facilitate obstacle avoidance. First, a fan-shaped array of five ultrasonic sensors is used to locate obstacles potentially in the vehicle's path. In the event that an ultrasonic sensor detects an obstacle, an algorthim is run in order to determine how the vehicle's path should be changed. This routine considers the vehicle's distance to the obstacle, and which ultrasonic sensors in the array detected the obstacle.

The tactile sensors consist of three push-bars on the front of the vehicle. These sensors are used to indicate that the vehicle has collided with an obstacle. In the event that a tactile sensor is actuated, an interrupt routine immediately stops the vehicle and performs a reverse maneuver. The reverse maneuver steers the vehicle based upon which tactile sensor was actuated, in order to relocate the vehicle further from the obstacle. The vehicle then returns to normal operation and continues on its path around the track. The ultrasonic and tactile sensor arrays can be seen in Figure 5.

COMPUTER AND SOFTWARE - The vision system is implemented using a pentium 100MHz PC and an FF1 DSP Frame Grabber from Current Technology, Inc. All vision algorithms were developed using standard C and the frame grabber's C libraries. Standard C is also used to code and compile the software necessary to communicate between the computer and the microcontroller.



Figure 5: Ultrasonic and Tactile Sensor Arrays

The vision system is based on two cameras looking to the left and right of the vehicle, but only one camera is active at a time. When the active camera obtains a series of three consecutive poor images, the system switches to the view on the other side. Image processing is performed on a view of the course directly adjacent to the vehicle. Of this whole image, only two smaller regions of interest (ROI) are considered to be significant. All image processing is then performed on these ROIs only, in order to save computation time. The frame grabber first thresholds the gray-scale camera image to convert the image to binary black and white. The thresholding routine is dynamically adjusted for each ROI to recognize the white paint on grass as the brightest portion of the image. This means that the ROI images used to navigate consist of white line segments on an all-black background. The vision code then uses these two ROI to detect a forward and a rear segment of the course line. In order to discern the line, the routine finds the grouped blobs which it extracts as a segment of the line. The location of the centroids of these blobs are then compared to those found when initializing the vehicle in the center of the course. This information is used to determine the vehicle's distance and angle relative to the line defining the course. The superposition of this zeroth order and first order path information results in a desired steering angle. This angle is then processed to be transmitted to the microcontroller.

The microcontroller integrates the desired steering angle (based solely on vision) with information from the ultrasonic and tactile sensors. Priority must be assigned to obstacles potentially in the vehicle's path; therefore tactile and ultrasonic sensor information is, in some cases, weighted more heavily than the vision in determining the actual steering angle. Consequently, the microcontroller combines the three inputs to determine the path of the vehicle. In effect, the microcontroller performs all onboard navigation.

SENSORS, ELECTRONIC LAYOUT AND SYSTEM INTEGRATION - Figure 6 is a schematic diagram showing the general architecture and integration of the sensors, electronics and computers used on-board CALVIN. Once an image has been processed, the PC sends a steering angle to the microcontroller via a serial connection. If the course boundary line is not clearly detected by the vision system in several successive frames, the PC also sends a command to the microcontroller to switch to the second camera, and the vision system begins tracking the opposite line. Ultrasonic signal processing as well as integration and direct control functions are handled by the Motorola 68HC11-based microcontroller. This microcontroller also executes closed-loop control of the steering and drive motors using pulse-width-modulated signals. A schematic layout of the battery connections and power control elements is shown in Figure 7.

DESIGN PROBLEMS AND SOLUTIONS - Throughout the design process for CALVIN, precise design tools and techniques were used whenever possible. Occasionally, however, an event would arise where trial-and-error methods were necessary. The first arose from the imprecise geometry of the base frame. Calvin began as a golf cart, but came with no factory dimensions. Measurements were made and CAD drawings constructed, but the angles of the frame were difficult to accurately measure, and small differences had a great effect on the positioning of integral parts. This forced a trial-and-error technique on mounting the drive motor, steering actuator and tachometer.

Electrical problems are also unavoidable on a vehicle with so many components. Floating grounds and ground loops are primary contributors to these difficulties. During the layout and wiring of the vehicle, an effort was made to keep the wires as short and organized as possible to facilitate quick troubleshooting. Still care had to be taken to reduce the interference in many of the electrical components, most notably the cameras and the monitors. To do this, ground wires were consolidated, and shielded coaxial wire was used when possible.

PERFORMANCE PREDICTIONS

BRAKING AND SPEED CONTROL - The most critical performance issues were those that involved safety requirements and meeting contest guidelines. Braking and speed control, especially on an inclined surface, are obviously of great importance. An earlier section of this paper describes the design of the braking system and how the foot pedal input was replaced by a pneumatic actuator. Acceptable performance was assured by selecting an actuator that supplied the same brake cable tension as a typical user would supply through the brake pedal in an emergency condition (roughly 500N).

Accurate speed control is important for safe vehicle operation and for assuring constant performance of the vehicle on sloped surfaces. Several past contestants reported control problems that were directly attributed to speed variations. Velocity is controlled on CALVIN using a tachometer attached directly to the output shaft of the drive motor to provide a voltage signal proportional to vehicle speed. This signal is fed back to the microcontroller, which issues a pulse-width-modulated control signal to the motor driver. By modeling the vehicle as a mass moving on a frictionless surface, control system gains can be selected to give stable velocity control with an adequate response time.

RAMP-CLIMBING CAPABILITIES AND SAND PIT PERFORMANCE - Competition rules permit ramps or inclined terrain to have a grade of up to fifteen percent. This information, coupled with the 8.04 km/hr maximum allowed speed and the vehicle's estimated weight, can be used to directly compute the minimum required power of the drive motor (since power equals force multiplied by velocity in the direction of the force). A twenty

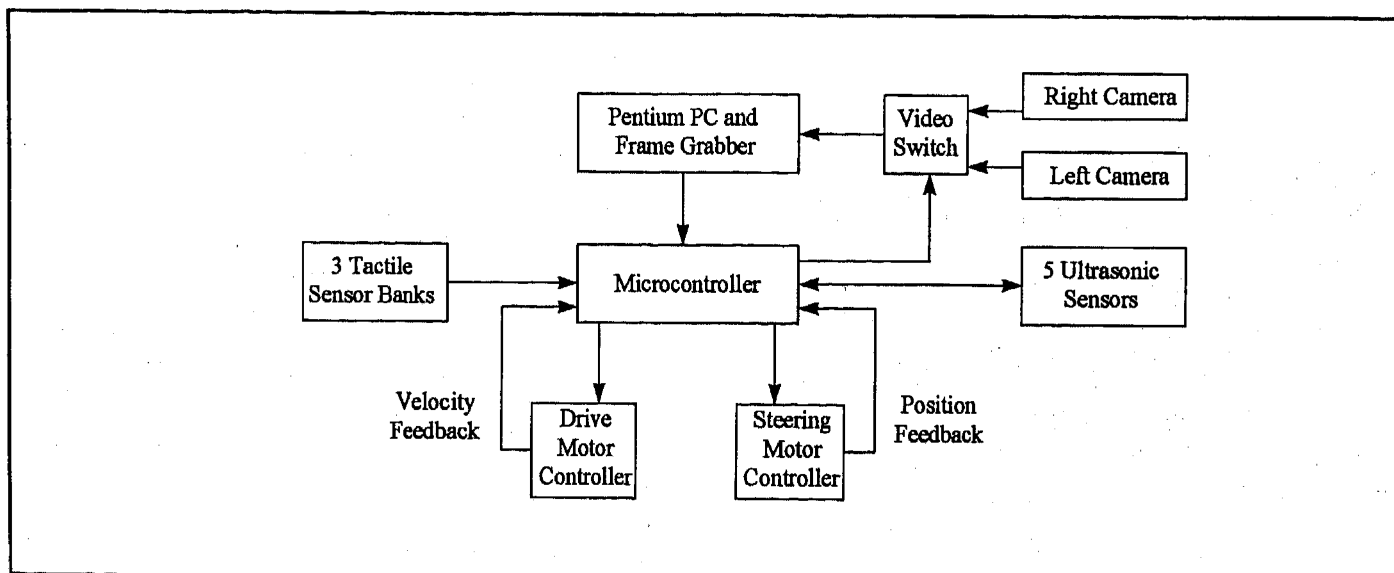


Figure 6: Sensor and control integration

percent grade and a 8.04 km/hr vehicle speed produce a vertical velocity component of 1.61 km/hr. Assuming a 1780 N vehicle weight and no friction results in a required drive motor of about 0.746 kW. Recognizing that steep inclines would be traversed at 1.61 km/hr or less, this was taken to be a conservative estimate. A 0.746 kW continuous, 1.49 kW peak, 24 volt DC drive motor was selected and purchased from a surplus catalog. Using a speed reduction of about 10:1 through a timing belt drive and the golf cart differential, the desired 8.04 km/hr peak speed and ramp-climbing ability were achieved as predicted. Designing a vehicle to traverse a sand pit was a less precise prediction. Goodyear, who donated the vehicle tires, provided assurances that these wide, soft tires would perform well in sand and would not damage the grass portion of the course.

REACTION TIMES AND LINE-TRACKING PERFORMANCE - With any digital control system, the faster the update speed of the controller, the more stable the performance of the controlled device under rapidly changing conditions. The computer vision and ultrasonic acquisition and computation requirements obviously limit the overall update rate of the controller. The team attempted to partially address this issue during conceptual design by separate, parallel processing of the ultrasonic and vision feedback. The vision feedback processing is accomplished directly on the frame grabber card or in the host Pentium 100 MHz personal computer. Operation of the ultrasonic sensors and the associated processing are accomplished in the Motorola 68HC11-based microcontroller. This allows a maximum update rate to the steering and drive motors of about 6 Hz. At peak vehicle speed, this amounts to an update about once every 0.3 m of travel. Through testing, this has been determined to be marginal for robust line tracking; one or two bad images can result in the vehicle traveling out of bounds. Nevertheless, within the constants of the existing equipment, this is a reasonable update rate. At competition, we intended to run CALVIN at peak speeds of about 3.22 km/hr to maintain stable operation.

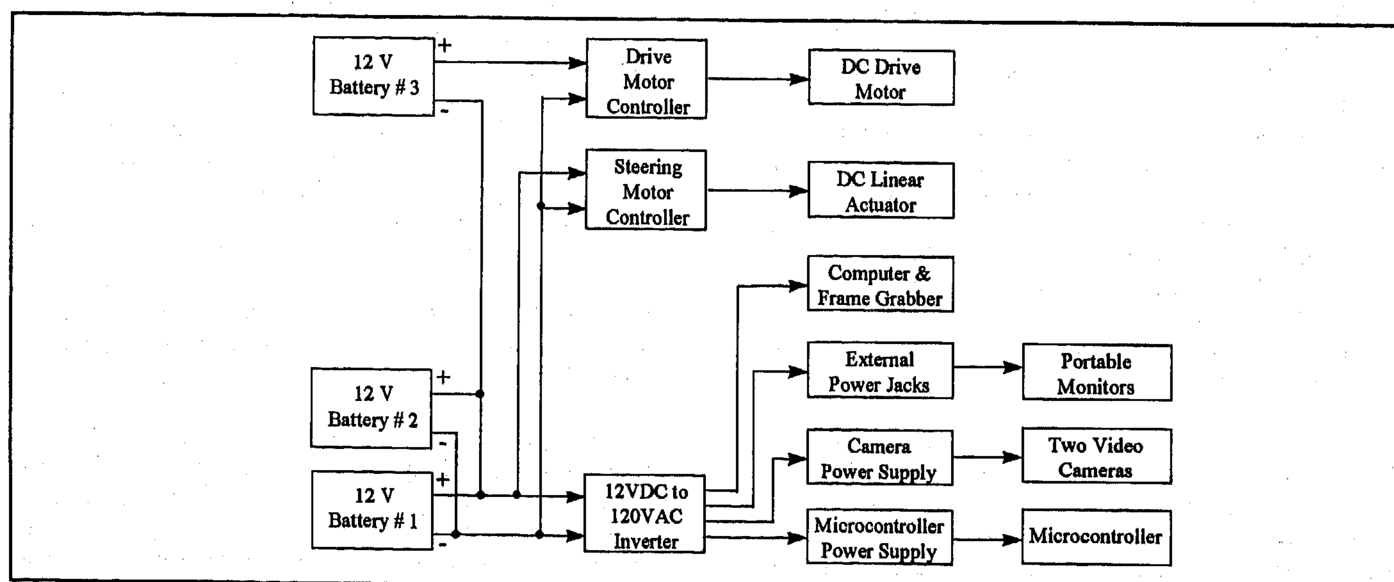


Figure 7: Power supply diagram

INNOVATIVE ASPECTS OF THE DESIGN

The design philosophy used in developing CALVIN did not specifically include innovation as a design objective. Our primary objective was to design the most competitive possible vehicle under the constraints of the governing contest rules and limited financial resources. Nevertheless, the resulting design is believed to be innovative in several significant respects. In all cases, however, the innovative features are intended to address problems or shortcomings reported in similar vehicles from previous competitions.

A simple example of this was the design of a system that uses only three 12-volt deep-cycle batteries as the source for all on-board power. Power for the PC and peripheral equipment was supplied using an 400 Watt inverter to generate 120 volt AC power. Other DC power requirements were met through small on-board power supplies or through series connection of the batteries. This power supply arrangement was previously shown as Figure 7. While this may seem to be a minor issue, a number of previous teams, including the first-place 1995 Colorado team [Gifford, 1995], cited the maintenance of several dissimilar batteries as a significant disadvantage. The University of West Virginia team also cited the weight penalty and potential danger associated with their eight- battery, 96-volt DC bus system [Banta, 1995].

A second, and more significant, innovative feature of CALVIN is the use of two independent cameras for line tracking. This innovation is an attempt to address, in a cost-effective way, the problems noted by 1994 and 1995 competitors with single-camera systems [Murphy, 1995, Gifford, et. al., 1995]. Originally, the team hoped to use two cameras and two independent frame grabbers. With such a system, each camera could continuously track one of the two course oundary lines. Unfortunately, two frame grabber cards or a single multi-input frame grabber proved to be too expensive, and an alternative innovative solution was pursued. The end result was to use a simple relay-actuated video switch to change camera views when the course boundary lines disappeared or became difficult to detect. Using a line-integrity check based on the size and grouping of white image areas, the computer vision system makes three attempts to find a line with the current camera. If two successive attempts fail, the vision software sends a command to the microcontroller to switch to the camera on the other side of the vehicle and begin tracking the opposite line. Based on a video update rate of at least six frames per second (bad images can often be rejected quickly, resulting in a higher rate), and a maximum speed of 8.04km/hr, the vehicle will travel no farther than 0.244 m in this time.

A third innovative feature of CALVIN's design is the use of a radio controller to generate the same PWM signals as the microcontroller for the steering and drive motor controllers. This allows simple switching between autonomous and manual modes, which greatly facilitates vehicle setup and testing. During normal operation, all user interaction with CALVIN is through the radio controller or through a single weather-resistant control panel mounted on the left rear quarterpanel of the vehicle. This control panel is shown in Figure 8 below. While this is again a simple feature, it is a clear improvement over the interfaces used by vehicles in previous competitions, many of which required awkward interactions including manual pushing for placement and setup.

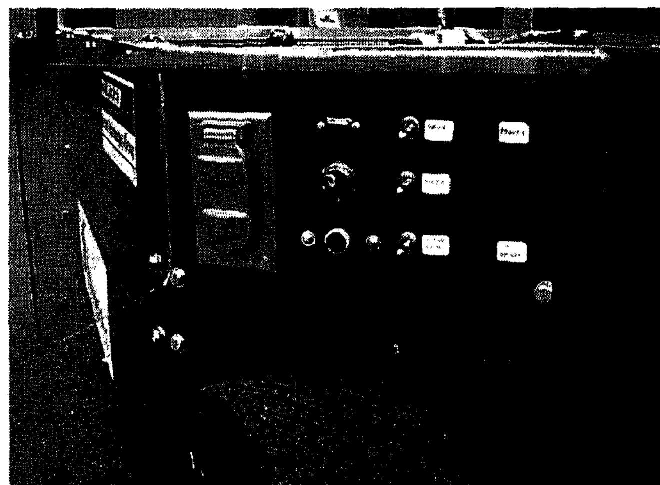


Figure 8: The control panel

A fourth innovative feature of CALVIN is the separation of data processing between the PC and the microcontroller. Since data processing speed has historically limited other vehicles maximum speeds [Cheok, 1994, Nagy et. al., 1995], a system was designed to process data in parallel. By processing closed-loop velocity and steering control, as well as ultrasonic and tactile sensor data on CALVIN's microcontroller, the PC is used solely for the computationally intensive vision algorithms. This separation of data processing greatly reduces the total processing time, therefore increasing the maximum sustainable vehicle speed.

VEHICLE COST AND FUNDING SOURCES - The table at the end of the paper lists the ready-made components that were purchased or donated. Separate columns list both the retail cost and the purchase price of the component. Note that many items were either donated or sold to the project team at a substantially reduced cost. Also note that all fabricated components, such as mounting brackets, sensor housings, framework, bumper and the Lexan outer shell, were built by members of the design team in the student shop of the Mechanical Engineering Department. In addition, team members did a great deal of electronic design and circuit fabrication. It is estimated that approximately 5000 person hours have been invested in this project. This includes time spent on activities such as project organization, fundraising, publicity, and research.

COMPETITION RESULTS

The 1996 Unmanned Ground Vehicle competition held at Walt Disney World, Epcot Center July 13-15, was composed of two sub-competitons. The first was the SAE Design Competition, which evaluated the written report, an oral presentation and a static viewing for each vehicle.

The second competition was the obstacle course race. All vehicles were given three attempts at completing the course, and the vehicle to complete the course in the shortest amount of time, or to proceed the furthest in any one run would be declared the winner of this competition.

Virginia Tech's CALVIN vehicle was declared the winner of the SAE Design Competition. Virginia Tech' s other vehicle, BOB, scored third place in the SAE Design Competiton. The

winner of the obstacle course race was Ohio State University's vehicle. Virgina Tech' s obstacle course race results were 13th and 6th for CALVIN and BOB respectively.

ACKNOWLEDGMENTS

Thanks to the many corporate sponsors and to the Departments of Electrical and Mechanical Engineering for making this project possible. Special thanks to Mr. William Agnew of SAE for carefully reviewing this document before final submission.

Part Name and Description	Project Cost ($)	Retail Cost ($)
Base Vehicle - Gas Powered E-Z Go golf cart (used)	500	500
Drive Train - 12 VDC 0.764 kw Thermo King Motor (surplus)	100	100
Drive Train Components (belts, pulleys, etc.)	75	75
Computer - Pentium 100 Mhz PC	1125	1125
DSP Frame Grabber (Current Technologies)	donated	900
Motorola 68HC11-based Microcontroller (Coactive Asthetics)	125	250
Assorted Electronic Components and Cables	150	150
Polariod Ultrasonic Sensors (six)	350	350
Linear Actuator - 12 VDC 6" stroke (Motion Systems)	donated	300
2 Cameras & Lenses (DEI / Professional Security Alliance)	400	1200
PWM Motor Controllers (VANTEC)	450	950
Aluminum for Frame and other structures (R.J. Reynolds)	donated	200
Batteries (Sears)	150	300
Air Tank (Steel-Fab)	donated	100
400 W Inverter (Tripp-Lite)	donated	100
Lexan for Outer Shell (Piedmont Plastics)	donated	350
TOTAL	3,425	6,950

Vehicle Cost Breakdown

May 1995	Aug. 1995	Oct. 1995	Dec. 1995	Jan. 1996	Mar. 1996	May 1996	July 1996
Autonomous Robotic Vehicle Project Introduction at Virginia Tech Preliminary design team assembled	Golf cart acquired for base vehicle Academic year begins, large project team organized	Base vehicle operational and prepared for modification	Conceptual design/frame modifications completed Fail-safe brake system implemented	Spring semester teams organized System components ordered Drive and steering systems tested	Computer vision and ultrasonic sensors tested Automatic drive control specified	Formation of summer teams Systems integration completed Testing and subsequent modification begins	Inter-team competition between CALVIN and BOB Design paper submitted Departure for Epcot 7/12 4th Annual UGR competition in Epcot

Project Timeline

REFERENCES

Murphy, Robin R., An Artificial Intelligence Approach to the 1994 AUVS Unmanned Ground Robotics Competition, IEEE International Conference on Systems, Man, and Cybernetics, Oct., 1995, Vancouver, B.C., Paper 0-7803-2559-1/95.

Gifford, K. K., Deeds, M., Van der Hoek, A., Henning, F., Freeman, J., Haussman, G., Kuzminsky, S., Stoller, S., RoboCar: Software, Hardware, and Mechanical Design Issues for the University of Colorado Autonomous Rover Vehicle, SPIE Vol.2591, Sept., 1995.

Banta, L. E., "Undergraduate Robot Design at West Virginia University," Mobile Robots X, SPIE, Vol. 2591, 1995, pp. 202-208.

Cheok, K. C., "Autonomous Unmanned Ground Robotic Vehicle Competition: An Intelligent Control Challenge," Proceedings of the American Control Conference, Vol. 1, 1994, pp. 383-387.

Gifford, K. K., Deeds, M., van der Hoek, A., Henning, F., Freeman, J., Haussman, G., Kruzminsky, S. and Stoller, S., "RoboCar: Software, Hardware, and Mechanical Design Issues for the University of Colorado Autonomous Rover Vehicle," Mobile Robots X, SPIE, Vol. 2591, 1995, pp. 228-238.

Matthews, B., Ruthemeyer, M., Perdue, D. and Hall, E. L., "Development of a Mobile Robot for the 1995 AUVS Competition," Mobile Robots X, SPIE, Vol. 2591, 1995, pp. 194-201.

Murphy, R. R., Hoff, W., Blitch, J., Gough, V., Hawkins, D., Hoffman, J. C., Krosley, R., Lyons, T., Mali, A., MacMillan, J. and Warshawsky, S., "Colorado School of Mines Behavioral Approach to the 1995 UGR Competition," Mobile Robots X, SPIE, 1383, November 1990, pp. 436-447.

Nagy, P. V. and Bock, T., "The Northern Illinois University Autonomous Mobile Robots," Mobile Robots X, SPIE, Vol. 2591, 1995, pp. 209-219.

970176

New Thermal Infrared Sensor Techniques for Vehicle Blind Spot Detection

John W. Patchell and R. Steven Hackney
A.L.I.R.T. Advanced Technology Products
Toronto, Ontario

ABSTRACT

Vehicle Blind Spot Detection (BSD) represents a logical first step towards a comprehensive Collision Avoidance strategy as envisioned in plans for the Intelligent Transportation System. BSD systems have been developed based on a number of technologies, but cost and performance issues have prevented widespread commercial success, particularly for light vehicles.

This paper presents a new Vehicle Blind Spot Detector. The device is based on a recently developed, novel thermal infrared sensor technology.

A description of the operation of the sensor is provided, together with a brief comparison with existing technologies. Experimental data are also presented, showing results of operation of the sensor.

INTRODUCTION

Collision Avoidance (CA) systems will form an integral part of the Intelligent Transportation System initiative [1]. Stand-alone Blind Spot Detection represents a logical first step in the development of fully integrated CA [2].

Considerable testing has been done on related technologies [2,3], along with evaluation of prototype and commercial systems [4]. While some systems are available for heavy trucks and other large vehicles, BSD systems are not yet common and are not commercially available for light vehicles (passenger cars and small trucks).

Typically, systems with highly reliable performance are unacceptable because of total installed system cost. Conversely, most inexpensive systems cannot provide required performance across all ranges of environmental conditions and target types.

This paper presents a novel approach to infrared sensing technology with application to Blind Spot Detection. This technology appears to be relatively low cost, and holds the potential to be implemented with high operating performance.

OPERATION OF THE NEW SENSOR

The ALIRT BSD uses long wavelength (7 to 14 μm) infrared technology which senses <u>emitted</u> thermal energy of a vehicle. This is in contrast to <u>reflected</u> energy sensed by active sensors (radar, ultrasonic or short wavelength infrared).

The pyro-electric detectors used in the ALIRT BSD respond only to a change in temperature. If the temperature of the object in the sensing field is constant, the output of the

detector is zero. Essentially, the detector is a differentiator that gives an output proportional to the rate of temperature change within the field of view: no temperature change equals no output. Thus, a BSD based on a single simple infrared detector will give zero output (no detection signal) for a vehicle in the host vehicle's blind spot if there is no relative motion, even though both vehicles may be moving. Conversely, a change in the temperature of the target region, for example an overpass shadow, may trigger a "false alarm."

This limitation is overcome by employing fields of view (FoV's) of two types and alternately comparing them using proprietary techniques (international patents pending). One type of FoV is aimed in the conventional manner at the blind spot while the second is aimed so that only a representative sampling of the road surface is visible -- for example immediately behind or below the host vehicle.

If no vehicle is in the blind spot, then both types of fields of view see the road surface. The difference in temperature between the two FoV's is consequently small and the output of the detector is small.

However, when a vehicle is present in one FoV and not in the other, then the output is large. When a vehicle enters the blind spot there is a positive differential between the detection and reference zones of the sensor. This differential is used to trigger an output signal (for example a light or buzzer) to warn the driver of the lane change hazard.

In the case of a change in the road temperature, as generated by the shadow of an overpass, for example, both the reference zone and the detection zone will change temperature at approximately the same time. Therefore the differential will remain zero.

One possible FoV arrangement is shown in Figure 1. The unit is shown mounted on the left side rear view mirror. The road-only element views the road directly behind the car while three lens elements are used to cover the driver's blind spot. In the initial prototype, only one of the three illustrated Detection FoV's has been implemented.

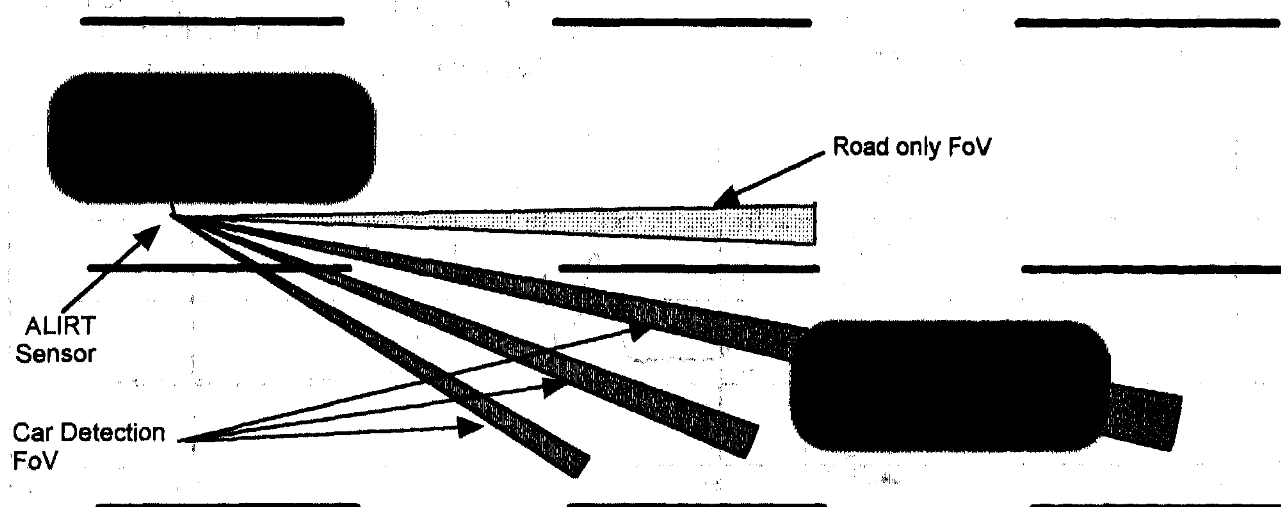
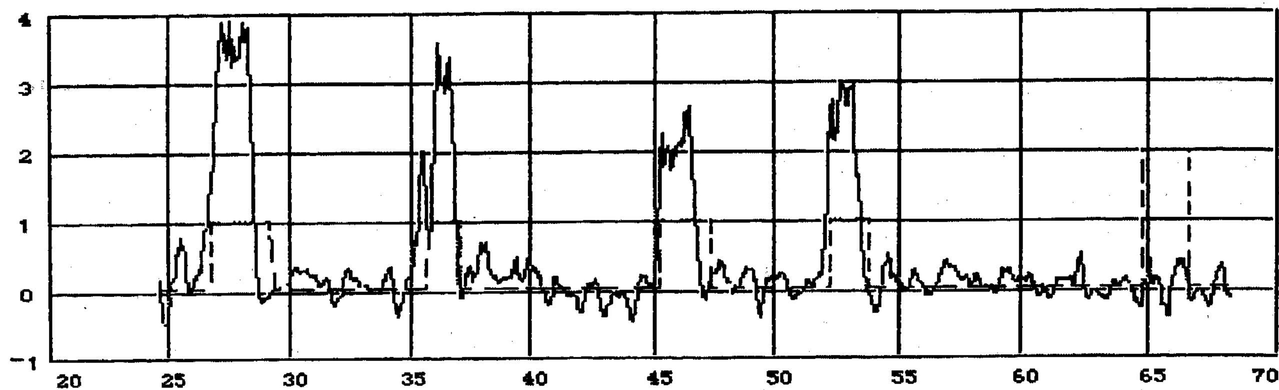


Figure 1 - Arrangement of Fields of View



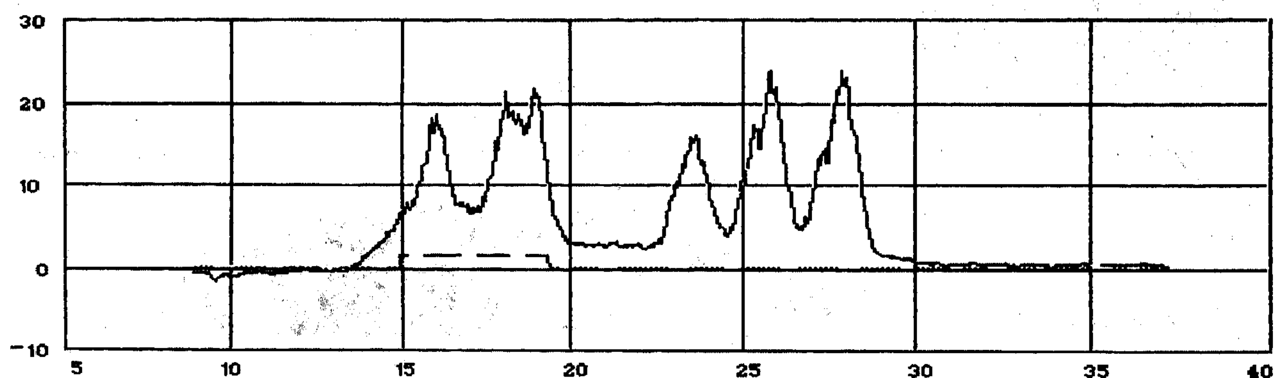
Time: 3:40 PM
Weather: Sunny, 25C
Comment: The first four events are cars. The final event (at about 65 seconds) is a bridge shadow.

Figure 2 - Detection of Cars

EXPERIMENTAL RESULTS

TEST RESULTS ACHIEVED Early testing has indicated that portions of operating vehicles are reliably warmer than the environment by at least 1 to 2C. This has been found to be true of the engine compartment, radiator and exhaust system: testing has also shown that tires will achieve approximately 2C of heating in approximately 1 km of driving in city condition. Therefore, a moving vehicle will emit significantly greater infrared than its surroundings.

Considerable testing has been conducted using a prototype device. The following graphs are typical of the results obtained: similar data have been collected across a range of



Time: 10:00 AM
Weather: 17C, Partly sunny
Comment: Transport truck signal. The first peak at 16 sec is front wheels, 18-19 sec is dual axles on rear of cab and 23, 26, and 28 sec the three axles on the trailer.

temperatures, weather conditions and times of day.

In figure 2 and 3 , the solid line is the BSD output. The broken line is a manually entered "event" channel used as a marker to record the occurrence of a vehicle in the blind spot or other significant occurrence. The horizontal axis is time in seconds.

The above data were collected using a sensor configured with a single detection field of view and a single reference field of view. The detection zone is approximately 300 mm diameter at the road surface.

Consequently, under some circumstances this configuration of the device will not detect large vehicles since the tires of the vehicle actually straddle the detection zone. This can clearly be seen in the case of the trailer of the transport truck. During a portion of the time that the trailer is resident in the blind spot, no detection is registered.

A second implementation of the device has been constructed with revised reference and detection zones . As indicated in figure 4, this version of the BSD features two distinct detection zones. This has largely eliminated the "straddle" problem found with the single-zone implementation of the original prototype.

Figure 5 shows the detection of a passenger bus using this version of the detector. With the current threshold alarm trigger of approximately 0.6C, the vehicle is continuously detected during the entire time it is in the blind spot of the host vehicle.

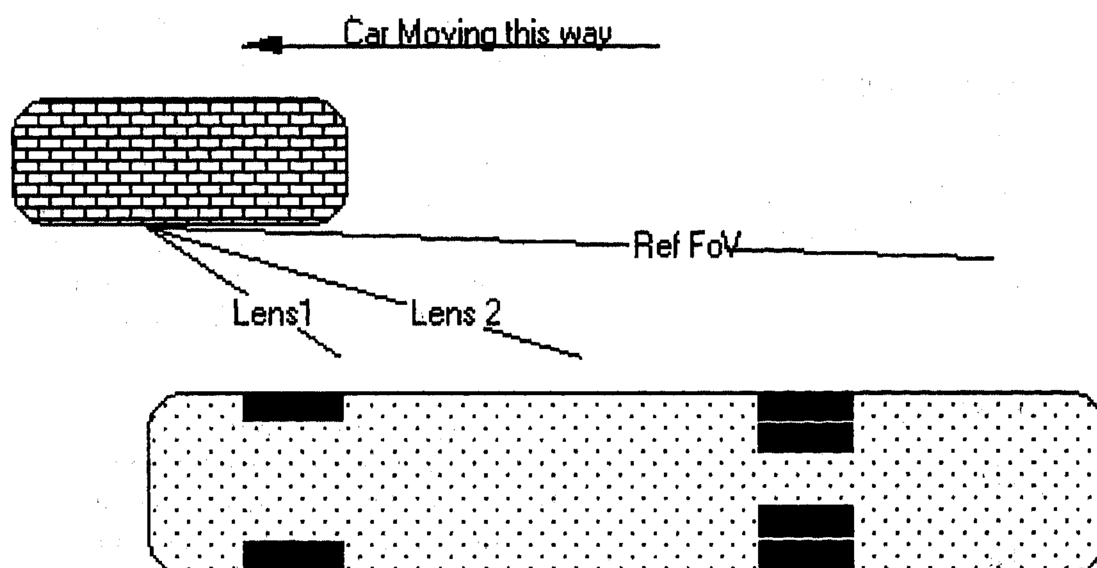


Figure 4 - Configuration of Multiple Detection-Zone BSD

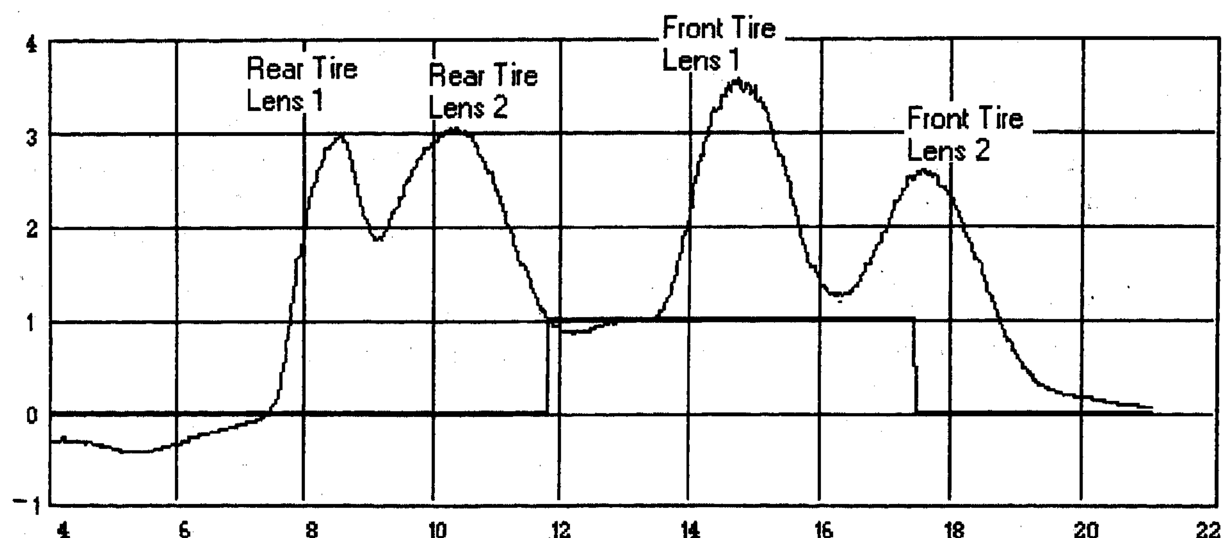


Figure 5 - Detection of a Bus Using Two Detection-Zone BSD

As can be seen from these figures the ALIRT BSD is very effective in detecting vehicles in real road conditions.

COMPARISON WITH EXISTING TECHNOLOGIES

A number of technologies have been employed in attempting to solve the BSD problem, including:

- Ultrasonic
- Active Infrared
- Laser
- Radar

These technologies are all "active" -- that is they all make use of a sensor that detects a signal which is transmitted from the host vehicle and reflected off of the target vehicle.

By contrast, the ALIRT BSD technology is passive. This strategy is expected to offer the following advantages:

System Cost: Because a passive system is inherently less complex than an active system, advantages are expected in the areas of system cost, installation cost and general reliability. Total sensor cost is estimated to be less than twenty dollars with volume manufacture.

Reliability: Near infrared has been generally found to be environmentally stable. Testing has shown the sensor to be effective in conditions of rain, winter and summer temperatures, snow and daylight or darkness. The system detects operating vehicles independently of colour, size or shape. It does not respond to parked vehicles or other non emitting roadside objects. IR transmissions are

relatively unaffected by airborne moisture. Fog will affect the sensor operation less than it will affect the driver's vision.

CONCLUSIONS

A vehicle Blind Spot Detection system has been developed based on a new passive infrared technology. Testing to date indicates that the detector will be able to provide a high level of performance at relatively low cost.

REFERENCES

[1] M. Upton, "Techniques for Distance Measurement," Society of Automotive Engineers Technical Paper No. 952085, 1995

[2] S.J. Mraz, "On the Road to Smarter Cars," Machine Design, September 14, 1995

[3] E. N. Mazzae, W. R. Garrott and M. A. Flick, " Human Factors Evaluation of Existing Side Collision Avoidance System Driver Interfaces" Society of Automotive Engineers Technical Paper No. 952659, 1995

[4] "Side and Rear Object Detection Systems," Truck Engineering, November 1995.

970190

Active Matrix LCD Performance in the Automotive Environment

Mark R. Vincen
Delco Electronics Corp.

Abstract

The environment of the automobile poses severe challenges to AMLCD technology. Among automotive display applications, the primary instrument, or instrument cluster, is the most demanding due to the safety concerns of display shortcomings or failure.

This paper will address and summarize issues involving the application of AMLCD display technology to the primary instrument. The results of environmental testing of ruggedized automotive AMLCD modules will be presented, without revealing the display module manufacturers' names.

The paper will also reveal the results of supplemental testing of these AMLCD modules. Tests that characterize AMLCD display performance such as contrast, brightness, response time and viewing angle performance vs. ambient temperature will be presented.

Introduction

As already mentioned, the primary instrument is the most demanding display application in the automobile. This is the case because safety related information is displayed by the instrument cluster. Vehicle gear position, speed, and critical telltales or warning indications such as brake activation, etc., are examples. The federal government regulates specific aspects of these critical functions via the FMC (Federal Motor Vehicle System Standards) regulations. Aspects such as size, color, functionality, and readability are addressed. Importantly, if the vehicle is derivable, these critical functions must be operative.

Requirements for the Primary Instrument Display

The following requirements for a primary display are the most severe for a display in the automobile. These requirements address optical readability and comfort as well as the environmental performance and durability attributes necessary for high reliability and safety.

Optical Requirements:

DISPLAY BRIGHTNESS (white)	150 Footlamberts
CONTRAST RATIO (CR)	100/1
DISPLAY DIMMING RATIO	100/1
VIEWING ANGLE	+/- 60 deg. H,V
RESPONSE TIME	50 ms @ 25 deg. C
DEFECTIVE PIXELS	1 per display 'ON"
	1 per sq. in. "OFF"
RESOLUTION	80 color groups/in.
HIGH AMBIENT CONTRAST RATIO	10/1 min. white CR
	@ 3000 FC

Environmental Requirements:

OPERATING TEMPERATURE RANGE	-40 to +85° C
PERFORMANCE TEMPERATURE RANGE	-20 to +70° C
STORAGE RANGE	-40 to +95°C
POWERED TEMP	-40 to +85°C
HIGH TEMP ENDURANCE	85°C, 250 hrs.
HUMIDITY	65° C, 94% RH
VIBRATION	3.2 G, Multi-axis
EMC	GM 9114P
ULTRAVIOLET EXPOSURE	500 HRS (Xenon device)

The preceding optical and environmental requirements have been addressed in a set of Pre-validation tests. These tests are based on an actual Validation test set for a production

primary instrument cluster. A primary cluster using the AMLCD technology would be required to pass these Validation tests in order to be qualified for production. The actual Production Validation tests require certain minimum test sample quantities; we used smaller quantities to lower costs and save time. The purpose of this Pre-validation test was to identify potential problem areas with the test modules, not to qualify them for production. AMLCD modules from several manufacturers were tested. The following is a block diagram of the Pre-validation test procedure:

Pre-Validation Test Sequence:

Power Temperature Cycling (PTC)	500 - one hour Cycles -40 to +85 C 15 minute dwells and transitions Device power 5 minutes ON, OFF
High Temperature Endurance (HTE)	+85 C Steady State, 250 hours
Biased Humidity Test (BHT)	+65 C, 94% R.H. Steady State (non-condensing)
Ultraviolet Exposure Test (UVE)	500 hr. Exposure in Xenon-lamp Device
Electro-Magnetic Compatibility (EMC)	GM 9114P Test Profile
Frost Test	-40 C Soak, then +35 C @ 94% R.H device must be operational for 30 minutes
Vibration Test	3.2 g, Multi-axis 0-1000 Hz.

Results of Environmental Testing

The results of the environmental testing on several manufacturer's AMLCD modules revealed problems with the display polarizers and the cold cathode fluorescent tube (CCFT) backlight system. The following summarizes these problems:

Display Polarizers

Our testing revealed polarizer problems in two areas. The first problem identified involved moisture penetration into the polarizer. After the Biased Humidity Test, moisture absorption into the polarizer caused strange optical effects including the appearance of a pronounced "X-Pattern" becoming visible in the display. This pattern moved with the observing angle of the viewer. Some polarizers are more susceptible to this problem than others. One polarizer type showed abnormal front surface reflection patterns in addition to the "X-Pattern" seen through the display.

The second problem found was the delamination of the polarizer from the display glass substrate. This problem was found after the High Temperature Endurance Test (+85°C, 250 hrs.). Again, not all of the polarizers used on the various modules tested had this delamination problem. The polarizer adhesive is obviously suspect here.

The polarizers used in these AMLCD modules are the poly-vinyl alcohol (PVA) type. This type of polarizer has good optical performance, including color neutrality, but it has high temperature limitations. We are told the upper temperature limit for PVA type polarizers is about 95°C. This is only 10°C above the +85° C temperature we are testing the modules at. The polarizers, especially the rear one, are being heated by the module backlight system. This temperature rise behind the display is in the 10°- 30° C range. This means the rear of the display can be at 115°C when the module is in a +85°C ambient as it is in the Pre-Validation tests. This temperature is clearly above the maximum temperature of the PVA type polarizer. Interestingly, some of the polarizers tested were able to pass the High Temperature Endurance (HTE) test indicating the PVA polarizer can withstand the high temperatures for a limited period of time. This is greatly helped if the particular module backlight imparts a relatively lower temperature rise to the polarizers.

CCFT Backlight System

The next major problem revealed by the testing involves the cold cathode fluorescent (CCFT) backlight system. The Power Temperature Cycling (PTC) test damaged some CCFT backlight systems, resulting in catastrophic failure. The backlights still functioned, but were

very dim and orange in appearance. Analysis indicated the lamps had been damaged in two significant ways. One, the lamp electrodes were damaged by activation at -40 C when the lamp mercury is in the liquid state. The activation of the plasma flow inside the lamp without the mercury in the gaseous state results in eventual damage to the electrodes. The PTC test took the CCFT lamps down to -40°C 500 times (and activated the lamps) during the test, which is enough to cause failure with some backlight systems. The second failure also involved the mercury in the liquid state. At -40°C the liquid mercury collects at the bottom or lowest section of the CCFT tube. The chances are good that the collection point is coated with phosphor. The liquid mercury can contaminate the phosphor, damaging the lamp's function even after the lamp returns to room temperature.

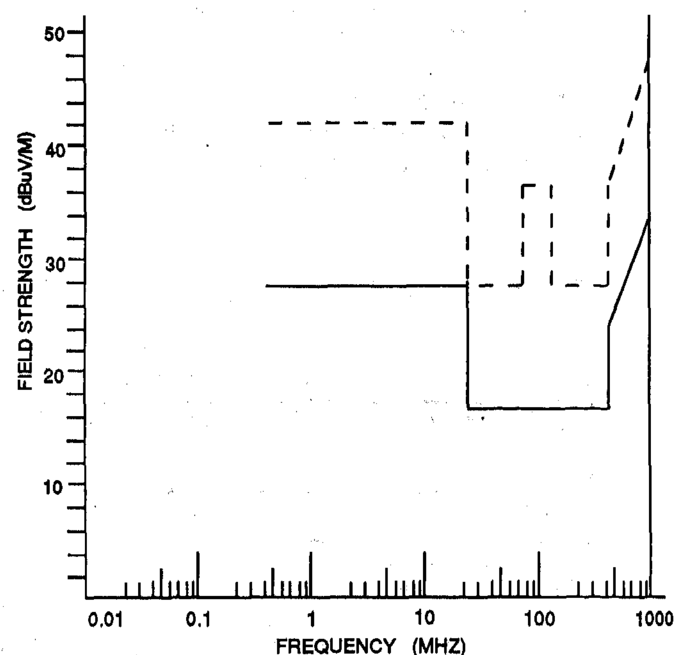
These two failure mechanisms are major drawbacks to CCFT system applications at low temperature extremes. To minimize some of these problems it would be necessary to heat the lamps and delay trying to operate them at these extreme temperatures.

Electro-Magnetic Compatibility (EMC)

All of the AMLCD modules had very high RF emissions; both radiated and conducted through the wiring harness. Figure 1. is a description of the acceptable levels of Radiated Emissions per the GM 9114P specification.

The results of our EMC testing on several manufacturers' AMLCD Modules indicate that RF emissions by these modules is a serious problem. The CCFT backlight system is a major source, contributing to most of the emissions at the lower frequencies including the AM radio band. The remaining module electronics, including the video circuitry and row/column drivers, produce most of the RF emissions at the higher frequencies including the FM and Cellular phone bands.

None of the tested modules used RF suppression techniques (as far as we could determine) which would make a significant reduction in RF noise levels. However, the RF levels are so high without suppression techniques that we have concern that the AMLCD modules will still ultimately have relatively high RF emission levels. This is especially problematic because vehicle manufacturers are requiring lower and lower emission levels.

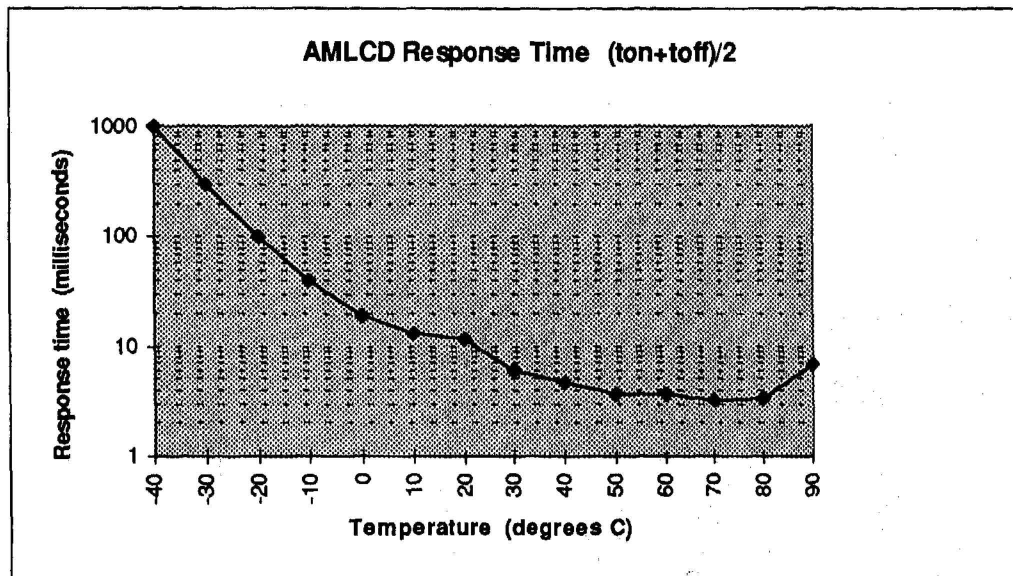


Frequency Range (MHZ)	Amplitude (dBuV/m) Narrowband	Broadband
0.4 - 75.167	28	41
20 - 75.167	15	28
75.167 - 119.132	15	38
119.132 - 375	15	28
375 - 1000	25 - 34.5	38 - 47.8

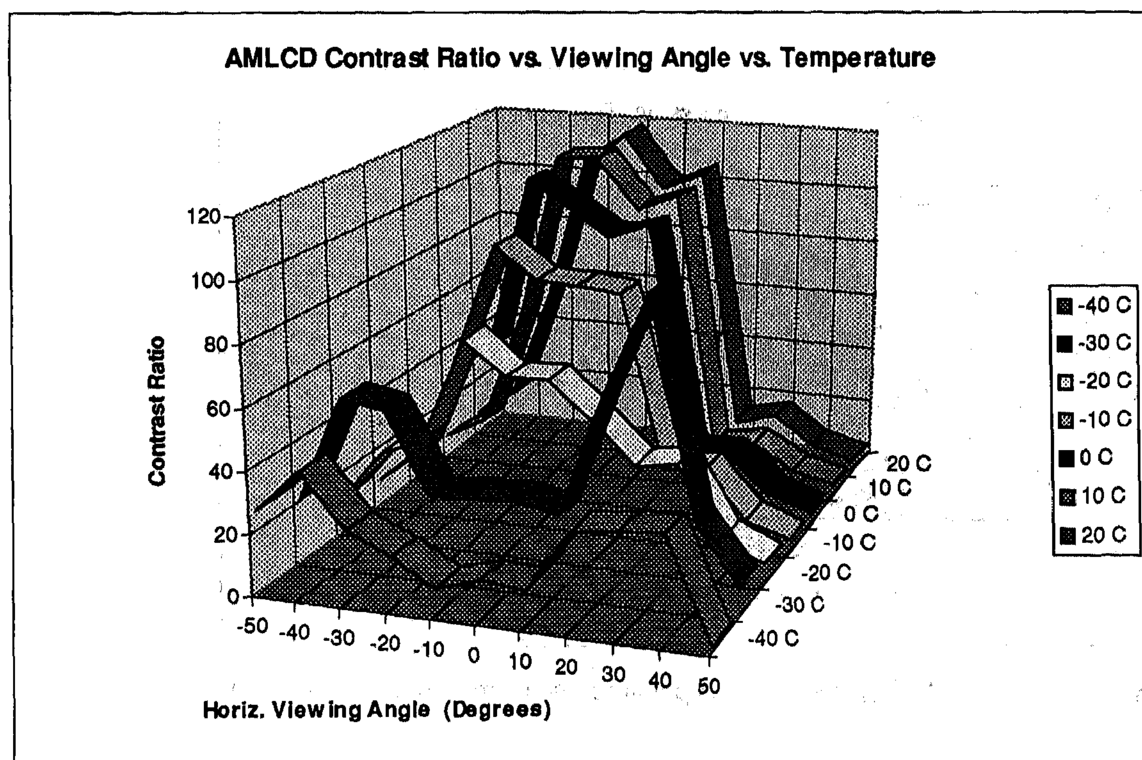
Figure 1. - Radiated Emissions Performance Objectives

Supplemental Testing and Evaluation

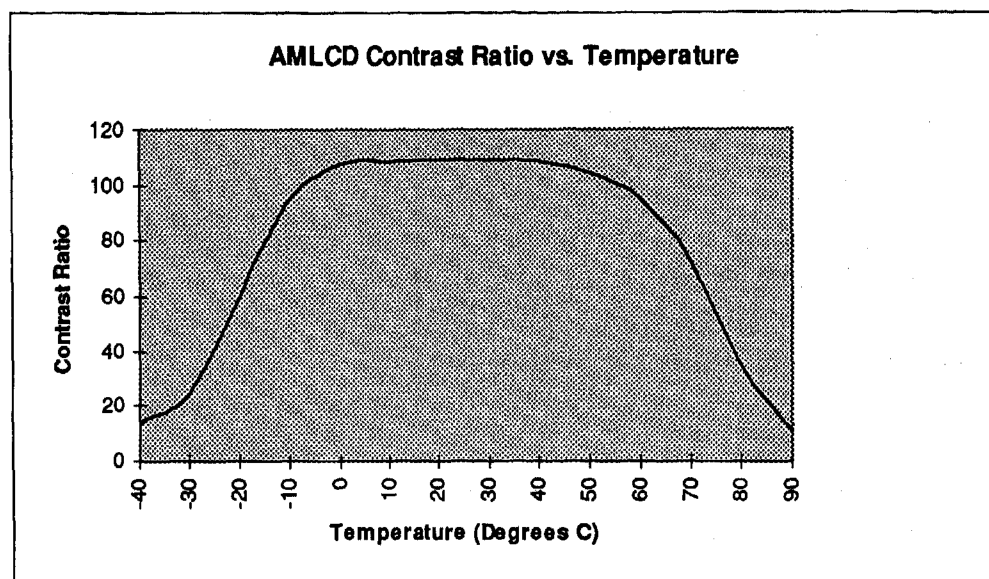
Additional optical and EMC testing was performed on some of the AMLCD modules. The optical testing involved the measurement of display brightness (incl. lifetime), contrast, response time, and viewing angle performance of the display vs. temperature. Graphs 1 - 7 show the results of this evaluation effort. The AMLCD modules that were evaluated did not have the new wide-angle optical film. Ruggedized AMLCD modules with this film were not available at the time of this writing.



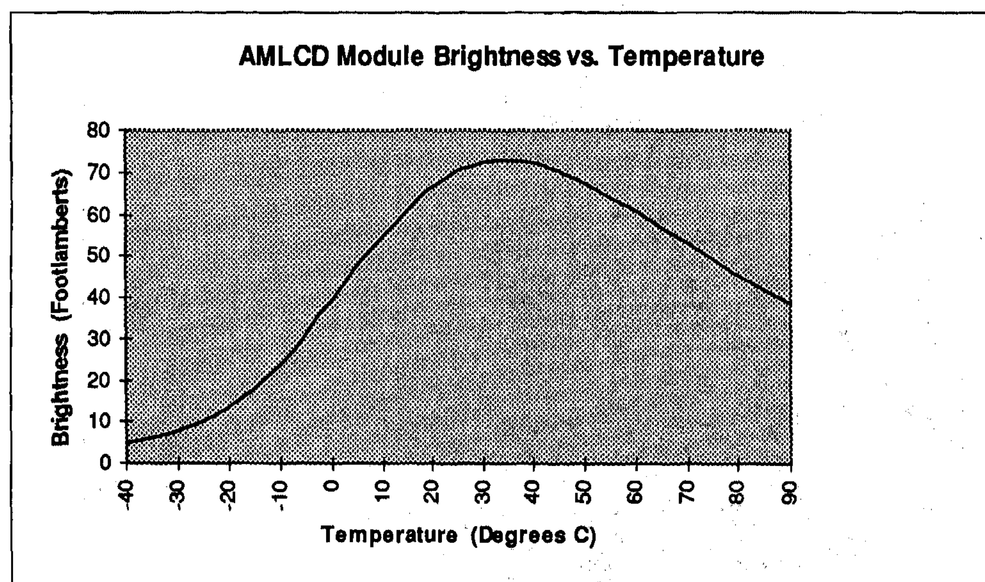
Graph 3. AMLCD Response Time vs. Ambient Temperature



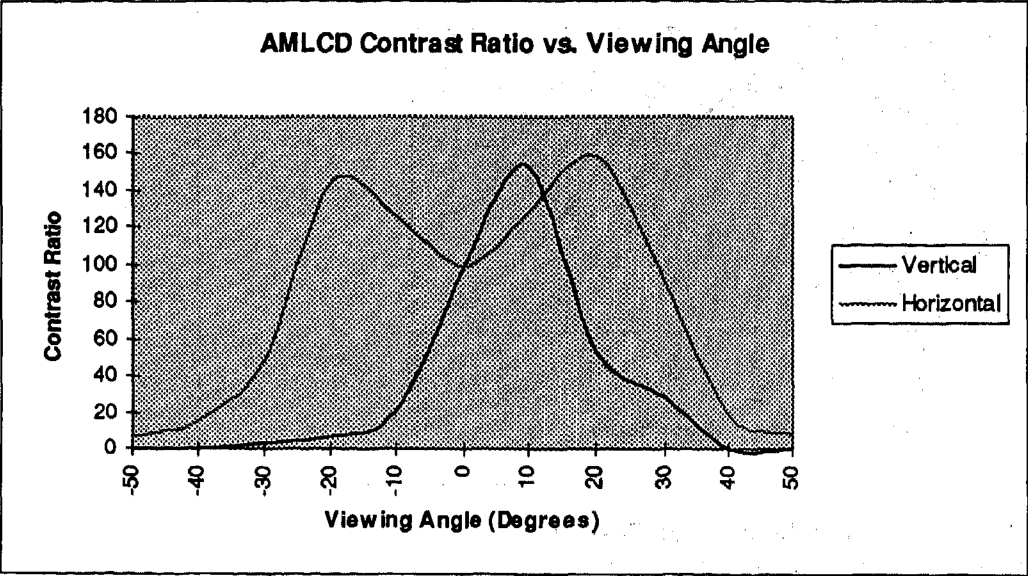
Graph 4. AMLCD Module Contrast Ratio vs. Viewing Angle vs. Temperature



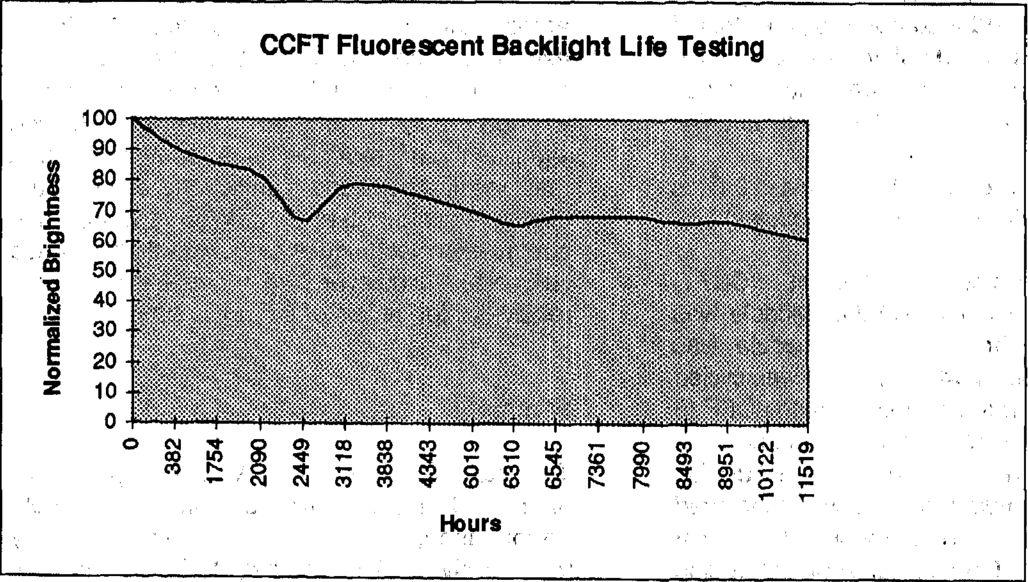
Graph 1. AMLCD Contrast Ratio vs. Temperature (Measured Normal to Display)



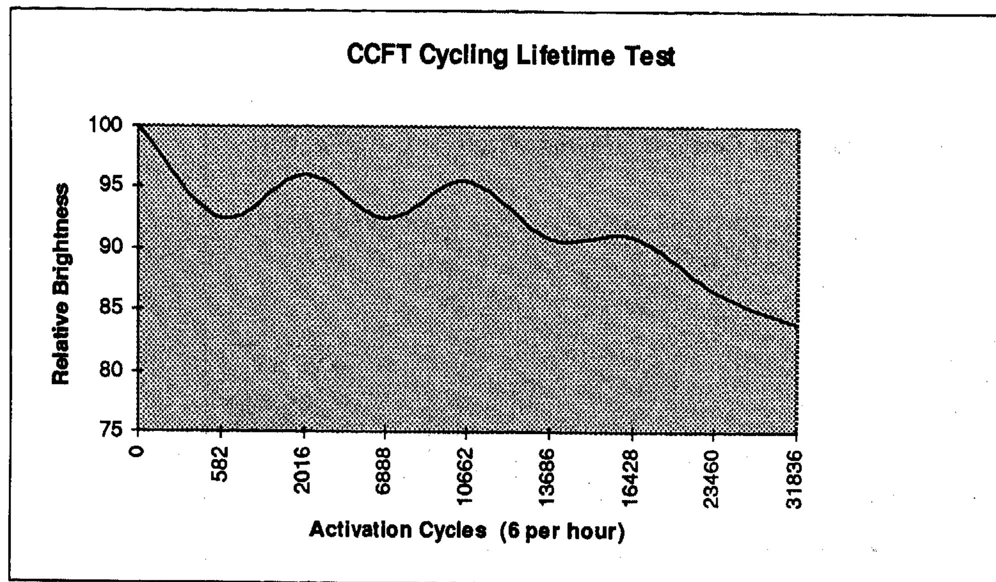
Graph 2. AMLCD Module Brightness vs. Temperature (measured Normal to Display)



Graph 5. AMLCD Module Viewing Angle vs. Temperature



Graph 6. CCFT Fluorescent Backlight Life Test (Continuous Operation)



Graph 7. CCFT Fluorescent Backlight Cycle Life Testing (Lamp ON 5 minutes, OFF 5 Minutes [6 cycles\hour])

The results of these supplemental tests indicate the sensitivity of AMLCD technology to the ambient temperature. In the case of the Primary Instrument application, this degraded performance, (i.e, reduced contrast, constricted viewing angle, slow display response, and dimmer display brightness), demands the use of a heater to bring the AMLCD display and the

CCFT backlight system up to at least -20 C as quickly as possible (with present modules).

The Cold Cathode Fluorescent (CCFT) backlight for a particular AMLCD Module was evaluated in a Life Test. A module was powered continuously while another was cycled at the profile of 5 minutes ON, minutes OFF (6 cycles\hour). Both tests were conducted at room temperature (23°C). The display brightness of the two modules was measured periodically. The continuously ON test evaluates the CCFT phosphor while the cycling test evaluates the CCFT electrodes. The results indicated the CCFT backlight is capable of a 10,000 hour life (to half initial brightness) in both a constant and cycling mode at room temperature.

We conducted numerous supplemental EMC tests on these same AMLCD modules. The result of the baseline (as-is out of box) EMC testing showed very high levels of RF noise across the measured spectrum. Continued experimentation revealed that there was, as expected, significant RF noise conducted through the wire harness. Tests also showed significant RF radiation was coming through the module case, and significantly, through the display glass itself. We applied RF suppression techniques to substantially reduce RF noise. However, the module still exhibits relatively high levels of RF noise emission.

Summary

The next generation of AMLCD displays will offer substantial improvements in viewing angle performance, contrast, and display transmissivity. This will help to lower backlight brightness (and temperature rise) and will make the display more sunlight readable.

Improvements on the horizon include the integration of logic and row/column driver

circuitry on the display glass for lower cost and higher reliability.

The present state of AMLCD technology does not appear to be fully ready for automotive primary cluster applications at a reasonable cost (as measured against the standards outlined in this paper). Improvements need to be realized in both optical performance (especially vs. temperature) and environmental ruggedness.

However, AMLCD technology is continuing to evolve and improve. Up to this point, the computer and consumer electronics applications have fueled AMLCD growth and influenced its development. We see a new wave of growth on the horizon for the automotive application of AMLCD technology. As automotive applications increase and higher volumes appear on the horizon, AMLCD R&D will focus more and more on the automotive issues and challenges.

Acknowledgments:

The author wishes to thank Karl Stone for his support and general consultation and Mark Helton and Greg Watson for their consultation on EMC issues.

970227

Fundamental Issues in Automotive Veiling Glare

E. N. Boulos, D. Jack, R. Surowiec, and J. L. Bomback
Ford Motor Company

S. Subramanian, C. J. Simmons, and J. H. Simmons
University of Florida

Copyright 1997 Society of Automotive Engineers, Inc

ABSTRACT

The veiling glare effect in automotive vehicles consists of diffuse and specular scattering of sunlight onto and from the windshield. This effect occurs over a wide range of solar elevation angles and increases with increased degree of inclination of the windshield. Thus its effect on visual acuity must be considered in automotive design. The present research on the subject of veiling glare only addresses scattering from a clean windshield and ignores the larger effect of scattering from dust, dirt or haze on the front and back faces of the windshield since the latter is operator dependent (can be removed by cleaning the windshield). In this paper, we present an analysis of autmotive veiling glare that takes into account windshield reflectivity without and with coatings, and the characteristics of dashboard cover materials. The analysis develops a new reflection model for rough surfaces, based on laboratory measurements of reflectivity as a function of wavelength and angle of incidence, and on reflection and/or scattering angle.In the case of windshield reflectivity, single-film, 3-film and graded-index coatings are considered for both s and p polarizations and a range of windshield rake angles, and calculations are compared to laboratory and full scale measurements. In the case of dashboard reflectivity, surface texture and material ligthtness are considered. Specular and diffuse reflection components are analyzed in terms of angle of incidence, color and polarization direction. It is found that texture has a large effect on the reflected light, especially at large angles of incidence and causes a noticeable forward scattering component which includes specular reflection. Experimental measurements and computer-generated models show that this forward scattering effect can be as large as 10-20 degrees. The presence of forward scattering has the result of increasing the range of solar elevation angles that produce veiling glare. Windshield and dashboard reflectivity effects are then combined and analyzed to determine how they impact the veiling glare process.

INTRODUCTION

The stray light, not associated with the incident image, that is seen in the windshield by the driver of a vehicle is called glare. It results from light reflected and/or scattered by the windshield. This light can come from basically two sources: (1) the reflection of sunlight from the dashboard cover material, and (2) the scattering of sunlight from deposits on the windshield of the vehicle. The latter is easily removed by the vehicle operator by simply washing the windshield. The former is often what is referred to as veiling glare. This paper will discuss the results of a comprehensive model developed to simulate the contributions to veiling glare from the windshield with and without anti-reflective coatings and from the texture of dashboard cover materials.

Veiling glare must be controlled in order to assure clear visibility of objects ahead. Attempts to reduce veiling glare have played an important role in automotive design. Because the major component of veiling glare has been reflection by the windshield of light scattered or reflected by the instrument panel cover, its effect has been considered in the design and color selection of instrument panels and dashboard materials for the broad range of windshield rake angles used today.

Although a sandwich of polymer film between two glass panes, the windshield may be treated optically as a single glass panel, because the refractive index of the polymer layer matches that of the surrounding glass. It is well known that the reflection of light from an uncoated glass plate exhibits a strong dependence on wavelength as well as angle of incidence. This is shown in Fig. 1 where the reflection is calculated from both inner and outer surfaces of the windshield for s- and p-polarized incident light. Note that for windshield rake angles (angles of incidence) of 60-65 degrees, the reflectivity coefficient approaches 20%. Of course, that is 20% of the light reflected from the dashboard cover, not the light incident on the windshield from outside. Therefore, the reflectivity

Table 1 - Relationship of solar elevation angles to windshield and dashboard angles

Wind\Dashb	-5°	0°	5°	10°	15°	20°
50°	90	80	70	60	50	40
55°	80	70	60	50	40	30
60°	70	60	50	40	30	20
65°	60	50	40	30	20	10
70°	50	40	30	20	10	0
75°	40	30	20	10	0	-10

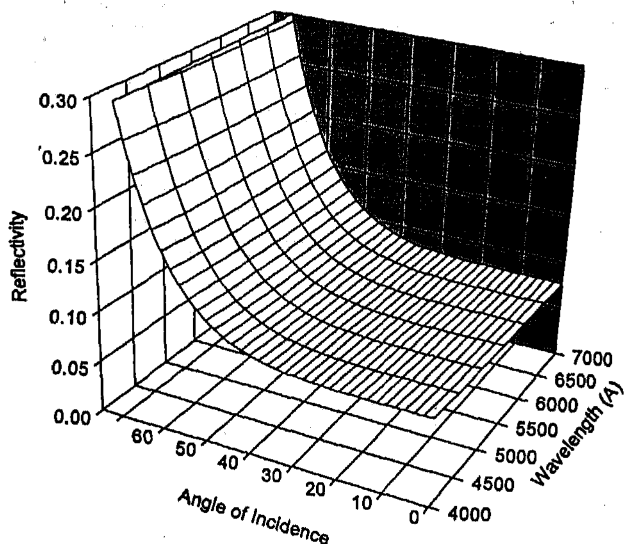


Fig. 1 - Reflectivity from both sides of the windshield, summing s and p components.

of the dashboard cover material is of utmost importance in the evaluation of driver visual performance. As a result, anti-reflection (AR) coatings and/or low reflection - low scattering dashboard materials have been considered in automotive designs that have large windshield rake angles.

AR coatings fall into two classes: interference or graded. The interference coatings consist of a single film or a stack of films designed to reflect light destructively for a given wavelength and angle of incidence. Graded coatings are designed to reduce the reflected light by decreasing the sudden jump in refractive index between air (n=1.0) and the windshield (n=1.5). In general, single-film interference coatings have a narrow reflection minimum as a function of wavelength, and angle of incidence. This can be improved by making multilayer coatings. By contrast, graded-index coatings yield a broader minimum, but the reflectance minimum depends directly on the refractive index at the outer edge of the coating. This number is limited by MgF_2 (1.38) which is about the lowest limit of glass refractive indices.

In most windshield designs, only the inner surface is coated with AR films. Therefore, the inward reflected light from the outer surface of the windshield contributes the greatest portion of the veiling glare, and it determines the minimum reflection achievable.

When considering the geometry of the windshield-dashboard-driver system, it is possible to calculate the solar elevation angles which will maximize veiling glare for a variety of windshield inclination and dashboard angles as shown in Fig. 2. The solar elevation angles are shown in Table 1. In the table, the solar elevation and the dashboard angles are measured from the horizontal while the windshield angle is measured from the vertical. It is clear that solar veiling glare occurs at relatively high elevation angles for windshield angles of 55-65 degrees. As the solar elevation angle decreases to the horizontal, the solar veiling glare becomes inconsequential since the sun will also be directed into the driver's eyes.IN this case, the driver will take action to avoid looking into the sun, also eliminating the veiling glare.

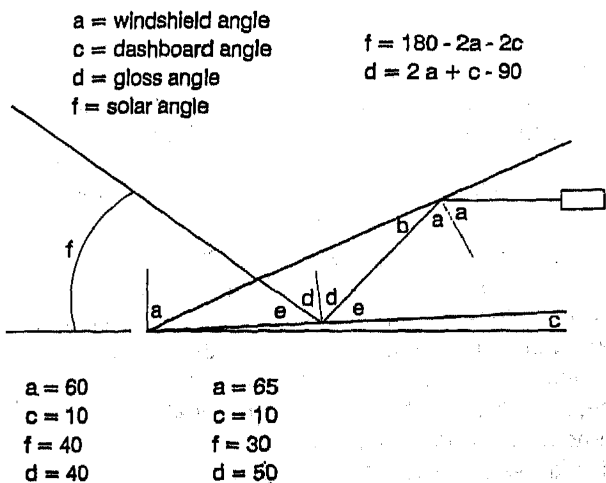


Fig. 2 - Sketch of windshield, dashboard and solar elevation relationships.

Methods for mitigating the amount of veiling glare received by the driver must rely on the use of anti-reflective windshield coatings as discussed above, and on the use of non-reflective dashboard cover materials. In this paper, we discuss a comprehensive model designed to simulate the effect on veiling glare of windshield AR coating materials and of surface roughness in the instrument panel covers.

Main Section

ANTI-REFLECTIVE WINDSHIELD COATINGS - In a previous paper,[1] we calculated the reflection coefficient for windshields without and with anti-reflective coatings on on the inner surface for s and p polarized light. A menu-driven computer program was developed in Excel format to calculate and plot the total reflectivity expected from uncoated windshields and those coated with any combination of up to 10 layers. The studies showed that s- and p- polarized reflected light can be reduced with single-layer, 3-layer and graded-index coatings on the inner surface of the windshield. The graded-index coatings were developed using a proprietary sol-gel fluorination process that reaches a minimum refractive index of 1.41 at the outer surface of the coating. One of the major advantages of graded-index coatings is that they show a much weaker wavelength dependence than the multi-layer coatings, thus the coated material does not exhibit a tinge of color.

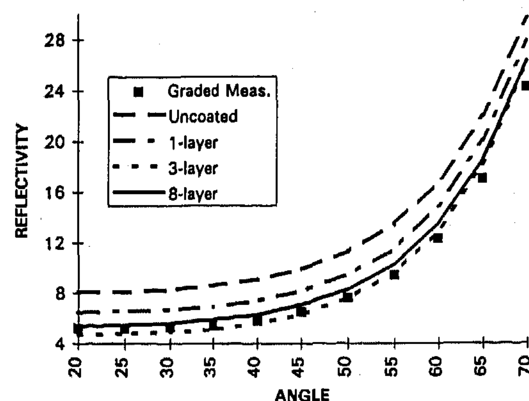


Fig. 3 - Percent reflectivity vs angle for green light compared between uncoated windshields and windshields coated with 1-layer, 3-layer and a graded-index coating (represented by 8 layers).

The results of reflectivity measurements are shown in Fig. 3 for measurements at the wavelength corresponding to the green color (550 nm). Two 3-layer coatings were tested. In the first, we used index and thickness relationships available in commercial AR coatings. These, however, are best used with light at normal incidence. AR behavior can be improved by using index and thickness values more appropriate for angles of incidence between 50 and 70 degrees. The results of using these new coatings are shown in Fig. 3, along with the reflectivity of an optimized 8-layer coating. Note that this formulation gives a lower reflectivity than the graded-index coating only at shallow angles of incidence. However, as stated above, the graded-index coating gives less wavelength dependence and avoids any color in its reflection spectrum.

DASHBOARD MATERIAL ROUGHNESS -In order to develop an understanding of the role f roughness and lightness of dashboard cover material on reflectivity, we conducted exploratory laboratory measurements followed by the development of a new model. The model was then reduced to a menu-driven Excel-based program which can be combined with the windshield and coating calculation program to yield a comprehensive simulation approach to modelling the entire veiling glare system. In the experimental study, we examined several dashboard materials of different color and roughness. The roughnesses were measured with a gloss meter at 60°. Reflectivity measurements from the dashboard materials for light incident at 60-65 degrees showed an unexpected forward shift in the reflected light from the specular angle as shown in Fig. 4 for both s and p components. The shift in reflection angle from specular was observed to be as large as 17 degrees. This forward shift was more pronounced in the p-component of the light for all materials. We find that this shift increases the range of solar elevation or rake angles that maximizes solar veiling glare. Thus it extends the effect through a greater daytime period.

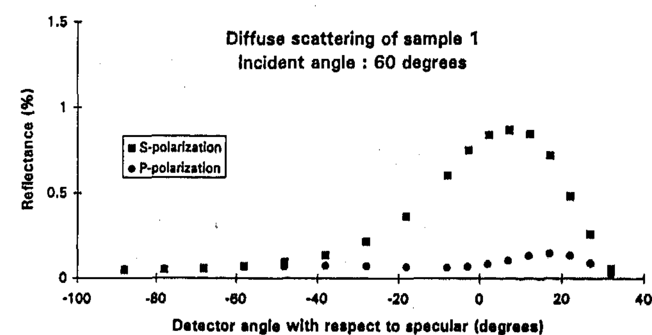


Fig.4 - Reflected light from sample 1 cover material, showing the shift in the reflection peak for s and p-polarized light

An analysis of the source of the forward shift in the reflection peak determined that it was due to the fact that a rough surface exhibits a distribution of reflection angles. We modelled this process by generating Gaussian distributions of reflection angles centered about the windshield rake angle. Calculation of the angular dependence of the reflected light showed that for most reasonable entrance apertures, there will be a forward shift in the reflection angle as shown in Figs. 5a and 5b, for a variety of distributions. Measurements of the roughness of the dashboard cover surface were conducted using a contact profilometer and the resulting angular distribution is shown in Fig. 6. The curve is best fitted by two Gaussian distributions (one very broad and one very narrow). Even though the simulations resulted in

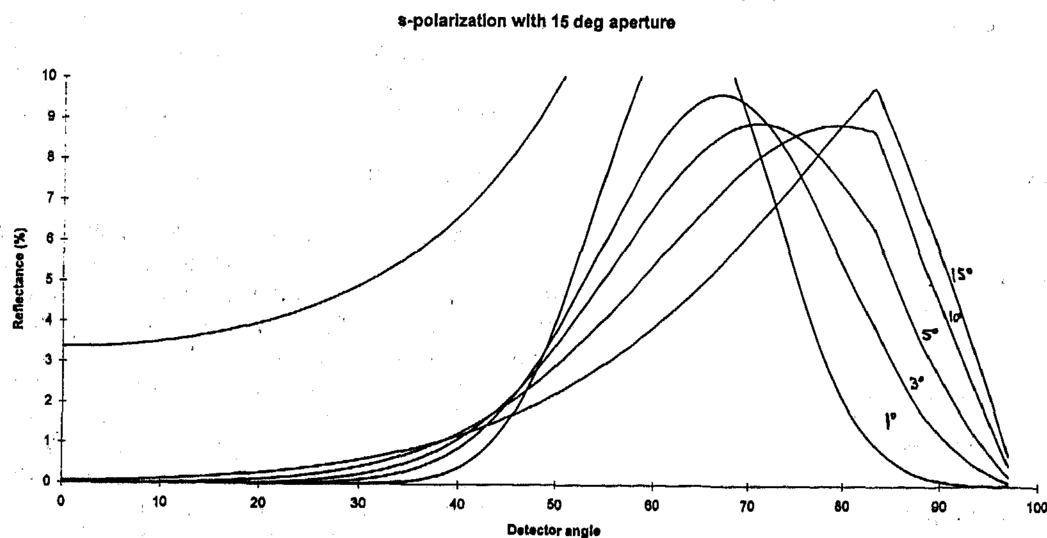


Fig. 5a - Calculations of forward scattering of the s-polarization for a variety of angular distributions corresponding to different roughnesses (angular distribution widths of 1°, 3°, 5°, 10° and 15°).

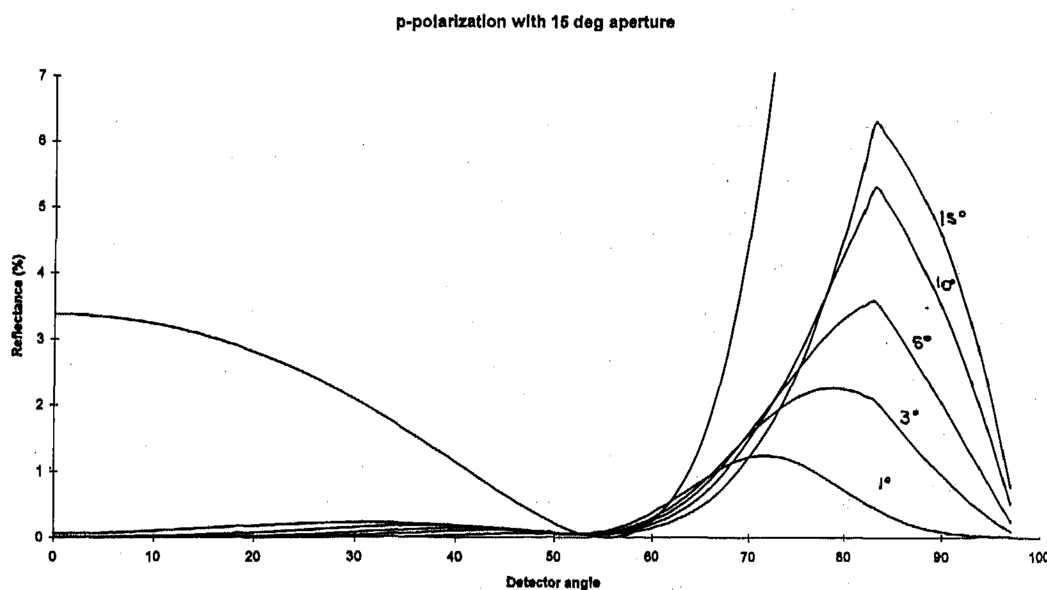


Fig. 5b - Similar plot for the p-polarization.

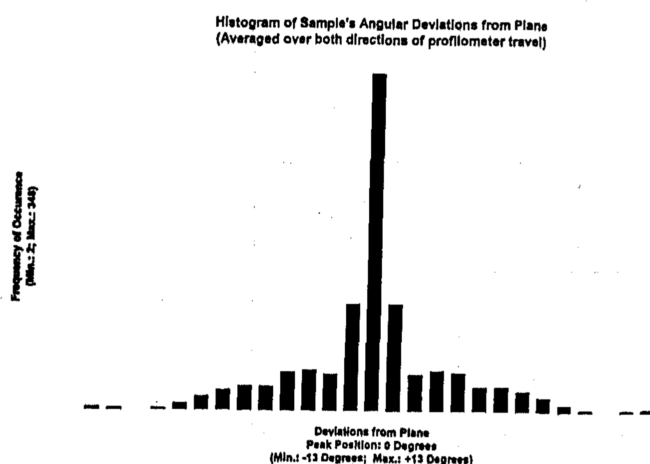


Fig. 6 - Histogram of dashboard roughness.

very close agreement with the measurements of the angular dependence of the reflection coefficients, tests were conducted to investigate the effect of smaller angles of incidence. Clearly, if the forward reflection is due to a combination of a broad distribution of surface angles and the asymmetry in the angular dependence of the reflection coefficient as shown in Fig. 2, the s and p components will behave differently around Brewster's angle of incidence. Measurements were conducted at 50° (5° below Brewster's angle) and at 55° (Brewster's angle). The results showed a backward shift in the reflection from specular at a 50° angle of incidence in only the p-component while the s-component exhibited the expected forward shift in scattering angle. At Brewster's angle, the p-component showed two peaks, one above and one below specular and zero reflection at the specular angle (see Fig. 7). This strictly follows the behavior predicted by our model and shows without doubt that the forward scattering measured

forward scattering measured at 60° and 65° angles of incidence is a direct result of the roughness of the dashboard cover.

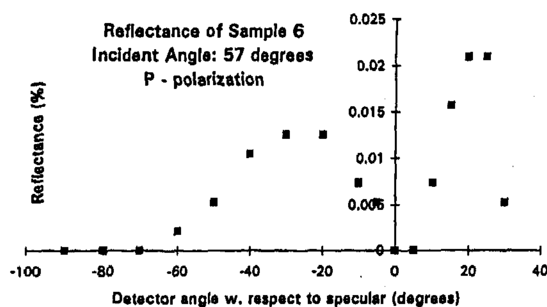


Fig. 7 - Reflectivity of the p-polarization at Brewster's angle.

Dashboard materials tested are listed in Table 2. The term "soft" in the color column indicates a subdued color rather than vibrant. The 60° Gloss column lists the Gloss Meter measurements.

Essentially, the results showed no correlation between the lightness and the 60° gloss measurements, while there was good agreement between the 60° Gloss measurements and the directly measured reflectivity (see Table 3). The specular reflection maxima, as well as the forward scattering angles, are shown in Table 3 for both the s- and p-polarization components.

The table shows larger forward shift values for the p-component than for s due to the steeper increase in reflectivity with angle of the p-component for angles of incidence near 60°. Fitting the forward scattering curves and examination of the profilometer traces showed that the distribution of angles of incidence is relatively narrow (eg. less than 5° at half-max). It should be noted that the forward shift is sufficiently large to significantly increase the range of solar elevation angles which produce direct veiling glare. Thus increases of 10-20 degrees in the solar elevation angles will produce veiling glare long before the sun appears in the line of sight of the driver. Increasing the windshield rake angle to 65° will increase the angular shift of the forward scattering for both the p- and s-components of the reflected light.

Table 2 - Dashboard sample characteristics

Sample #	Color	Texture	60° Gloss	Lightness
1	Gray	Semi-smooth	2.1	40.08
2	Dark Blue	semi-smooth	2.7	26.88
3	Gray	Rough	0.8	39.37
4	Soft Green	Semi-smooth	2.2	35.13
5	Soft Green	Rough	0.6	34.19
6	Currant	Semi-smooth	2.9	27.65
7	Tan	Medium	3	37.01
8	Soft Burgundy	Semi-smooth	1.6	33.46

Table 3 - Specular reflection components

Sample #	s-Angle	s-Value	p-Angle	p-Value	s/p
1	7°	0.87	17°	0.15	5.9
2	7°	1.45	22°	0.17	8.73
3	12°	0.28	17°	0.06	4.4
4	7°	0.66	17°	0.11	6.19
5	12°	0.23	17°	0.06	3.5
6	12°	1.06	22°	0.16	6.54
7	7°	1.16	17°	0.13	9.21
8	7°	0.65	17°	0.12	5.62

CONCLUSIONS

We have examined the components of veiling glare associated with the windshield reflectivity and with the dashboard reflectivity. Clearly, the addition of AR coatings to the inner surface of the windshield is desirable to reduce veiling glare. We have examined the properties of the dashboard materials and how they contribute to veiling glare. It is clear that rough materials will reduce the total scattered light, however, they increase the forward scattering of the specular light which increases the range of solar elevation angles that produce veiling glare. In the opposite case, shiny and smooth surfaces increase the amount of scattered light, while they reduce drastically the range of solar elevation angles that produce veiling glare. In materials selection issues, it is still better to reduce the total amount of scattered light, even though it produces veiling glare over a longer period of time during the day.

Finally, this work has led to an understanding of the relationship between surface roughness and reflectivity, and to the development of a comprehensive simulation model that covers uncoated and coated windshields and any dahsboard cover materials of known roughness for any light condition and any windshield rake angle. The simulation program is easy to use and menu-driven. Work is currently under way to calculate directly the corresponding Veiling Glare Index.

References

1. "GRADED-INDEX AR COATINGS FOR LARGE ANGLES OF INCLINATION," J. H. SIMMONS, S. J. PAGANO, L. K. DOWNIE, B. G. POTTER, C. J. SIMMONS, E. N. BOULOS AND M. F. BEST, J. NON-CRYSTALLINE SOLIDS 178, 166-175 (1994).

Acknowledgements

The authors wish to thank H. Moeckler and C. Schardt for their help with the reflectivity measurements.

970228

Reflection Characteristics of Automobile Interior Materials

Frank P. Nanna
TRICOR Systems, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Automobile interior surface reflections and gloss have presented problems for users and manufacturers alike. Eye-fatiguing glare and windshield reflections can range from minor irritations to major safety concerns. Matching various interior components produced by different organizations can also produce problems.

This paper describes gloss/reflection measurements of samples of materials commonly used by manufacturers. Natural environment and laboratory measurements are performed and their results compared.

The measurement results show that the geometric composition of grain can affect both the amplitude and directionality of the reflection and that current measurement techniques may be inadequate.

BACKGROUND

There are two readily apparent criteria for the surface finish of automobile interiors, *appearance* and *functionality*. Manufacturers attempt to provide a pleasing appearance by utilizing grainy materials to simulate the look of much more expensive leather. They also attempt to match or harmonize all viewable components, including grainy materials produced by different manufacturers, which has created problems. Most often the appearance mismatch is determined by humans and, as such, is subjective.

Functionally, the interior is intended to diminish glare so as to not impair the driver's vision or fatigue his/her eyes. Care must also be taken to avoid geometric configurations that can produce objectionable imaging in the windshield.

The prediction of glare and windshield imaging is a relatively straightforward optical problem as long as the interior surfaces are smooth. The introduction of a grainy surface changes the problem considerably. The nature of the grain can alter not only the amplitude but also the direction of the reflected illumination.

Some manufacturers use coatings in an attempt to control the reflective/gloss characteristics. The coatings are designed to reduce the gloss and possibility of glare. Many manufacturers utilize gloss measurement devices designed to the American Society for Testing and Materials (ASTM) standard D523 to evaluate the effectiveness of the coatings. A discrepancy between the measurements made by the device and a manufacturer's human evaluation of the gloss led to this paper.

OBJECTIVE

The objective of the tests and analyses were to determine relationships between illumination, product orientation and observer viewing angle. Other studies[1] showed a relationship of illumination source size to perceived gloss. This study looks at the geometric relationship in a natural environment. The tests performed were not meant to be all inclusive but rather to determine gross effects that warrant further examination.

DESCRIPTION OF THE TESTS

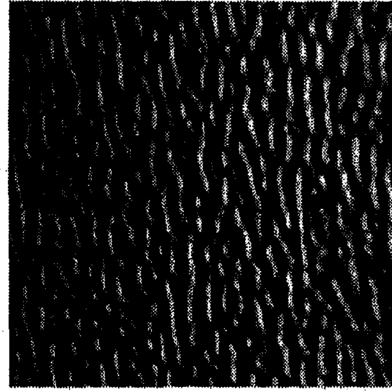
Qualitative and quantitative testing was performed both indoors and outdoors. Measurements were made on five different grain samples commonly used in automobile interiors and a non-grained control sample (Figure 1). Each of six black samples were coated at the same time to control variations in reflectance/gloss. The manufacturer of the material supplied the samples along with gloss measurements at 60° and 85° as measured on an ASTM D523 instrument.

Crude, short-range qualitative tests were conducted that confirmed the manufacturer's opinion of the discrepancy between human and device evaluation of the gloss of the samples.

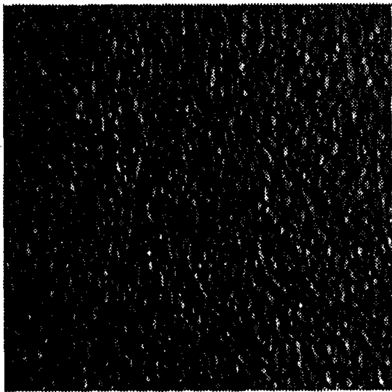
A second psychophysical test was performed with 12 subjects to rank the gloss/brightness of the samples at a distance where the grain could not be perceived. The subjects were viewing the samples perpendicular to a large (1.0 m x 1.25 m) illumination source from a distance of 5.2 m. The height of the subjects' eyes were maintained at a height to provide a 30° elevation look angle. Statistically, the subjects identified the gloss/brightness ranking of the samples consistent with the device measurements. The subjects experienced considerable difficulty in the ranking even though the results appeared consistent. During the course of the test, it was noted that a change in look angle changed the ranking order of the samples. Various pairs of samples reversed their relative gloss/brightness appearance when the observer moved to a different look angle. Sample testing with a photometer confirmed the transition. At that point the decision was made



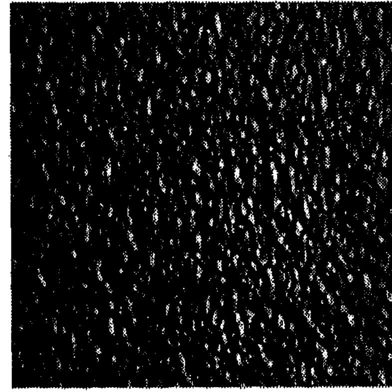
Sample 1



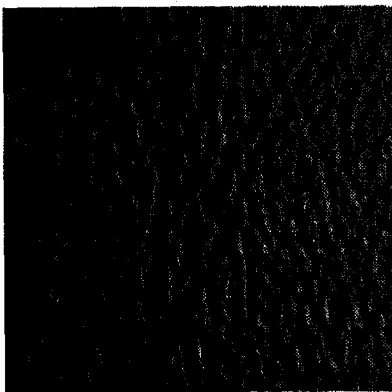
Sample 2



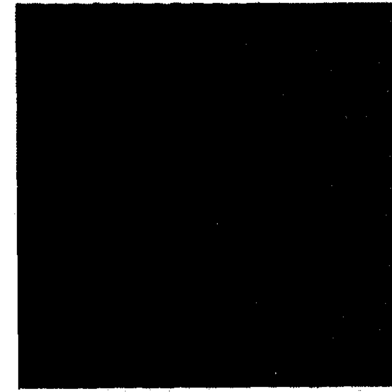
Sample 3



Sample 4



Sample 5



Sample 6

Figure 1 Grain Samples

to collect qualitative data in the natural environment where the material would be used.

Testing was accomplished outdoors with the samples solar illuminated and in shadow. Solar-illuminated measurements were limited to the sun residing in the forward hemisphere to coincide with the most severe environment encountered by a driver and an instrument panel.

Azimuth and elevation measurements were made in accordance with Figure 2. Solar-illuminated measurements were limited to two solar angles. Our experience in optical stealth and camouflage dictated that data on more solar angles is better accomplished by computer simulation[2,3]. A calibrated barium block standard was used to provide a reference brightness measurement with the varying solar intensity levels.

Tests of the samples were conducted in an indoor environment using the large light source and a video camera-based gloss measurement system. The results of these tests and corresponding outdoor tests are compared to the measurements provided by the ASTM D523 device.

SOURCE OF ERROR

Solar location was determined by time and a computer program for the longitude and latitude of the tests, however, a finite time was required to make the measurements. Some variation of results occurred due to the physical location of the photometer field of regard on the sample.

TEST RESULTS

Test 1 (Figure 3) was conducted at a solar elevation angle of 29°. The photometer was at the angle of reflection of 29° and then moved at 15° intervals in azimuth, maintaining the 29° elevation relationship with the sample.

On axis (0° Az) the samples reacted as would be anticipated. The non-grained sample exhibited the highest brightness. All samples exhibited specularity with readings exceeding the barium which has a total hemispherical reflectance approaching 100%. At 15° azimuth all but one of the grainy samples exceed the brightness of the non-grained sample and varied from approximately 19% to 52% of their on-axis values. The trend continued as the off-axis angle was increased for four of the five samples.

Test 2 (Figure 4) was a repeat of Test 1 except the solar elevation and measurement angle were at 50°. It is immediately seen that the sun's angle with respect to the samples significantly changed the specular content of the reflection. The sample measurements were reduced significantly to 25% of their value at 29° in the on-axis condition, although the solar intensity increased approximately 60%. The pattern of diminished brightness approximated the 29° measurements, except the 45° azimuth angle was reached before the brightness levels of four of the five samples exceeded that of the non-grained sample.

Tests 3 and 4 (Figures 5 and 6) were designed to test the brightness sensitivity of the samples with respect to elevation look angle. The two solar elevation angles tested were 30° and 50°. In both cases, the azimuth angle was held constant on axis (0° Az).

The Test 4 samples changed values as a function of look angle. Only sample 1 and the non-grained sample peaked at the 50° reflection angle. The peaks of the other samples varied along with their values relative to each other and the non-grained control sample. The 50° solar and elevation look angle strongly suggested that the various grains exhibit directional reflectivity. The effect is not as obvious but does appear to be present in the 30° readings.

Test 5 (Figure 7) was conducted with the samples in a shady/overcast condition. Only elevation look angle was changed. The results suggest that the grainy surfaces were more capable of distributing surround light in all directions than the non-grained sample.

Test 6 was conducted employing a (1.0 m x 1.25 m) illumination source. The source was constructed with fluorescent lamps covered by diffusing glass. The samples were placed approximately 30 cm in front of the source and

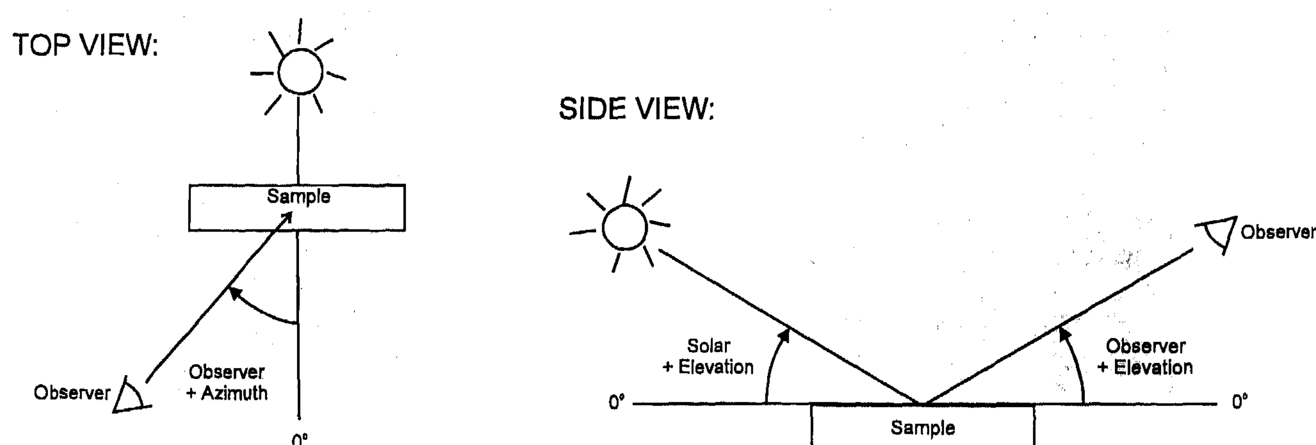
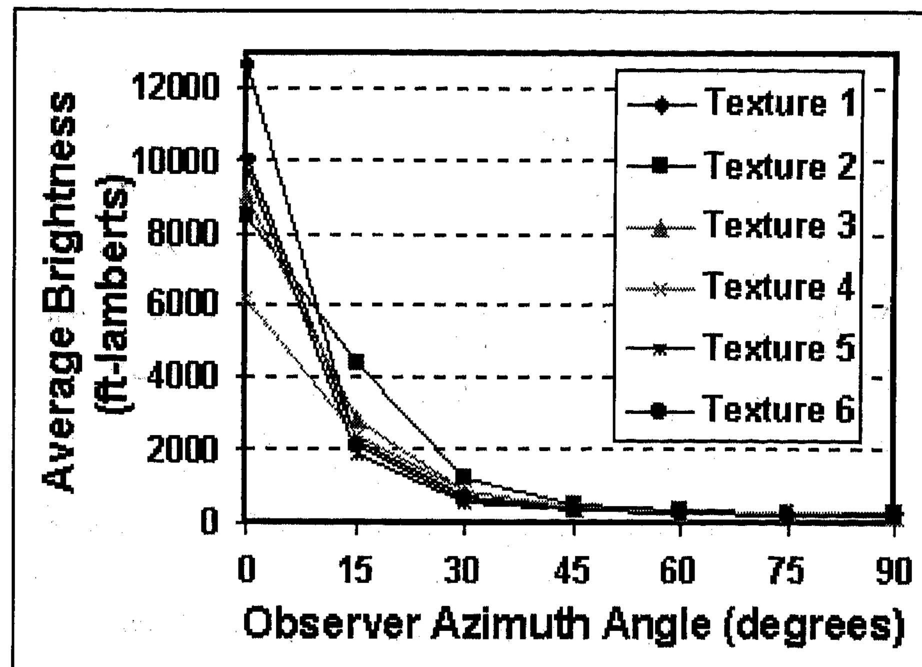


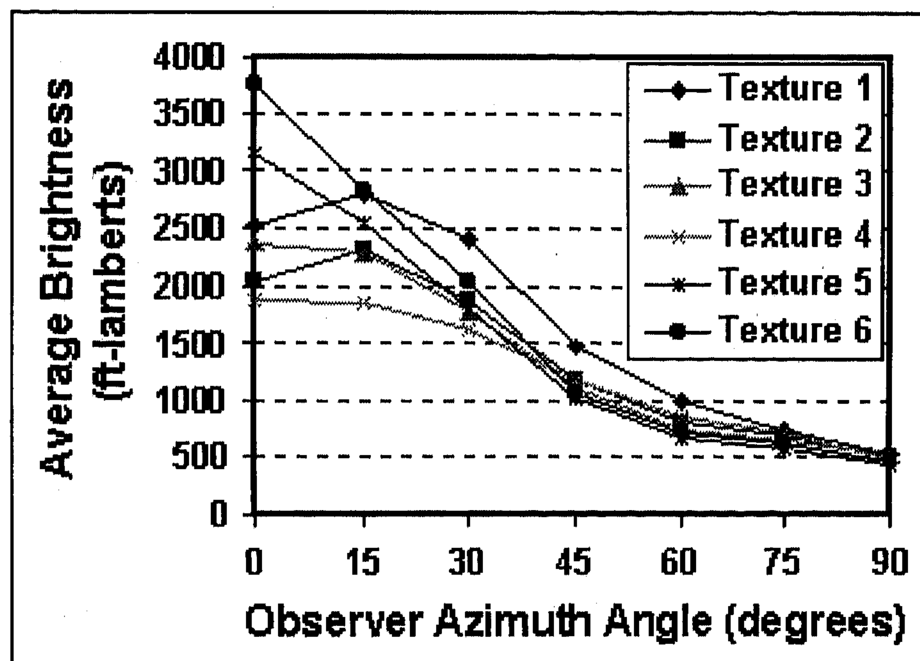
Figure 2 Azimuth and Elevation Geometry



Solar Elevation = 29°   Observer Elevation = 29°   Observer Azimuth = "Angle"   Units = fL

Angle	Barium	#1	#2	#3	#4	#5	Non. Gr
0	5475	10060	8470	9051	6208	9740	12610
15	5505	2250	4370	2850	2450	1840	2075
30	5417	732	1240	830	828	590	625
45	5390	430	520	438	440	350	348
60	5312	322	342	312	317	268	257
75	5236	278	282	270	265	235	234
90	5312	258	267	243	250	220	224

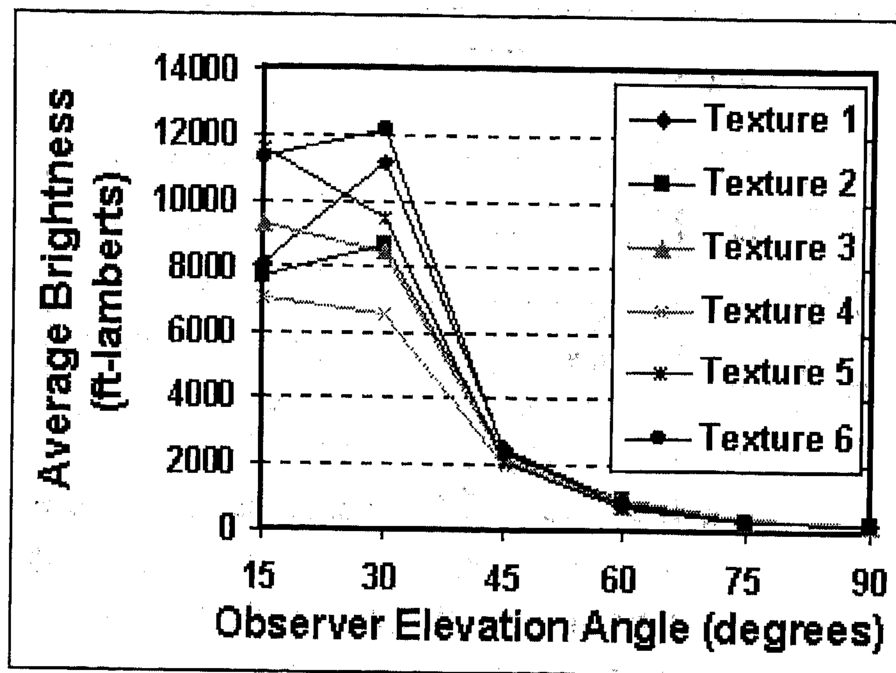
Figure 3 Test 1 Outdoor Azimuth Test



Solar Elevation =50°   Observer Elevation = 50°   Observer Azimuth = "Angle"   Units = fL

Angle	Barium	#1	#2	#3	#4	#5	Non. Gr
0	8368	2541	2024	2374	1871	3175	3742
15	8385	2810	2331	2290	1852	2545	2835
30	8368	2425	1865	1775	1635	1804	2039
45	8435	1479	1182	1116	1172	1020	1073
60	8377	985	815	737	860	651	715
75	8362	750	667	627	715	566	613
90	8438	518	498	515	477	427	446

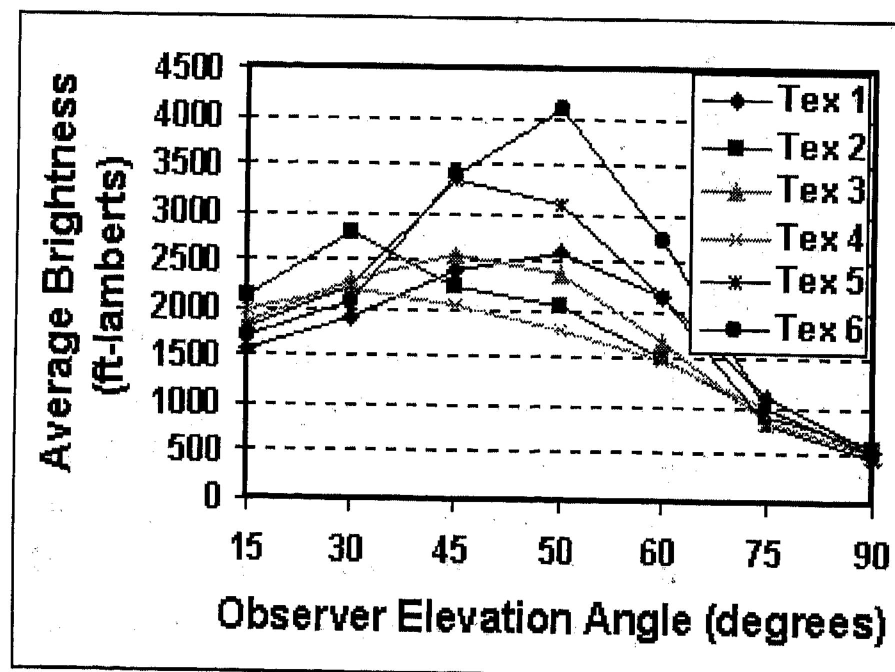
Figure 4 Test 2 Outdoor Azimuth Test



Solar Elevation =30° Observer Elevation = "Angle" Observer Azimuth = 0° Units = fL

Angle	Barium	#1	#2	#3	#4	#5	Non. Gr
15	5183	8053	7699	9308	7002	11600	11340
30	4918	11190	8701	8441	6584	9512	12190
45	4927	2432	2292	2105	1940	2051	2388
60	4973	834	973	770	921	702	815
75	5008	310	351	309	380	311	345
90	5116	233	242	235	265	235	240

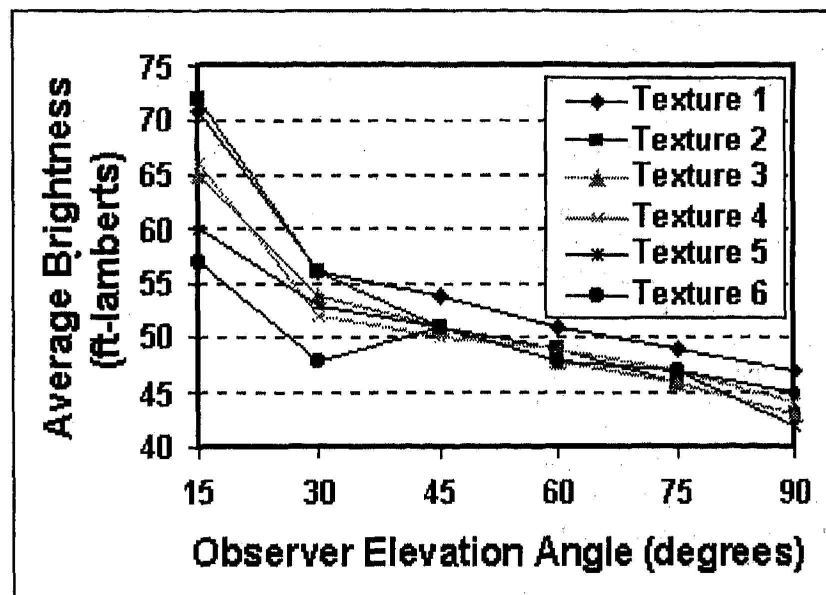
Figure 5 Test 3 Outdoor Elevation Test



Solur Elevation =50° Observer Elevation ="Angle" Observer Azimuth =0° Units = fL

Angle	Barium	#1	#2	#3	#4	#5	Non. Gr
15	8155	1573	2121	1859	2002	1807	1713
30	8172	1923	2795	2300	2204	2244	2064
45	8155	2402	2226	2570	2040	3330	3386
50	8172	2621	2036	2367	1813	3094	4104
60	8163	2162	1499	1695	1491	2166	2743
75	8339	1116	910	850	1010	831	992
90	8371	536	589	507	543	459	508

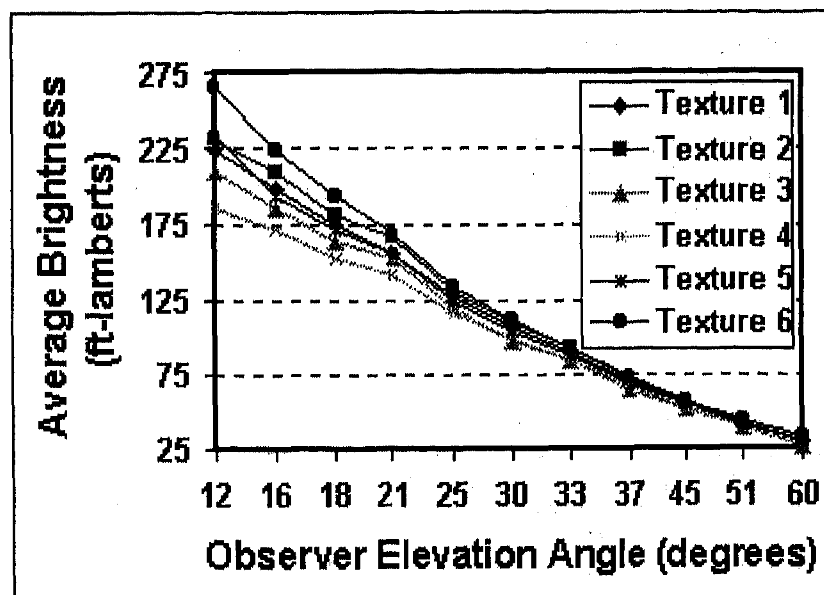
Figure 6 Test 4 Outdoor Elevation Test



Solar Elevation =N/A (Shade) Observer Elevation ="Angle" Observer Azimuth =0° Units = fL

Angle	Barium	#1	#2	#3	#4	#5	Non. Gr
15	611	71	72	65	66	60	57
30	560	56	56	54	52	53	48
45	572	54	51	51	50	51	51
60	577	51	49	48	49	48	48
75	581	49	46	46	47	47	47
90	569	47	43	43	44	42	45

Figure 7 Test 5 Outdoor Shade Elevation Test



Solar Elevation = "Angle" Observer Azimuth =0° Units = fL

Angle	Barium	#1	#2	#3	#4	#5	Non. Gr
12	272	223	231	210	186	235	266
16	270	199	210	186	171	194	223
18	269	176	182	163	152	171	194
21	265	156	167	152	141	156	169
25	255	128	131	120	115	124	133
30	247	105	109	96	98	102	111
33	242	88	92	84	84	87	92
37	236	70	72	64	67	68	73
45	231	54	55	51	52	54	56
51	231	42	42	39	41	41	44
60	231	29	29	27	29	30	32

Figure 8 Test 6 Large Illuminator Test

measured with a photometer. The illuminator panel provided a relatively uniform source at an approximate overall source size of 120° in azimuth and 70° in elevation although the effective size was less. The data shown in Figure 8 is relatively benign.

Test 7 employed a TRICOR Model 801A Surface Analysis System. The device uses a video camera capable of providing 240,000 data points on a sample which is illuminated by an extended (18 cm x 23 cm) illumination source. The device is calibrated to the same black glass plate used to calibrate the ASTM D523 device and is assigned a gloss level of 255. The unit is capable of dividing the levels of gloss into selectable bands based on a percentage of sample points. The data shown in Figure 9 shows the average value of the 20% highest and 100% (average value) of gloss data points. This selectability allows a simulation of observer viewing distance under various lighting conditions. For example, at short range the observer can resolve the grain and gloss projected from that grain, whereas at longer range the individual grain variations are not resolvable due to the resolution limits of the eye.

Utilizing some percentage of the brightest data point in effect shows the potential capability of the surface gloss/reflection. A small illumination source cannot produce the light rays over an angle sufficient to produce a reflection into the camera's field of view due to the normals of a grainy surface.

Figure 10 shows gloss measurements from an ASTM D523 device. The device incorporates a small collimated light source and a photodiode focused on the surface reflection of the source from a surface. Calibration of the unit is accomplished by assigning a value of 100 to a measurement of black gloss standard of a prescribed index of refraction.

The ASTM D523 device was designed to measure the gloss of flat samples and makes no other claim. Its use for measuring curved or textured surfaces is highly suspect due to its physical construction. Figure 11 shows its proper use and the effects of curved and irregular surfaces. The intent is that the collimated light emitted by the source is collected by the relatively small photodiode. When the surfaces are curved or irregular, the reflected ray's directions are changed, reducing the number collected. The figures at the right are intensity maps of the surface of the photodiodes for the different conditions.

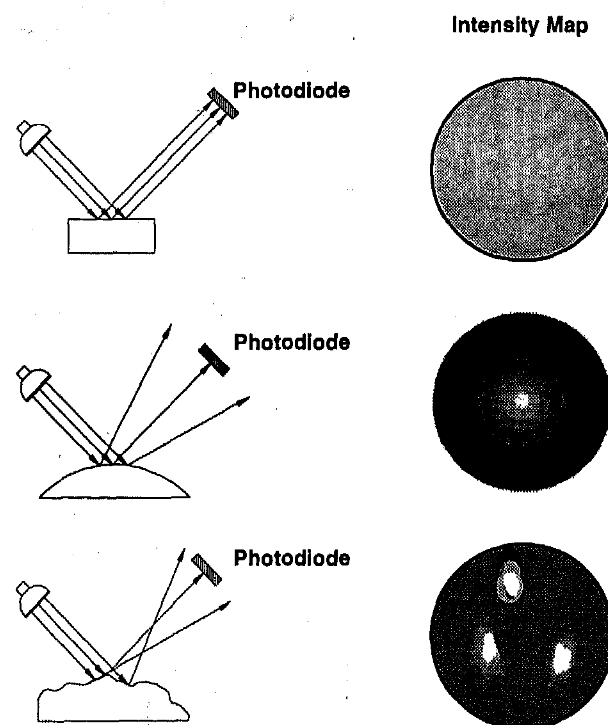


Figure 11 Measurement Variances Using Photodiodes

Light Source Elevation = 33° Sensor Elevation = 33°

Group	#1	#2	#3	#4	#5	Non. Gr
20%	127	124	111	101	125	125
100%	107	104	99	82	112	121

Figure 9 Test 7 Video Camera Test Results

Light Source Elevation = Sensor Elevation = Angle

Angle	#1	#2	#3	#4	#5	Non. Gr
30 deg	3.8	3.0	3.5	2.5	4.4	5.7
5 deg	8.2	3.4	5.8	3.3	11.5	24.7

Figure 10 ASTM D523 Device Measurements

MEASUREMENT COMPARISONS

Table 1 shows a comparison of 30° on-axis measurements of gloss and brightness outdoors solar illuminated (Tests 1 and 2) and shady test (Test 5), indoors with the ASTM D523 device, the large illuminator (Test 6) and the 100% and 20% values of the video camera system (Tests 7 and 7a). All values were normalized to the non-grained sample reading for that test to allow a comparison.

The data clearly shows that the grainy sample's brightness/gloss with respect to the non-grained sample directly increases with the effective illumination source size confirming our previous report[1]. It is obvious that a greater number of the surfaces provide reflection to the sensor as the source size is increased. The effect is maximized when the entire hemisphere becomes the source and the grainy surfaces equal or exceed the smooth non-grained surface.

SUMMARY OF TEST RESULTS/CONCLUSIONS

The testing showed that grainy surface reflections/gloss are significantly influenced by the environment, the geometric composition of the grain and the viewing angle of the observer. The data indicates that the geometric composition of the grain affects the directionality of the reflection from the surface.

A lower solar elevation angle (<50°) reflected brightness levels can reach objectionable levels. On-axis levels can exceed 10,000 Foot Lamberts (fL) at low sun angles which is approximately the value of a cloud edge in close proximity to a high noon sun. Brightness levels exceeding 2,000 fL are prevalent even with the sun off-axis at typical driver viewing angles. The 2,000 fL corresponds to the horizon on a bright day or the sun reflecting from the desert.

The data suggests that gloss-reducing coatings may require a more complex analysis to achieve their objectives on grainy surfaces. The impact of the findings on the automobile interior manufacturing industry appears to be the following:

1. Appearance and functionality should not be independent parameters in designing the grain employed in interiors in general and on the top of instrument panels in particular. The directional reflection of various grains can increase susceptibility to glare or windshield imaging.
2. If harmony is defined by grainy surfaces being produced by different manufacturers being geometrically alike and presenting the same brightness, it can be accomplished at the uninstalled level on identically shaped material. Once the geometry of the surface or its spatial orientation is changed, the perceived brightness of the various parts will change with the environmental conditions.
3. The ability to reproduce and match grainy materials manufactured by different entities cannot be accomplished by using an ASTM D523 device. The device was designed to measure flat surfaces and performs well in that capacity. To ensure reproducibility and matching materials, the geometric characteristics of the grain along with the gloss need to be monitored.
4. Optimum functionality of interiors from a glare and imaging aspect would appear to be most easily accomplished by computer simulation. Allowable variations in basic interior geometry in combination with grain directionality could be rapidly assessed under all pertinent environmental conditions and viewer aspect angles.

REFERENCES

1. Nanna, Frank P., "Variables Affecting Human Perception of Gloss," Inside Automotives, Sept 1996.

2. Akerman, A., Conrad G. et al, "Camouflage Handbook (U)," Air Force Wright Avionics Laboratory (AFWAL) AFWAL-TR-86-1028, April 1986.

3. Allen, P. and Conrad, G., "Signature Analysis Software," TRICOR Document No. TSI 557-101A, Sept 1990.

ABOUT THE AUTHOR

Mr. Frank Nanna is a Senior Engineer at TRICOR Systems Inc., Elgin, IL. After receiving his BSME from Purdue University in 1984, he worked briefly for Northrop Defense Systems Division in the Infrared Countermeasures Department. Since joining TRICOR in 1985, Mr. Nanna has been involved in visual stealth and camouflage programs conducting and participating in numerous field and psychophysical tests and developing sophisticated data analysis and digital image processing software. As the principal designer, he is a co-holder of patents on TRICOR's Model 801A Surface Analysis System.

Table 1 30° Comparison On-axis Measurements

	Gloss Meter	Outdoors		Shady	Large Illuminator	Video Camera System	
Sample No.	D523	Test 1	Test 2	Test 5	Test 6	Test 7	Test 7a
1	66.6	79.7	91.2	116.7	94.6	88.4	100.2
2	52.6	67.1	71.4	116.7	98.2	85.9	99.2
3	61.4	71.7	69.2	112.5	86.4	81.8	88.8
4	43.8	49.2	54.0	108.3	88.3	67.8	80.8
5	77.1	77.2	78.0	100.0	91.9	92.5	100.0
NO GRAIN	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Barium = 247							

970229

Considerations on a Feeling of Troublesomeness Regarding Automotive Head-Up Displays During Driving

Kazumoto Morita, Jinichi Mashiko, and Takeo Okada
Traffic Safety & Nuisance Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The authors have carried out two types of tests regarding head-up displays (HUDs) for the purpose of determining the relationship between display position and a feeling of troublesomeness for the driver.

The first test was a basic experiment to evaluate how troublesome the HUDs were to an observer sitting in a chair outdoors. The collected data was used as the basis for carrying out approximation by means of two-dimensional normal distribution.

In the second test, the observer was actually in a car and a mock-up HUD making use of clear film was used to carry out the experiment in investigating how the feeling of troublesomeness changed depending on whether the vehicle was in a stop state or in a moving state. Compared to a stop state, the results in a driving state showed a shift toward the left side for the distribution of a feeling of troublesomeness. The reason is probably because the distribution of the line of vision of the driver moves toward the left when driving in Japan.

INTRODUCTION

One method of transmitting information to the driver when driving a motor vehicle is via head-up displays (HUDs) which display vehicle speed on the windshield. The devices have been mounted on market-sold vehicles since around 1988[1,2]. Up to now, HUDs have been used mainly to indicate vehicle speed. More recently, however, there have been proposals to use HUDs as a display device for navigation systems[3,4]. In the coming days, it is expected that HUDs will be employed for the purpose of displaying a variety of information.

When it comes to the technology of HUDs, the situation is different from that in jet fighters where professional pilots with skills operate the equipment. In automobiles the users are ordinary drivers. That means the equipment must be easy to read and use. Considering how one of the characteristics of the HUDs is that they overlap with the scenery in the front of the vehicle, an important question is the position on the windshield where the HUDs are located.

The display position of a HUD must be determined for a variety of reasons. We might readily assume that, in order for the driver to quickly read and understand the display contents, the ideal HUD position would be directly in front of the driver. Indoor tests carried out by Okabayashi and others have studied accuracy in simultaneously apprehending the information regarding the scenery in front of the car and the displayed information according to the display position[5]. Results show that there is a high accuracy rate for grasping information in a position directly in front of the observer and in a range of vision within several degrees.

Nevertheless, when actually driving a vehicle, if HUDs are directly in front of the driver they overlap with the scenery to the front and are perceived as troublesome by the driver. In considering comprehension of the information, it is naturally desirable that HUDs be located as much as possible in front of the driver. However, considering the feeling of troublesomeness to the driver, it is also important to locate the device at a position that is somewhat removed from a position directly in front of the driver.

For these reasons, the authors carried out two types of experiments to determine the extent to which the driver experienced a feeling of troublesomeness depending on the display position of a HUD. Of the two types of tests, the first was a basic experiment in which the HUD was displayed on plate glass on a desk and the observer sat in a chair in order to evaluate the feeling of troublesomeness. In the other type of test, the observer rode in a compact vehicle and there was evaluation both in a stopped state and in a driving state in order to evaluate the extent of a feeling of troublesomeness.

METHOD OF BASIC EVALUATION TEST

EXPERIMENTAL METHOD - An observation test was carried out at a test course located in Kumagaya City, Saitama Prefecture in Japan. Figure 1 shows the overall test conditions. An observer seated in a chair (right side of figure) observed a HUD image displayed on a flat display glass located on a desk. The person conducting the experiment (left side of figure) changed the display position of the HUD in carrying out the experiment. The observers placed their chins on a chin rest (height: 1.00 meters) to define the position of the head. In addition, a compact car was located at a point about 50 meters in front of the observer and observers were instructed to look in that direction.

As Figure 2 shows, there were 20 different display positions. The point directly in front of the observer was considered the origin point. The direction was exposed as the angle of vision. As for the horizontal direction, there were a total of 5 different positions ranging from 6 degrees to the left to 6 degrees to the right at intervals of 3 degrees. As for the vertical direction, there were a total of 4 different positions ranging from 6 degrees in a downward direction to 3 degrees in an upward direction at intervals of 3 degrees. Regarding the angle in a horizontal direction the right side was plus. Regarding the angle in a vertical direction, the upper side was considered plus. The experiment was designed so that the center of the HUD image was displayed in the display position shown on the coordinates of the horizontal and vertical directions.

Figure 3 shows an example of display conditions (display position located 6 degrees to the left and 6 degrees downward). As Figure 3 shows, the HUD image (created by using a fluorescent display tube) is the number 88. The color of the display image was green (chromaticity coordinates: x=0.242, y=0.436) and the display luminance measured in the direction of observation was 1600 cd/m^2 (all actual measurement values measured in dark room). The distance from the observer's eye to the HUD image was about 1.2 meters. The general size of the HUD image of the market-sold vehicles was used as a reference in making the vertical angle vision of the display image (88) seen by the observer to be 0.7 degrees and the horizontal angle of vision to be 1.2 degrees. The display plate glass was a plate with a reflectivity of 20% after coating the surface with oxidized titanium. This was set at an incline of 31 degrees from the horizontal plane. The transmissivity of the plate glass from the observer's direction was 75%.

In the three-day test carried out in the present experiment, the weather was both fair and cloudy. The horizontal illuminance measured by facing the receptor of the illuminance meter towards the zenith was about 5,000 to 80,000 *lx*. The road surface luminance measured at a position about 50 meters to the front of the observer was in a range between about 200 and 3,700 cd/m^2.

Regarding the order of the experiment, for each display position the HUD image was shown for 10 seconds to the observer. The observer was then asked to give a

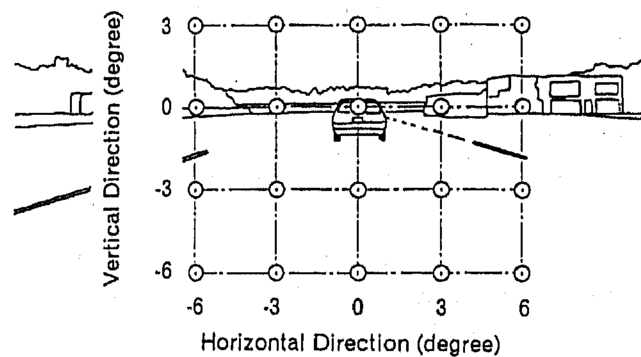


Fig.2 Display positions of the HUD



Fig.1 Basic evaluation test scene
-an observer (right side) and an examiner (left side)-

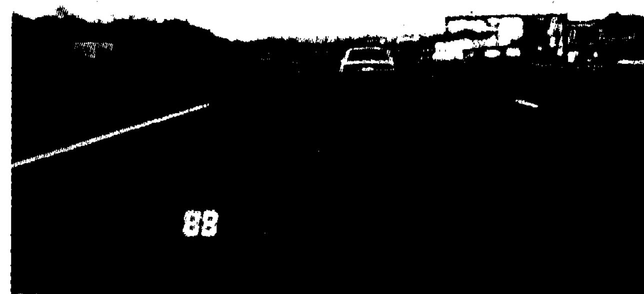


Fig.3 An example of HUD image

subjective evaluation of a feeling of troublesomeness according to the integers from 0 to -4 as shown in Table 1. (It is admitted by the authors that how they feel depends on their own judging scale.) The 20 types of display positions were selected at random and evaluation was carried out two times for the same display position.

There were a total of 15 observers with normal color vision and with an average age of 34.1 years (7 observers in their 20's, 3 observers in their 30's, 4 observers in their 40's, 1 observer in his 50's). That meant there was a total of 30 evaluation data obtained for a single display position.

Tab. 1 Rating Scale for a Feeling of Troublesomeness

Evaluation value	Meaning
0	Not troublesome
-1	Slightly troublesome
-2	Rather troublesome
-3	Very troublesome
-4	Extremely troublesome

EXPERIMENTAL RESULTS - Of the evaluation standards over 5 levels regarding a feeling of troublesomeness, when it came to determining the extent to which a position was troublesome to the observer, various judgments can be made. In this report, we will consider the total of the answers including "rather troublesome," "very troublesome" and "extremely troublesome" in order to examine the feeling of troublesomeness. Speaking conversely, we can say regarding the evaluation results "not troublesome" and "slightly troublesome" that the observers did not have any problem in this respect.

If we obtain the total of the above three types of replies for each display position, we obtain a figure like that in Figure 4. At the origin point, the total for the three replies of "rather troublesome," "very troublesome" and "extremely troublesome" was equal to the total number of 30 cases.

Examining the distribution of this feeling of troublesomeness we see that the feeling was greatest at the origin point and that the feeling decreased as one moved either to the right and left or up and down.

The evaluation results are discrete for the display positions determined for every 3 degrees of the angle of vision. This data is used as the basis in attempting to approach the distribution of a feeling of troublesomeness by means of a two-dimensional joint normal distribution as shown in the following formula (EQ (1))[6]. Here, the horizontal direction is the x axis and the vertical direction is the y axis.

The five parameters included in the above formula are obtained as follows.

$$\left.\begin{aligned} &E[x] = \mu_x \\ &E[y] = \mu_y \\ &E\left[(x-\mu_x)^2\right] = \sigma_x^2 \\ &E\left[(y-\mu_y)^2\right] = \sigma_y^2 \\ &E[(x-\mu_x)(y-\mu_y)] = \sigma_x\sigma_y\rho_{xy} \end{aligned}\right\} \cdots\cdots(2)$$

(E[•] means an operation of deriving the expected value.)

If we use the evaluation results as the basis for calculation, these parameters are as follows:

μ_x=0.03
μ_y=0.32
σ_x=3.56
σ_y=1.89
ρ_{xy}=0.006 .

The average of x and y corresponds basically with the origin point.

Remaining now is the question of whether the two-dimensional normal distribution approximated with EQ (1) is fitting as the actual distribution of evaluation results. Thus, it is possible to verify this problem by carrying out an χ^2-test[7] regarding the level of compatibility. The total of the replies for "rather troublesome," "very troublesome" and "extremely troublesome" becomes the observed frequency fi (i=1 to 20). The total value for the observed frequency fi was 187 cases. In addition, two-dimensional normal distribution is used as the basis for deriving $z(x,y)$ for the 20 display positions by means of EQ (1). The

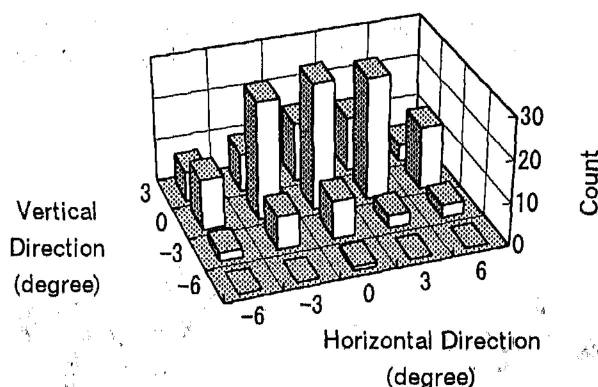


Fig.4 Total of replies including "rather troublesome", "very troublesome" and "extremely troublesome"

$$z(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\exp\left[-\frac{\sigma_y^2(x-\mu_x)^2 - 2\sigma_x\sigma_y\rho_{xy}(x-\mu_x)(y-\mu_y) + \sigma_x^2(y-\mu_y)^2}{2\sigma_x^2\sigma_y^2(1-\rho_{xy}^2)}\right]\cdots(1)$$

expected frequency Fi (i=1 to 20) for each position is obtained from this value so that the total is the figure of 187 cases as mentioned above. The observed frequency and expected frequency are used to obtain the value of X^2 by means of the following equation.

$$X^2 = \sum_{i=1}^{k} \frac{(fi - Fi)^2}{Fi} \cdots\cdots (3)$$

The distribution X^2 is χ^2 distribution in terms of approximation. By comparing the two it is possible to check the level of compatibility. If the value for X^2 as shown in EQ (3) is obtained from the present data, X^2=17.47. On the other hand, regarding the level of freedom of χ^2, the constraint conditions in order to obtain the expected frequency are 6 items (the 5 constraint conditions generated to derive the prior five parameters and the single constraint condition generated by the total number of data). The figure of the level of freedom is thus 20-6=14. If the level of significance is 5%, the value is χ^2=23.68. A comparison of the values for X^2 and χ^2 shows that $X^2 < \chi^2$. As a result, we could confirm that it is possible by means of the above two-dimensional normal distribution to approximate the actual distribution of evaluation results.

Figure 5 shows this two-dimensional normal distribution in a range from -10 degrees to 10 degrees in a horizontal direction and from -6 degrees to 6 degrees in a vertical direction. In order to obtain generality, there is expression in terms of the ratio of replies rather than the number of replies regarding the z axis. In other words, the total of the replies for "rather troublesome," "very troublesome" and "extremely troublesome" for each display position is expressed as the ratio after dividing by the total replies for those display positions. In Figure 4, the total of the replies for the above three evaluation values at the origin point was basically equal to the total of 30 cases. Taking this into consideration, we see that the maximum value for axis z in Figure 5 is 1.0. The figure also shows the contour lines at an interval of 0.2 regarding the feeling of troublesomeness. Figure 6 shows the same two-dimensional normal distribution as a plan view. In this case, too, there is display of contour lines at an interval of 0.2 regarding the feeling of troublesomeness. In Figure 5 and Figure 6 the range from 0 to 0.2 for a feeling of troublesomeness is shown by the gray section. In other words, this shows the range in which the total for the three evaluation results of "rather troublesome," "very troublesome" and "extremely troublesome" does not exceed 20% in relation to all replies. Stated otherwise, it shows the range where the total of evaluation results for "not troublesome" and "slightly troublesome" exceeds 80%.

Looking at the results of computing with EQ (2), we see that the correlation coefficient ρ_{xy} of x and y is almost zero. As a result, this is considered separate regarding x and y. In other words, if we obtain the evaluation results independently in a horizontal and vertical direction, it is possible to estimate the two-dimensional normal distribution for the whole. If we use the present evaluation results as the basis, we obtain the following. That is, if we obtain the average and standard deviation from the measurement evaluation results on a horizontal line at the position y=0, the values are as follows: μ_x=0.18, σ_x=3.62. And if we obtain the average and the standard deviation in the same way only from the measurement evaluation results on a vertical line at position x=0, the values are as follows: μ_y=0.17, σ_y=2.13. Considering the scattering of data in the evaluation test, we assume that it is possible to obtain results that are basically the same as the values obtained from 2-dimensional total evaluation as in EQ (2). Given the above, if we obtain the distribution of a feeling of troublesomeness in the representative horizontal and vertical directions in the evaluation test, we can estimate the 2-dimensional distribution and thus achieve efficiency in data gathering during the experiment.

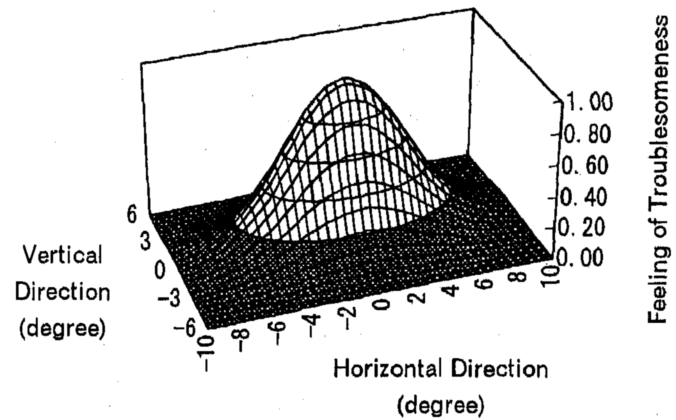


Fig.5 Distribution of a feeling of troublesomeness approximated by two-dimensional normal distribution

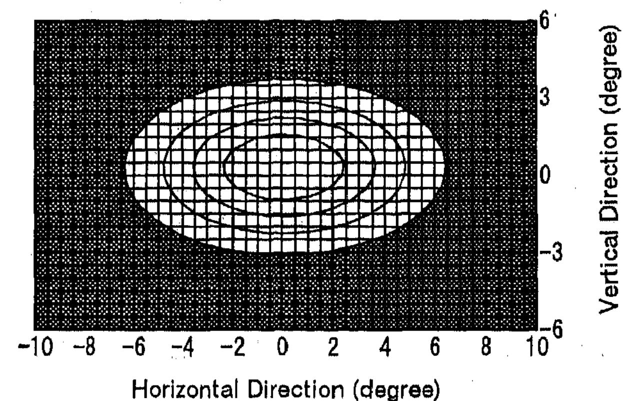


Fig.6 Plan view of two-dimensional normal distribution regarding a feeling of troublesomeness

As for the coordinates of the display positions analyzed above, they correspond to the central position of the HUD image when the display image has a value of 0.7 degrees in the direction of height and 1.2 degrees in a sideways direction. We can expect that, if the size of the display image changes, the results of evaluating a feeling of troublesomeness will also change.

EFFECT OF HEIGHT OF EYEPOINT OF DRIVER

Up to now, the discussion was based on an experiment in which the chin of the observer is supported by a chin rest so that the eye position of the observer is fixed. In actual life, however, people with different body types and driving positions are driving vehicles, which means the eye position changes. Regarding the HUD image, if one ignores the optical distortion of the display device, the position of the image does not change. As a result, depending on the position of the driver's eyes, the angle of the direction in which the image can be seen will change.

Considering how the HUD device is located in the vehicle instrument panel, the cases are more numerous in which HUDs are displayed in a downward direction in relation to the driver. Thus, we will consider the problems when HUDs are displayed in a downward direction vertically. Subsequent statements concerning the angle of vision will be expressed as the depression angle. If the depression angle is large, it means that the angle is large in a minus direction when using the method of expression used up to now for angle of vision.

If the HUD position is established according to the standard eyepoint (this point will be described later), the depression of angle in relation to the HUDs will increase for persons with an eyepoint higher than the standard eyepoint. Conversely, for persons with a lower eyepoint than the standard eyepoint, the depression angle will be smaller (Figure 7).

This being so, regarding a feeling of troublesomeness in the case of observers with a lower eyepoint than the standard eyepoint, the HUD image will become visible upon approaching the origin point directly to the front so that a feeling of troublesomeness will be felt to increase. The apparent change in the HUD position according to the eyepoint height is not only related to the height of the eyepoint but also to the horizontal distance from the eyepoint to the HUD image.

In order to investigate the change in the depression angle according to the eyepoint height, there is use of an "eyellipse"[8] which describes the elliptical shape of the driver's eye range. The conditions of the eyellipse used are as shown below when using the values of a normal compact vehicle as reference. This is a 95th percentile side view when the normal driving and riding seat track travel is 165 mm (Figure 8). The back angle is 25 degrees. The eyellipse setting method is as determined in ISO4513. Under these conditions, in this paper the standard eyepoint position is determined to be the intersection of the datum lines x-x and z-z in the eyellipse side view. In other words, the point located at 0.635 meters in a vertical direction above the seating reference point is the standard eyepoint position.

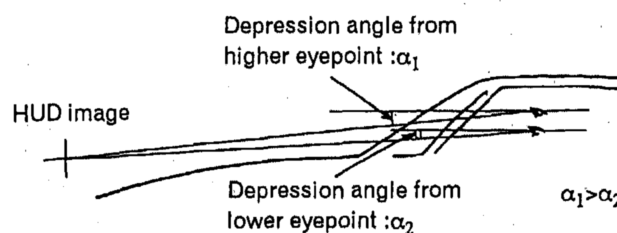


Fig.7 Change in depression angle owing to height of driver's eyepoint

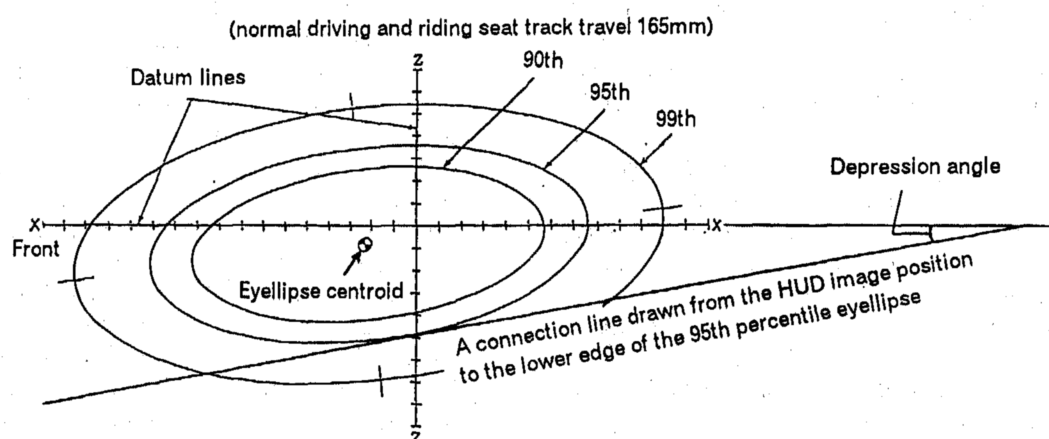


Fig.8 Side view of eyellipse template

The following is a discussion of the HUD image seen from depression angles of 4 degrees, 5 degrees, 6 degrees and 7 degrees from the standard eyepoint. A connection line is drawn from the HUD image position to the lower edge of the eyellipse of the 95th percentile. The horizontal distance from the standard eyepoint to the HUD image is changed from 0.8 meters to 4 meters.

Figure 9 shows the change in the angle (depression angle) of this connecting line from the horizontal line. As this shows, when viewed from a low eyepoint, the HUD image can be seen at the position of the depression angle that is smaller than the depression angle from the standard eyepoint. If the distance from the standard eyepoint to the HUD image is close, the depression angle is especially small. In other words, it is sensed as close to the position directly in front of the driver.

For example, regarding a display in which the depression angle is 5 degrees and the HUD image position is at a distance of 1 meter when seen from the standard eyepoint, if this is examined in terms of an eyellipse of the 95th percentile, it can be seen at a position where the depression angle is 2.1 degrees. To consider matters in terms of an eyellipse of the 95th percentile means that if a plane seen as a straight line in the side view is drawn tangent to the lower edge of the eyellipse, then 95% of the driver eye locations will be above the line and 5% of the driver eye locations will be below the line. That is 5% of the drivers with a lower eyepoint are viewing the HUD image at a depression angle that is even smaller than the depression angle of 2.1 degrees. Thus, considering the feeling of troublesomeness, although there is almost no problem for drivers with a standard eyepoint (Figure 5, Figure 6: The feeling of troublesomeness is 0.02 at a point located at 0 degrees in a horizontal direction and 5 degrees in a downward direction), for drivers with a low eyepoint this will be felt as troublesome. (In Figures 5 and 6, the

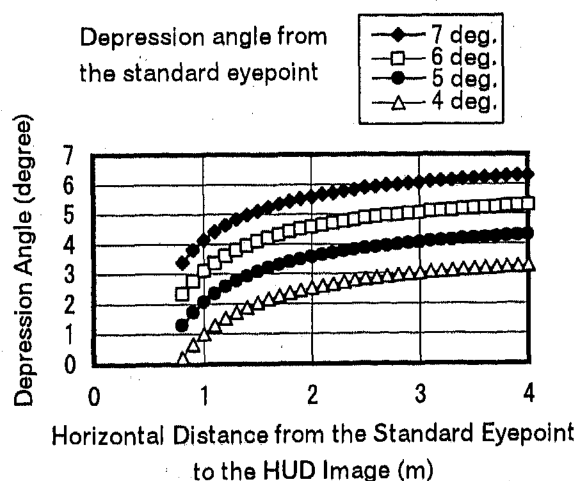


Fig.9 Change in depression angle when viewd from low eyepoint (in terms of the 95th percentile eyellipse)

feeling of troublesomeness is 0.47 at a point located at 0 degrees in a horizontal direction and 2.1 degrees in a downward direction).

As the above shows, even in cases where the eyepoint is low, in order to decrease change in the depression angle in relation to the HUD image, it is better to display the imaginary HUD image at a position far removed from driver and directly to the front. Nevertheless, there have been reports of a feeling of strangeness among drivers if it is displayed too far away[9]. For this reason it is desirable to display the image at an appropriate position. Considering Figure 9, we can consider it desirable to display the HUD image at a position that is about 1.5 meters or more distant from the driver so that marked changes in the angle of depression do not occur.

On the other hand, we can assume that in the case of drivers with a low eyepoint, if there is a structure in which the HUD position can be adjusted to a low position, this can act as one method of reducing changes in the angle of depression.

EVALUATION TESTS CARRIED OUT WHILE DRIVING A VEHICLE

EXPERIMENTAL METHOD - In the basic evaluation test, the observer sat in a chair and observed a HUD on a plate glass. In the present experiment, the observer rode in an actual vehicle (compact car) to evaluate the HUDs. The vehicles used in the test included the observation car in which the observer was riding and a vehicle located at a position about 50 meters in front. The vehicle to the front was used as a target when the observer was looking forward.

Because it was difficult to install an actual HUD at an arbitrary position in the vehicle, a mockup HUD using clear film was attached to the windshield.

The standard eyepoint of the observation vehicle was set as described above. In other words, a point located 0.635 meters above the seating reference point of the driver's seat in a vertical direction was established as the standard eyepoint.

As for the display position, the point directly before the standard eyepoint was made the origin point. In a horizontal direction (on the x axis) this was expressed in terms of the angle of vision from the standard eyepoint in establishing 7 points from -15 degrees to 15 degrees at intervals of 5 degrees. In a vertical direction (on the y axis), there were 5 points ranging from -6 degrees to 6 degrees at intervals of 3 degrees (Figure 10). In addition, as a reference, there was addition of the following four points: (x = -5 degrees, y = -3 degrees), (x = -5 degrees, y = 3 degrees), (x = 5 degrees, y = -3 degrees), (x = 5 degrees, y = 3 degrees). Because the origin points overlap in the horizontal and vertical direction, the total number of display positions was 15 points.

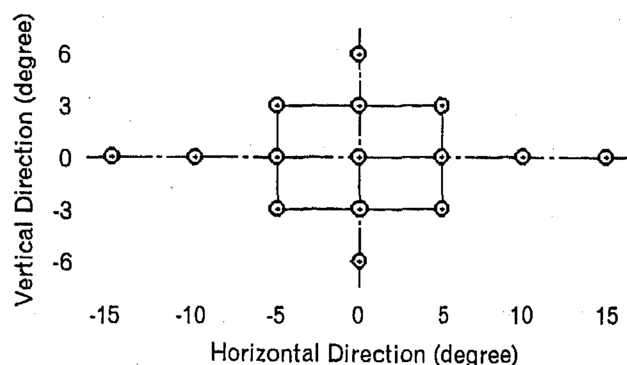


Fig.10 Display positions of the mockup HUD

As was the case already, the HUD image was a green figure 88. Regarding the size of the numbers, at the origin point directly to the front, the size was approximately 0.8 degrees in a vertical direction and approximately 1.3 degrees in a horizontal direction.

In the present experiment, because clear film attached to the windshield was used as a substitute, there were the following differences compared to an experiment with an actual HUD.

(1) Regarding the mockup HUD used this time, it was not a case of bright display as is true for an actual HUD. Instead, the numbers appeared darker than the background. The transmissivity of the green numerals on the clear film was 47%.

(2) By attaching the mockup HUD to the windshield, the distance from the standard eyepoint to the display position is shorter. For example, the distance from the standard eyepoint to the origin point display on the windshield is approximately 0.56 meters. With a normal HUD, the distance to the display position is generally longer.

(3) The windshield is at a slant in relation to the driver and also has a complex curved configuration. As a result, the distance from the standard eyepoint to the display position on the windshield changes in relation to the individual display positions. For this reason, if the size of the numerals on the clear film is the same, the display size seems to change somewhat according to the display position. In the present experiment there was use of two types of mockup HUDs for which the size on the clear film was slightly different. This was done so that the apparent difference in display size according to the display position would not be that pronounced.

There was the following additional problem with the sitting position. Regarding the up-down direction of the observer's eye position, the height of the driver seat was adjusted so that the windshield origin point height would agree with the position of the driver's head in the vehicle located 50 meters to the front. In a horizontal direction, there was adjustment of the sitting position of the observer in a left-right direction so that the position of the origin point would match the position of the driver in the vehicle to the front. (Because the distance from the eyes of the observer to the mockup HUD is short, if the observer focuses on the vehicle to the front, the two images of the mockup HUD in the left and right eyes of the observer will be seen on the right side and left side respectively. Thus, the driving position was adjusted so that the car in front can be seen between the two images.) Nevertheless, regarding the front-back direction, the driving position is left up to the observer so that his eye position does not definitely match the standard eyepoint. Regarding the HUD position, it is not necessarily true that all observers see the HUD image in the display position of the angle of vision that was set.

Because of the above restrictions, the results obtained in this experiment are not used to obtain distribution of a feeling of troublesomeness. Instead, they are mainly used to make a comparison of the feeling of troublesomeness between a stop state and a driving state.

The observers carried out two types of tests two times each: an experiment in which evaluation was made with the car in a stopped state, and an evaluation made when the vehicle was traveling at 60 km/h. The observer observed the HUD image for approximately 10 seconds and then replied concerning results of evaluation of feeling of troublesomeness using the same standards as in Table 1. The display order of the HUD positions was random.

There was a total of 14 observers with normal color vision and with an average age of 27.1 years. That meant there was a total of 28 evaluation data obtained for a single display position. The weather during the test alternated between fair and cloudy.

EXPERIMENTAL RESULTS AND CONSIDERATIONS - As was the case with the basic evaluation experiment, regarding a feeling of troublesomeness, there was an evaluation of the number of replies for "rather troublesome," "very troublesome" and "extremely troublesome."

Figure 11 shows the evaluation results when the vehicle was stopped. Regarding the vertical direction, this is more or less symmetrical with the horizontal line of y=0. As for the horizontal direction, the feeling of troublesomeness was distributed to the left in relation to the vertical line of x=0. Figure 12 shows the results of evaluation during driving. These results are similar to the evaluation results when the vehicle was stopped. However, an examination of details shows a large number of replies on the left side and the lower side.

If this distribution is examined in terms of average and standard deviation, the values when the vehicle is stopped are $\mu_x = -1.9$ and $\sigma_x = 6.4$ in a horizontal direction, and $\mu_y = 0.6$ and $\sigma_y = 2.9$ in a vertical direction. In this case, the values in a horizontal direction are derived only from the evaluation results on the horizontal line y=0; the values in a vertical direction are derived only from the evaluation results on the vertical line x=0. During driving, the values are $\mu_x = -2.6$ and $\sigma_x = 6.7$ in a horizontal

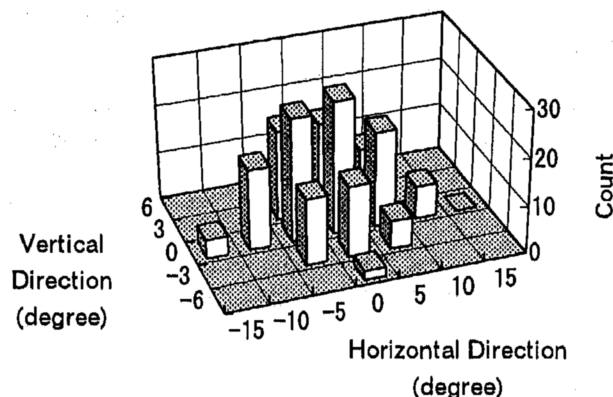


Fig.11 Evaluation results of a feeling of troublesomeness (when the vehicle is stopped)

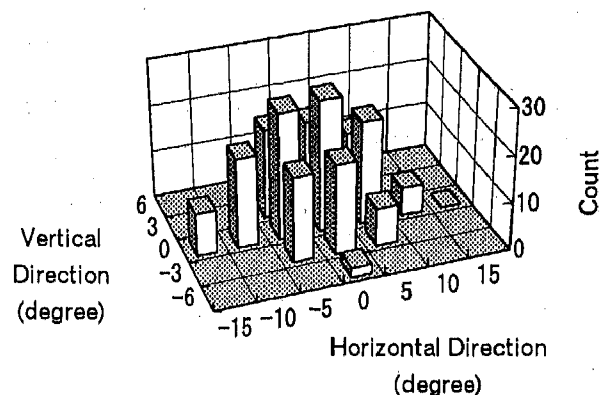


Fig.12 Evaluation results of a feeling of troublesomeness (during driving)

direction, and $\mu_y = 0.4$ and $\sigma_y = 3.0$ in a vertical direction.

Thus, the distribution of a feeling of troublesomeness shifted to the left during actual driving as compared to when the vehicle was stopped. On the other hand, regarding a vertical direction, the distribution during driving shifted slightly downward as compared to when the vehicle was stopped, although this was not a marked difference. As for the spread of distribution, no marked difference was ascertained between when the vehicle was parked and when it was driven.

In past research which studied the distribution of the line of vision of the driver during driving, it was clear that the line of vision of the driver was distributed toward the side of the street (sidewalk side)[10,11]. In other words, in cases like that in Japan where the cars (with steering on the right) are driven on the left side of the road, the line of vision is distributed toward the left side[10]. On the other hand, in cases such as in the USA where cars (with steering on the left) are driven on the right side of the road, the line of vision is distributed toward the right side[11]. Thus, when driving the car as in this case instead of sitting in a parked car, the reason why the distribution of a feeling of troublesomeness shifts to the left side is because the line of vision (attention) of the driver tends to shift to the left by force of habit so that HUDs located in that direction were perceived as troublesome.

On the other hand, regarding a distribution of the line of vision in a vertical direction, Okabayashi et al. have produced the following research results[5]. According to those results, the center of distribution of the line of vision in a vertical direction is roughly at the origin point on highways. On roads that are highly congested, this shifts to a position that is about 1 to 1.5 degrees below the origin point. In the present experiment, the traffic conditions allowed for leeway since the observation vehicle was located 50 meters behind the other vehicle on the test course. That meant the level of congestion was low and that conditions were close to those on a highway. As a result, even when driving, the amount of shifting of the line of vision to the bottom was small compared to when the car was stopped. When driving on roads with a high level of congestion, the distribution of the line of vision shifts downward. As a result, the distribution of a feeling of troublesomeness also is measured as shifting downward.

CONCLUSION

In order to clarify a display position of HUDs in which there is no feeling of troublesomeness for the driver, the authors carried out two types of tests. The first was a basic evaluation test in which plate glass was placed on a desk outside and an HUD was displayed on the plate glass. An observer seated in a chair were then asked to evaluate the feeling of troublesomeness regarding the HUDs. In the second test, an observer actually seated in a vehicle was asked to evaluate the feeling of troublesomeness of the HUDs. In this experiment a mockup HUD made of clear film was used to carry out the test in order to investigate whether there are differences in the feeling of troublesomeness when the vehicle is parked and during driving.

The results of the first basic test are as follows:

(1) The feeling of troublesomeness is highest in the display position located directly to the front of the observer. This decreases as the position moves away from that position either to the left and right or up and down. This distribution of a feeling of troublesomeness can be approximated with 2-dimensional normal distribution.

(2) When the feeling of troublesomeness is expressed in terms of 2-dimensional joint normal distribution, the correlation coefficient of the horizontal direction and vertical direction is almost zero. As a result, the

evaluation results for the horizontal direction and vertical direction are independent. In other words, by obtaining the values in a horizontal direction and vertical direction separately, it is possible to estimate the 2-dimensional normal distribution.

(3) When HUDs are displayed so that they can be seen below by a standard driver, the HUD image can be seen close to the position directly in front for drivers with a low eyepoint. Thus, we can estimate that the feeling of troublesomeness will increase for drivers with a low eyepoint.

The results for the test in which observers sat in actual vehicles are as follows.

(4) Compared to when the vehicle is stopped, the distribution of a feeling of troublesomeness during driving shifts toward the left. The reason is that, generally speaking, the distribution of the line of vision of the driver tends to shift toward the left. The reason why the line of vision shifts toward the left is the case of Japan where the cars (with steering on the right) are driven on the left side of the road. In cases such as the U.S. where cars (with steering on the left) are driven on the right side of the road, the distribution of the line of vision will shift conversely to the right. Thus, we can assume as a result, that the distribution of the feeling of troublesomeness will shift toward the right in the United States.

(5) As for the vertical direction, when it came to the present test where there was sufficient leeway in driving conditions on the test course, there was almost no difference recognized regarding the distribution of a feeling of troublesomeness between a driving and parking state. Past research on roads with a high level of congestion shows clearly that the distribution of the line of vision shifts downward. As a result, when driving on roads with a high level of congestion, we can estimate that the distribution of a feeling of troublesomeness will also shift downward.

In the present research, we revealed that in display positions close to the area immediately in front of the driver, the HUDs are felt as troublesome. However, regarding information of an urgent nature that must immediately be relayed to the driver, we can assume that the emphasis is more on speed of understanding and accuracy of understanding than troublesomeness. This may indicate that it is actually better to locate the HUDs closer to the area directly in front of the driver in the case of transmitting urgent information.

REFERENCES

[1] Shigeru Okabayashi et al.:"Development of Practical Heads-Up Display for Production Vehicle Application," SAE Paper 890559 (1989)

[2] M.Weihrauch et al.:" The First Head Up Display Introduced by General Motors," SAE Paper 890288 (1989)

[3] Jun'ichi Fukano et al."Automotive Head-Up Displays for Navigation Use," Proc. of the 14th International Technical Conference on Enhanced Safety of Vehicles, paper No.94-S2-O-02 (1994)

[4] Kazumoto Morita et al.:"Legibility of Displayed Place Names by In-Vehicle Navigation Devices," J.Light & Visual Environment Vol.20, No.1, Illum.Engng.Inst.Jpn. (1996)

[5] Shigeru Okabayashi et al.:"Driver's Ability to Recognize Forward View and Head-Up Display Images in Automotive Application Ⅱ ," J.Illum.Engng.Inst.Jpn. Vol.76, No.2, pp.35-44 (1992) (in Japanese)

[6] Y.K.Lin:"Probabilistic Theory of Structural Dynamics," McGraw-Hill,Inc. (1967)

[7] Takuro Kishine:"Statistics Theory from Beginning to Application," Yokendo, pp.207-216 (1976) (in Japanese)

[8] ISO 4513:"Road vehicles-Visibility-Method for establishment of eyellipses for driver's eye location" (1978)

[9] Shigeru Okabayashi et al.:"Driver's Ability to Recognize Forward View and Head-Up Display Images in Automotive Application Ⅲ ," J.Illum.Engng.Inst.Jpn. Vol.77, No.6, pp.285-295 (1993) (in Japanese)

[10] Toshiaki Miura:"Behavior and Visual Attension," Kazamashobo, pp.56-57 (1996) (in Japanese)

[11] David Shinar:"Psychology on the Road - The Human Factor in Traffic Safety," John Wiley & Sons,Inc. (1978)

About the Main Author

Kazumoto Morita was born in November, 1951 in Japan. He received a M.Eng. from Kyoto University (1976). His research interests are concerned with the safety of automobiles, especially with the automobile lighting, visibility, navigation systems, head-up displays, and elderly drivers performance. Fax No. +81 422 41 3233.

970230

New Optical Distortion Measuring Method Using Digital Image Analysis of Projection-Moire Patterns

A. S. Redner and G. K. Bhat
Strainoptic Technologies Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Optical distortion of images observed through windshields is due to the geometry to a small degree, but to a larger extent, due to effects occurring in production. The detection of manufacturing defects is of primary concern for on-line and off-line inspection procedures. Zebra patterns commonly used to detect serious windshield defects only yield qualitative and highly subjective results. This new method uses fine Moiré screens projected through the inspected windshield on a master grating, directly yielding a full-field fringe pattern directly revealing the line-of-sight deviation. PC-based Digital Image Analysis of the Moiré Fringes permits nearly instantaneous quantitative evaluation yielding a two-dimensional map of distortion levels and deviations. Complete derivation of Fringe-order relations to the measured distortion parameters is included in this paper, as well as a description of the software used in Digital Image Analysis.

Introduction

Quality control (QC) in manufacturing of automotive glazing includes an valuation of the optical quality, or "distortion". Certain amount of distortion is always present, since even most perfect curved glass placed at an angle to the line-of-sight will introduce some distortion (1). This geometry related distortion is taken into consideration in product design and is usually negligible.

The quality control is more concerned with detection of manufacturing defects and problems that are traceable to the float-glass tempering, bending, or in case of laminated panes, to the laminating process. Presently, the QC operations are primarily relying on visual inspection. Unfortunately, the approaches used for visual evaluation are subjective and not quantifiable. The new method developed by the author is designed to fill the existing gap, using PC-based vision technology in place of the human eye.

Optical Inspection Parameters:

Image-Displacement, Deviation and Distortion

Figure 1 below identifies the types of optical problems encountered in a transparency. The image DISPLACEMENT occurs when a ray of light crosses glass of uniform thickness placed at an angle. The displacement is very small and in fact, does not affect the vision or the quality of the observed image. The DEVIATION of line-of-sight α moves the image, similarly to the action of an optical wedge. A uniform deviation (or a perfect wedge) seldom occurs in practice, and does not distort the image quality.

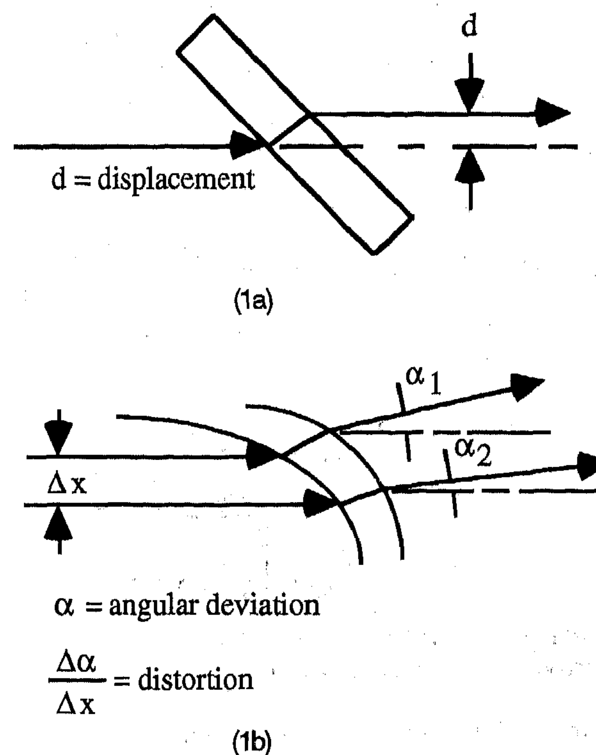


Figure 1: Optical Distortion Parameters

What is mostly of concern, is the DISTORTION defined as gradient of the deviation angle α, or $\Delta\alpha/_{\Delta x}$ (2) (3). The DISTORTION is actually the element that not only deviates the angular position of the observed object, but also deforms its shape and is clearly the most disturbing type of defect.

Presently used Methods

The most widely used method in QC is a "ZEBRA" board consisting of a brightly illuminated white screen containing black lines, (typically 1 inch wide) traced at 45°. The operator observes the "ZEBRA" through the inspected pane and decides the "QUALITY".

Although extremely simple in concept, the method is highly effective in locating gross distortions, and the only instrument used (the human eye) combines the lowest reproducibility with highest cost, that in a 3-shift production environment approaches 0.25M/year, without considering the cost of training and lapses due to operators fatigue or inattention.

The GRID METHOD (3) is more suitable for use in a developmental or design qualification test, since the analysis of the image is extremely time consuming. This method primarily designed for aircraft transparencies uses a large-scale grid-board containing 12.7 to 25.4mm square array of lines, and uses a double exposure. Measuring of deviation and distortion is performed manually, comparing photographs of the grid observed with and without the inspected pane. While the method is effective to reveal the angular deviation, it is very ineffective in attempting to quantify the distortion. It is interesting to note that the ASTM precision statement clearly indicates that " it is not practical or justifiable to specify the precision and bias...because of the operator-sensitive nature in the measurements of the test photos".

The RASTER METHOD described in EDE/324 (2) and ISO 3538C was developed primarily for windshields. It requires an examination of an array of circular dots projected through the examined windshield using a slide projector. (Figure 2)

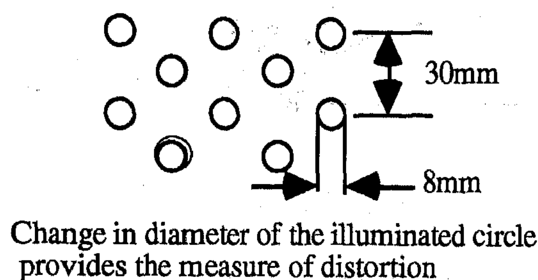


Figure 2: Raster Method of Measuring Distortion

The method is designed for an engineering laboratory, and reveals distortion as a change in image magnification of approximately 8mm diameter circles spaced approximately 30mm vertically and horizontally (2). The information is highly discontinuous and the process of data reduction is strictly dot-by-dot. Such process is inherently slow, even using machine vision. The large spacing between data points makes it inaccurate, since a 22mm defect can be totally missed by this system, making this concept unsuitable for Quality Control.

Visual application of this method (using a template) is yielding a margin of error up to 20% of the limit value of $\Delta\alpha/_{\Delta x}$ assuming that the operator can evaluate a change in the projected image within a fraction of mm.

Considering that the required distortion should not exceed 2' of arc (2), the increase of the 8mm circle at 2m distance becomes 1.16mm implying a distortion of 14.5% of the size, that is an extremely large change of magnification.

Moiré Method for Distortiometry

The optical schematic of the newly developed MOIRÉ DISTORTIOMETRY method is shown schematically in Figure 3.

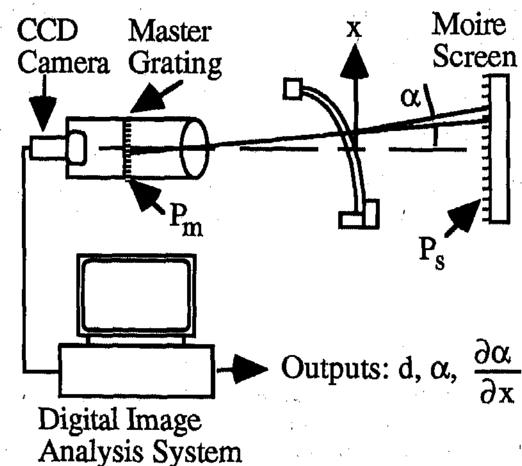


Figure 3: Optical Schematic of Moiré Distortiometry

Here, a MOIRÉ screen is projected on a high-density reference master through the inspected pane, forming MOIRÉ fringes. A Digital Image Analysis system (DIAS) performs the analysis the MOIRÉ pattern, retrieving a full-field information on image displacement, d, angular deviation

$$\alpha = \alpha(x, y)$$

and distortion $D = \partial\alpha/_{\partial x}$ and $\partial\alpha/_{\partial y}$.

The PROJECTION MOIRÉ FRINGES can be formed using an illuminated screen projected on the master, or in some applications, the master is projected on the screen, yielding identical results. In both instances, Moiré fringes reveal the relative displacement of lines, between the screen and the master due to deviations of rays crossing the inspected transparency.

In the derivation below, it is assumed that the magnification of the system is equal to the ratio of the screen and master pitches, $m = P_s/_{P_m}$. In absence of a tested pane, a constant fringe order is then observed throughout. A relative motion of projected lines due to the deviation generates Moiré fringes. The fringe order N is related to the deviation and pitch by:

$$N = L\alpha/_P \qquad (1)$$

The resulting light intensity modulation $I/_{I_0}$ is related to the fringe order N by:

$$I/_{I_0} = \sin^2 \pi N \qquad (2)$$

The displacement d, and the deviation of line-of-sight α is then obtained on a full-field view from the fringe pattern by:

$$d = Np$$

and the deviation becomes :

$$\alpha(x,y) = Np/_{L_1} = \frac{p}{\pi} \; arc \; \sin\sqrt{I/_{I_0}} \qquad (3)$$

The distortion in horizontal and vertical plane is quantified also on a full-field view.

$$\text{Distortion } (x,y) = \frac{\partial\alpha}{\partial x} \; and \; \frac{\partial\alpha}{\partial y} \qquad (4)$$

Since the detection of fringe order N is accomplished using the Digital Image Analysis within 1% of the fringe order N, the minimum detectable deviation angle $\alpha_{\min}$ using a 50 $l/_{in.}$ (2 lines/mm) screen placed at 2 meter from the examined pane, becomes:

$$\alpha_{\min} = \frac{Pmm}{100x2000_{mm}} = \frac{1}{2x100x2000} = 2.5x10^{-6}$$

$$radius\;(0.0086\ \text{min.}\ arc) \qquad (5)$$

The sensitivity of the Moiré method exceeds the sensitivity of the Raster projection method approximately 100 fold.

Software Description

The Moiré pattern revealing the distortion is quantified and detected by introducing "Carrier fringes." Carrier fringes are generated using a combination of optical magnification and master/screen pitch ratio. Figure 4 shows evenly spaced carrier fringes recorded using a CCD camera, in the absence of the pane,

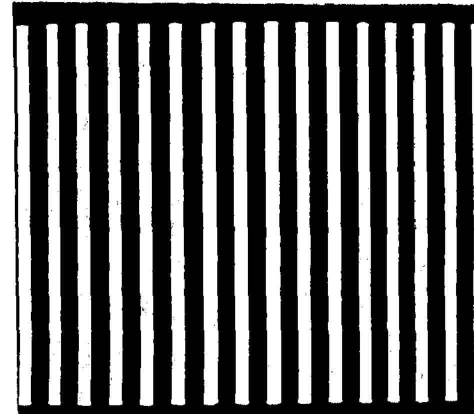


Figure 4: Carrier Fringes Recorded in Absence of the Inspected Transparencies

and Figure 5 shows the carrier fringes (distorted) with the pane in position.

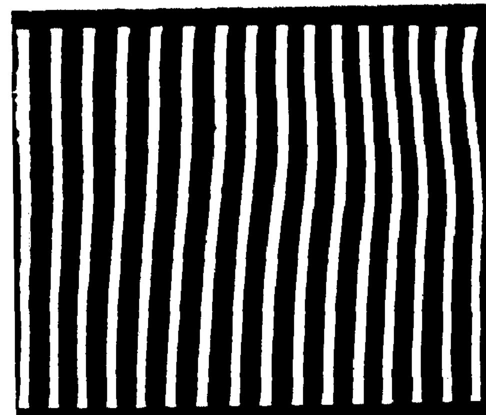


Figure 5: Moiré Fringes Recorded with the Transparency in the Field of View

Digital filtering and phase unwrapping techniques are employed to extract the carrier fringe orders. The deviations are obtained by computing the difference between the two carrier fringe orders. The distortion is then computed by differentiating the deviation field.

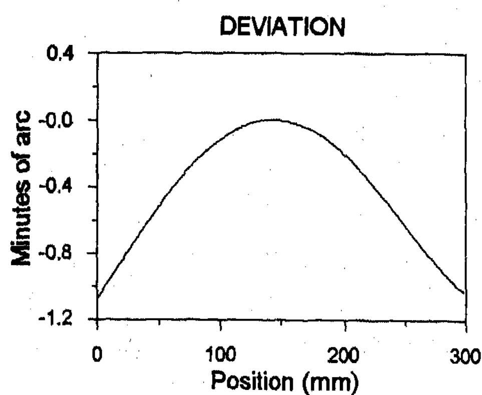
The software is menu driven and includes Graphical User Interfaces (GUI) for user interaction. The software includes two major options: (1) Quality control and (2) Engineering analysis.

Quality control option is used for on-line inspection of panes. When the operator selects this option, after placing the pane in position, the image of the Moiré fringe pattern is acquired and the computation of the deviation function d(x,y) is performed at all points. This function is used to generate a full-field pseudo-color map, illustrating graphically the distortion. A PASS-FAIL decision is automatically made based on a preselected threshold level. Figure 6, for example, shows full field map of deviation values

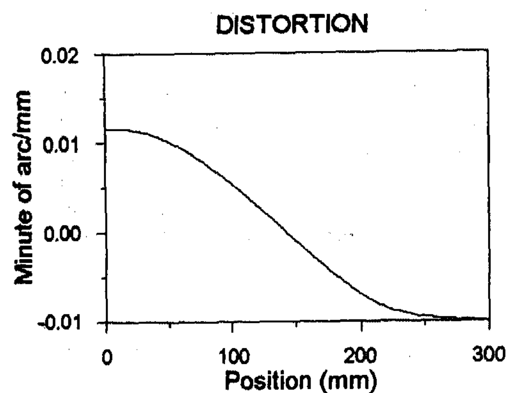


Figure 6: Full-Field Map of Deviation

and Figures 7a and 7b show respectively the deviation and distortion values along a cross section of the analyzed region.



(7a)



(7b)

Figure 7: Deviation and Distortion Along a Selected Cross-Section

Engineering analysis option is used for research and development purposes. The images are captured and analyzed as in quality control. However, no automatic decision is made. The operator has the choice of performing several types of analysis, including:

1) saving and retrieving images.

2) image enhancement by noise reduction and contrast stretching.

3) point analysis to obtain deviation and distortion values at any desired point in the image.

4) line analysis to obtain deviation and distortion values along any desired line in the image.

5) area analysis to obtain deviation and distortion values within any desired region in the image.

6) color coding to visualize different levels of deviation and distortion

7) saving numerical values of deviation and distortion for post processing and display.

<u>Conclusions</u>

A newly developed MOIRÉ DISTORTIOMETRY method was developed and its merits demonstrated. The method permits direct full-field data-acquisition revealing optical distortions due to light passing a windshield or aerospace transparency. The speed of response of the system is compatible with ON-LINE inspection requirements and the sensitivity is a 100 fold improvement over visual concepts.

References

1. Irland M.J., "Windshield Optics", SAE Paper #700480, May 1970.

2. Reg. No. 43, Annex 3, Uniform Provision Concerning the Approval of Safety Glazing and Glazing Material.

3. ASTM F733-90 "Standard Practice for Optical Distortion and Deviation of Transparent Parts Using the Double-Exposure Method", ASTM Book of Standards 15-03.

970237

A Look at the Automotive-Turbine Regenerator System and Proposals to Improve Performance and Reduce Cost

David Gordon Wilson and Andreas Carl Pfahnl
Massachusetts Institute of Technology

ABSTRACT

The adoption of turbine engines for automotive power plants has been hampered by the high cost, high leakage and high wear rate of present designs of ceramic-matrix regenerators. Proposals are made and analyzed here for design directions to achieve substantial improvements in all three areas. These include lower-cost extruded and pressed matrices; and clamping seals coupled with incremental movement of the rotary-regenerator matrix.

INTRODUCTION

The development of the ceramic automotive regenerator is briefly described. Next, a review of its present status discusses the current diverse-natured problems still inhibiting the commercial use of ceramic regenerators. Finally, a series of changes in the method of operation, manufacture, and design that could greatly improve the performance and reduce the cost are introduced and reviewed.

BACKGROUND. The designers of early automotive gas turbines recognized that, whereas aircraft turbine engines had achieved dominance using cycles incorporating high-pressure-ratio compressors and high-temperature air-cooled turbines but without heat exchangers to capture the heat from the turbine exhaust, this approach was not viable for automotive turbines (McDonald,1980[1]). Small compressors cannot produce pressure ratios of over 10:1 without suffering penalizing losses in efficiency because of the very small high-pressure passages required and the consequent relatively large blade-tip clearances. Two other adverse factors are that high-pressure-ratio compressors have very poor part-load efficiencies, and small turbine blades cannot be air-cooled. Accordingly, the principal alternative cycle was rightly considered the better choice: a low compressor pressure ratio (of the order of 4:1) and a high-effectiveness heat exchanger.

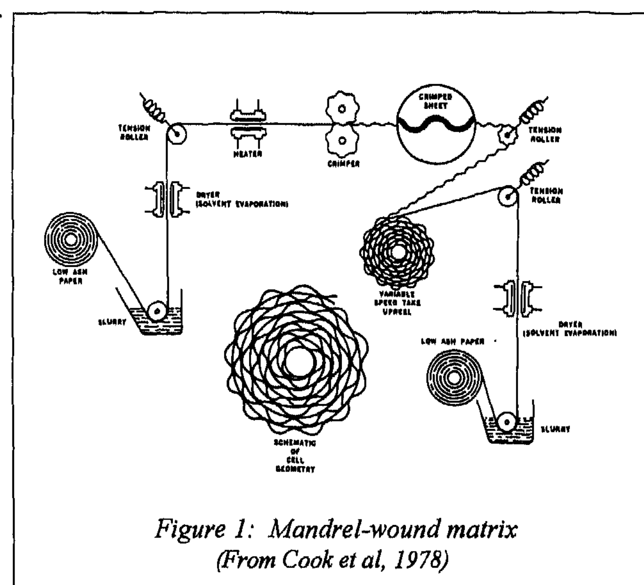


Figure 1: Mandrel-wound matrix (From Cook et al, 1978)

(The effectiveness - the proportion of potential exhaust heat actually recovered - started in the early 1950s at about 75% and has been continually increased) (Wilson, 1995 [2]). Early rotary regenerators were made of stainless-steel. Alternate bands, wavy and plain, were wound around a mandrel and brazed (figure 1). Metal regenerators of this type were used in the celebrated 1963 Chrysler demonstration when fifty turbine-powered automobiles were loaned to over two-hundred members of the public for up to three months each. However, the resulting large quantity of nickel-chrome alloys considerably increased the cost of the engine. The head of the Chrysler program, George Huebner, stated subsequently that the cost of the materials alone in its turbine engine was over eight dollars per horsepower versus $1.50 per hp (1963 dollars) for the total production cost of the then-current spark-ignition engine (op.cit.). Huebner had approached Corning Glass with regard to the possible use of ceramics for regenerator matrices, but it was a British rival, Rover, that put Corning Cercor regenerators to first use in its 1965 entry of a regenerative turbine car in the prestigious Le

Mans 24-hour race. The remarkable performance of the regenerators, as well as of the turbine engine itself, seemed to assure their future application in automobiles.

By this time, Ford had initiated its own effort to develop ceramic regenerators beginning with the lithium-aluminum-silicate (LAS) type in 1965. The most serious problem surfaced after years of extensive highway testing with its Ford "707" truck engine. It was found that sulfur from diesel fuel and sodium from road salt damaged and caused failures of the regenerators (Cook et al, 1978 [3], Day, 1979[4]). Notable challenging mechanical-design problems persisted in the sliding seals and drive system.

Although by 1974 Corning had developed two alternative ceramics, aluminum silicate (AS) and magnesium aluminum silicate (MAS), and although Ford had greatly improved on the mechanical systems and was close to full-production commitment, the project was terminated. The engine was eventually sold to Teledyne.

The ceramic regenerator was not sidelined, however. As turbine-rotor-inlet temperatures in small engines have increased to about 1650 K, boosted partly by the government-sponsored advanced-gas-turbine (AGT) and other programs, some form of ceramic material became mandatory for motor-vehicle regenerators. The coupling of high inlet temperatures to the relatively low cycle pressure ratio gave a turbine-expander exit temperature too high for metal heat exchangers. Development on both ceramic recuperators (tubular and plate-fin) and rotary regenerators continued, which for the latter heat-exchanger type meant improving primarily on the manufacturing methods, ceramic materials, and seals.

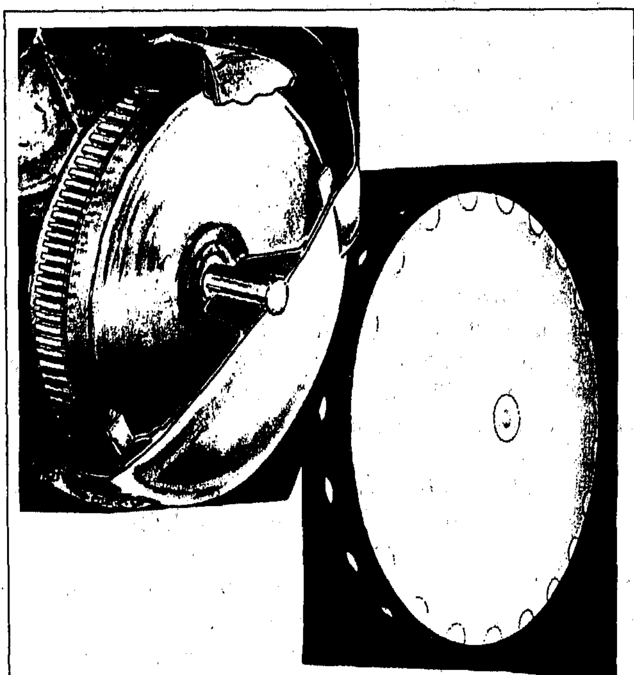


Figure 2 Corning "Cercor" disk regenerators

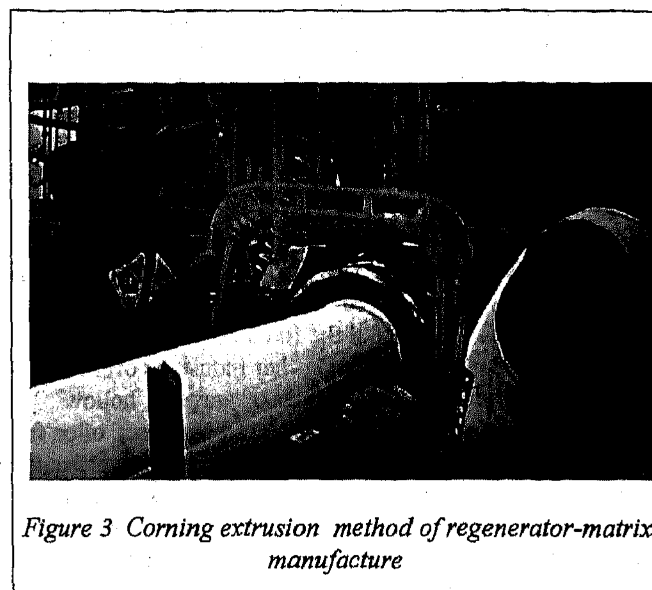


Figure 3 Corning extrusion method of regenerator-matrix manufacture

MANUFACTURING METHODS. Corning manufactured most of the regenerator disks in a manner similar to that described previously for the stainless-steel disks (figure 1), except that in place of metal foil, teabag-type paper soaked in ceramic slurry was used (Day, 1994[5]). The disks then underwent a firing schedule before the surfaces were machined. The final versions typically had a porosity (proportion of open area) of 0.7 and a hydraulic diameter of the passages of about 0.5 mm, a degree of fineness of manufacture that still inspires awe (figure 2).

The wrapping method of regenerator-matrix manufacture was too expensive for large-scale production. It also necessarily involved non-optimum passage shapes, and a flow resistance that could vary spatially (London and Shah, 1973 [6],

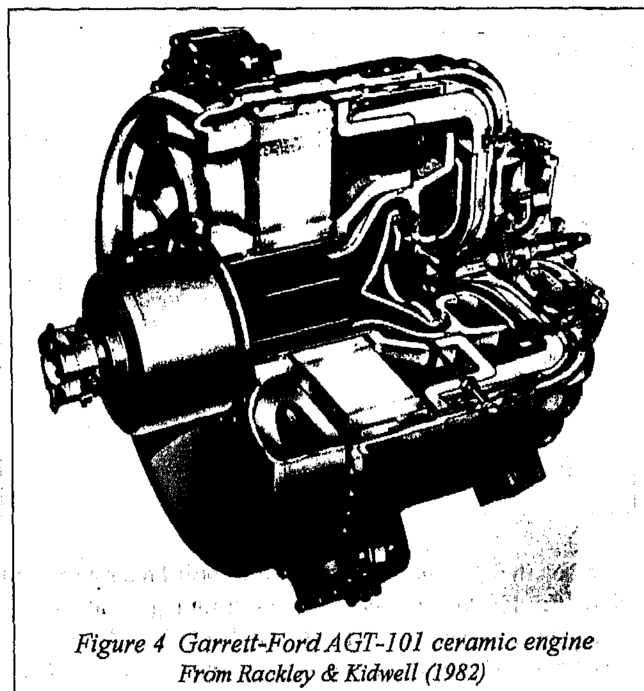


Figure 4 Garrett-Ford AGT-101 ceramic engine
From Rackley & Kidwell (1982)

and 1980 [7]). Corning developed an extrusion method of core manufacture that is potentially far lower in manufacturing cost, and allows more-advantageous square and rectangular passages of uniform size to be produced (Day, 1994, op.cit.) (figure 3). It was used, for instance, in the advanced Garrett-Ford AGT-101 engine (Rackley and Kidwell, 1982[8] (figure 4).

SEAL PERFORMANCE AND WEAR The rotary regenerator was invented by Fredrik Ljungstrom in 1920 (Hansson (1955)[9]) as a variation of the fixed-bed regenerators developed for the metallurgical and other industries over millennia. In these, a container of stones or gravel or hollow ceramic bricks is heated for a period by hot gases, often from a furnace exhaust. Then the combustion or blast air would be switched to pass through the hot solids to pick up heat. The three or four fixed-bed regenerators associated with blast furnaces are called "hot stoves" or "Cowper stoves", and usually two or three are being heated while one is transferring heat to the blast air.

Ljungstrom put the solids in a rotor that passed the inlet and outlet ports continuously. Thus the sliding seals acted in the place of the flow-switching valves. However, they could not seal against a large pressure, and Ljungstrom's rotary regenerators were applied principally as air preheaters for steam-generator combustion air, recovering heat from the near-atmospheric-pressure exhaust stream.

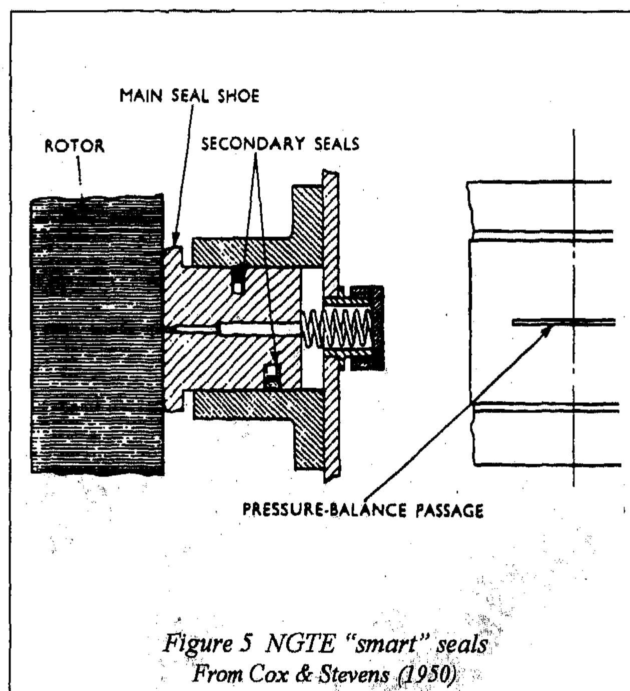


Figure 5 NGTE "smart" seals
From Cox & Stevens (1950)

When the National Gas-Turbine Establishment in Britain took on the development of the Ljungstrom regenerator for turbine engines, a seal was developed that followed the surface of the stainless-steel matrix as it passed under it (Cox and Stevens, (1950)[10]) (figure 5). Most other regenerator seal designs work towards achieving the same goal, of providing a means of pressing a low-wear rubbing seal on the regenerator core while allowing some play for it to follow the surface. The disk tends to warp while operating in the high temperature gradient. Seals supported by bellows [Norbye, 1975[11]] and by leaf springs were developed to allow for the flexibly following the matrix surface (Davis and Jackman, 1996[12]). All these designs have one commonality: they are rubbing face seals. Therefore they wear.

Wear has produced the problems that plagued these designs and those of today: leakage and a relatively short working lifetime. Leakage of the compressed air flow imposes a significant penalty on the cycle thermal efficiency (Pfahnl and Wilson, 1996[13]). Furthermore, the useful life is set by the increasing leakage producing an eventually unacceptable performance penalty (Shimada et al, 1993[14].

PROBLEMS OF PRESENT REGENERATORS

The substantial potential reduction in mass-production costs brought about by extrusion did not immediately make automotive turbine power viable. New seal designs and wear-face materials have so far not turned out to be low-cost, high-performance solutions. Many problems still persist in the ceramic regenerator that prevent low-cost implementation in a small automotive-size turbine. However, the following lists specific problem areas (including those mentioned above) where new ideas, concepts and design methods could significantly contribute to making ceramic regenerators a viable commercial heat exchanger for motor-vehicle use.

MANUFACTURE The extrusion method of matrix manufacture can produce cores to a present maximum diameter of about 300 mm, a limit brought about because of stress limits in the fragile extrusion die nozzle. This size limits the application to low-power engines (below 100 kW) or compromises the design of larger engines.

OPERATION Three problems are the following.

1. The leakage rates under the high-pressure sliding seals of regenerators, at a reported minimum of three percent of the compressor-delivery air, impose a high penalty on the overall engine thermal efficiency (approximately a one-point loss in efficiency for each one-point increase in leakage).

2. The seals and disk faces often wear rapidly, increasing the actual leakages to over 15% (Shimada et al, op cit).

3. The use of monolithic cores that can be only replaced, not repaired, increases the cost of ownership.

DESIGN Three design problems frequently found in current regenerators are these.

1. In most existing regenerator designs, the exhaust-side pressure drop is extremely high, often over ten percent, adding a large penalty on the engine thermal efficiency

2. On the compressed-air side of the matrix the pressure drop is often a small fraction of one percent, a value too low to produce uniform flow through the matrix. The effectiveness of the regenerator suffers seriously (Wilson, 1966[15]).

3. Designs having a low regenerator effectiveness penalize the engine thermal efficiency unnecessarily.

PROPOSED CHANGES

The following sections examine the potential of segmented construction of a rotary regenerators, coupled with discontinuous rotation, and other-than 50-50 hot-cold face-area split to address the previous listed regenerator-development problems.

DISCONTINUOUS ROTATION At MIT a regenerator design intermediate between the continuous-rotation Ljungstrom type and the switched fixed-bed stoves has been under study. An approximation to continuous rotation is a stepper-motor type of movement. The regenerator rotor could move rapidly through an incremental angle, pause for a period, and then move again. If the pause period were considerably longer than the period for rotation, the seals could be brought up hard against the matrix faces and held there to inhibit virtually all seal leakage. Withdrawing the seals for rotation should occur over such a small distance that simply relaxing the applied seal pressure may suffice. Seal wear should be drastically reduced and may become negligible, since wear is a strong function of the applied force. The effective seal-clearance height during rotation would, therefore, be less than that found under normal seals after a period of wear. Various projects have been undertaken to assess the feasibility and design of such a discontinuously-rotating regenerator.

Beck examined the discontinuous-rotation case to determine the penalty that would be incurred over the continuous-rotation case. He found that a penalty could be entirely avoided if the incremental-movement angle and the angle between the two radial high-pressure seals were chosen appropriately (Beck, 1993[16]). For instance, if the incremental movement is 30 degrees and the angle between the radial seals is 120 degrees there would be no effectiveness penalty.

In design studies of methods of achieving accurate incremental movement we have examined forms of Geneva mechanisms, geared stepping motors, and pneumatic actuators. There appears to be a high degree of design freedom. Although aesthetically one would like to produce a controlled acceleration-deceleration profile that would minimize matrix stress and perhaps conserve the sum of kinetic and potential energy, Long showed that an incremental-movement period that occupied only ten percent of the time between rotations would not produce damage-inducing stresses in a typical ceramic matrix (Long, 1992[17]).

Similarly there appears to be considerable design freedom in the choice of mechanisms for withdrawing the high-pressure seals through a chosen distance and for reapplying them (or for simply reducing the seal pressure).

An examination of many of the design aspects of discontinuous-rotation regenerators was made by Pfahnl and Wilson, 1996[18]. No areas of great difficulty were found.

The cost of introducing a discontinuous-movement mechanism into a standard rotating regenerator, at least at the design stage, seems small compared with the obvious benefits. A retrofit would be clearly more costly.

SEGMENTED MATRIX ELEMENTS. If the matrix moves in increments of 30 or 45 degrees, there is no longer a great need to use matrix elements that prevent through-core leakage under the seals, as at present. The matrix could be porous not only axially but peripherally, so long as there were barriers to through-core leakage where the seals clamp to the matrix faces.

This possibility opens up a great deal of design freedom. A matrix could be made of a collection of individual segments, each contained within a thin solid wall (figure 6). The matrix in this surrounding casing could be sintered of such low-cost elements as spheres or twisted square-section rods or tubes of various cross-sections.

Doing so could greatly reduce the manufacturing cost of the matrix. The segments could be mass-produced. Being of considerably smaller size than the total matrix, extrusion could be used on the segments which could then be assembled

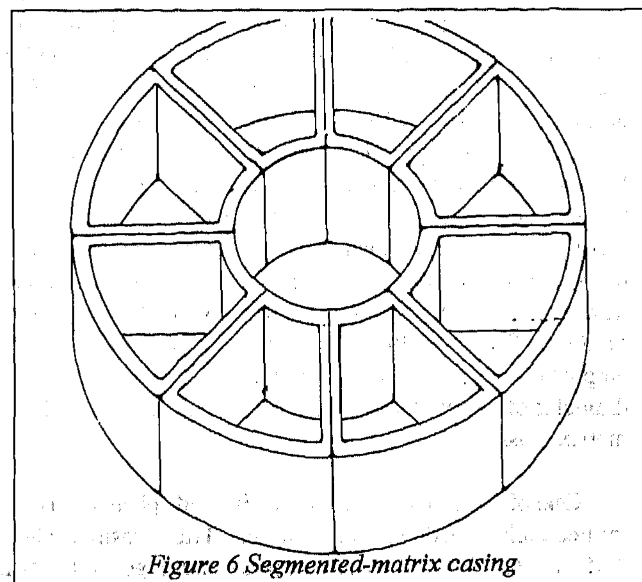


Figure 6 Segmented-matrix casing

into a disk that would be much too large to be extruded. Or just the cylindrical casing "shell" could be formed and filled with the matrix elements mentioned above. A cost of only a few percent of that for extruded elements is anticipated.

A large matrix built up in this manner and held together perhaps by one or two circumferential bands would have two additional advantages over a monolithic disk. The thermal stresses would be much lower. Thermal-expansion coefficients are very low for most ceramic materials, but the ductility is also very low, so that thermal stresses and fatigue become very important design criteria. And if a disk segment became clogged or otherwise damaged, it alone could be withdrawn, perhaps through a casing access panel, and replaced.

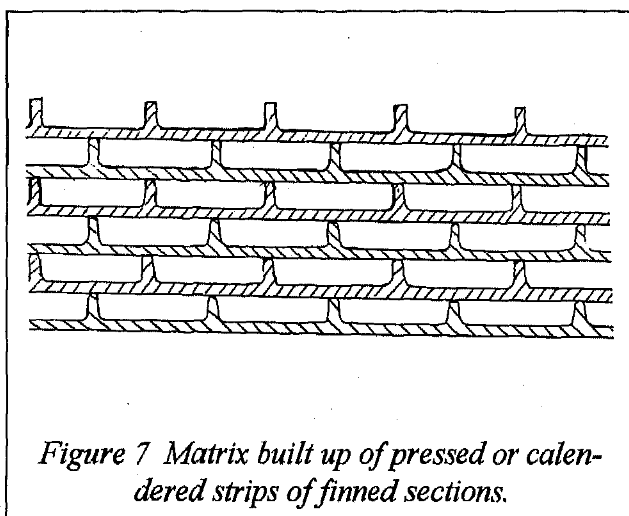


Figure 7 Matrix built up of pressed or calendered strips of finned sections.

RESULTS OF COMPARATIVE DESIGN STUDY.

We examined the performance of alternative regenerators for a 100-kW automotive engine using the basic method of Kays and London[19] with some additions given by Beck and Wilson (1996)[20]. We have incorporated refinements that allow us to specify the hot-side (low-pressure) pressure drop (we normally start at 3% of the local total pressure), and to choose at least the range of the cold-side pressure drop (we aim to get as close to 1% as possible). A further development is to allow the designer to specify the temperature gradient through the matrix. Too high a temperature gradient increases the axial conduction and the thermal stresses. The method chooses a surrogate, the temperature gradient of the hot-side flow. At present we have chosen to limit the thickness (or perhaps the "thinness") to the value giving a hot-flow gradient of no more than 7.5 degrees K/mm. The principal outputs of these calculations are the required hydraulic diameter of the passages, and the face area, rotation rate, and matrix mass.

One of the low-cost matrices, sintered spheres, was immediately revealed as unattractive. The porosity is about 0.35, so that the face area and mass of the regenerator disk(s) would double compared with a matrix of porosity 0.7. Packed but aligned circular tubes would produce a matrix composed of two types of passages, with one hydraulic diameter inside the tubes and another for the "outside" flow between the tubes. These nonuniformities would produce flow and heat-transfer nonuniformities. Open-cell "reticulated" ceramic foam, mentioned by Day (op. cit.) could be an advantageous low-cost matrix, but no heat-transfer or flow-friction data have been produced to the authors' knowledge.

One apparently low-cost method is suggested in figure 7. Finned ceramic sheets that would give rectangular passages of length/width ratio close to four (which appears to be near the optimum shape) could be produced quickly by pressing or calendering and fed under automatic control into the segment casings before sintering. Cook et al (op. cit.) investigated calendering among other alternatives for coiled matrices. A matrix formed in this way would apparently have no thermal or fluid-friction compromises compared with extruded matrices, which at present are the standard against which other approaches should be judged.

CONCLUSIONS AND RECOMMENDATIONS

We have used qualitative arguments (some quantitative analyses are in a companion paper) to conclude that rotating a regenerative-heat-exchanger matrix in increments of, say, 30 or 45 degrees under relaxed sealing pressure and clamping the seals during a succeeding rest period would simultaneously greatly reduce the seal leakage and the seal and rotor wear.

If this method of operation (for which MIT has a patent [21]) is adopted, a segmental form of construction of the regenerator rotor is proposed. Doing so would presumably reduce manufacturing costs, give more design freedom (especially with regard to regenerator- or engine-size limit), reduce thermal stresses, and give the potential of replacing only one segment of a regenerator matrix in the event of local damage.

This approach would also increase the face area and weight of the matrix to a small extent, and would require measures to prevent seal leakage along the necessarily chamfered edges of the segments.

ACKNOWLEDGMENTS

Early work on the concepts discussed in this paper was funded by Lincoln Laboratory. Several generations of undergraduate and graduate students have worked on associated hardware or analytical projects. The authors express their appreciation for all these contributions.

REFERENCES

[1] McDonald, Colin F. (1980). "The role of the ceramic

heat exchanger in energy and resource conservation ." Jl. of Engg. for Power vol. 102, ASME, NY, NY.
[2] Wilson, David Gordon (1995). "Automotive gas turbines: the pioneers; progress and setbacks; and government funding and the way ahead." A three-part series in Global Gas Turbine News, vol. 35 nos. 2, 3 & 4, IGTI/ASME, Atlanta, GA.
[3] Cook, J. A., C. A. Fucinari, J. N. Lingscheit, and C. J. Rahnke (1978). "Evaluation of advanced regenerator systems". DOE/NASA/0008/78/4, NASA report CR-159422, Cleveland, OH.
[4] Day, J. P. (1979). "A study of chemical reactivity in ceramic heat exchangers." Jl. of Engg. & Power vol. 101, pp. 270-274, ASME, NY, NY.
[5] Day, J. Paul (1994). "A rotary heat exchanger for automotive and other ground-based gas turbines." Paper no. 94-GT-124, ASME, NY, NY.
[6] London, A. L., and R. K. Shah (1973). "Glass-ceramic hexagonal and circular passage surfaces - heat transfer and flow-friction design characteristics." Paper no. 730100, SAE, PA
[7] Shah, R. K., and A. L. London (1980). "Effects of non uniform passages on compact-heat-exchanger performance". Jl of Engg. for Power vol. 102 pp. 653-9, ASME, NY, NY.
[8] Rackley, R. A., and J. R. Kidwell (1982). "The AGT-101 advanced automotive gas turbine." Paper 82-GT-91, ASME, NY, NY.
[9] Hansson, A. A. (1955). "Birger and Fredrik Ljungstrom - inventors." A. B. International STAL Company, Stockholm, Sweden.
[10] Cox, M. and Stevens, R. K. P. (1950). "The regenerative heat exchanger for gas-turbine power plants." Proc. Instn. Mech. Engrs., vol. 163 no. 60, London, UK.
[11] Norbye, Jan P. (1975). "The gas-turbine engine." Chilton Book Company, Radnor, PA.
[12] Davis, Leonard C., and Anthony J. Jackman (1996). "Some significant highlights in Allison regenerator development." Paper no. 96-GT-503. ASME NY. NY.
[13] Pfahnl, Andreas Carl, and David Gordon Wilson (1995). "Key issues in the general design of motor-vehicle gas-turbine engines." Paper no. 95-GT-393, ASME, NY, NY.
[14] Shimada, Kazuaki, et al (1993). "Advanced ceramic technology developed for industrial 300-kW CGT (ceramic gas turbine) research & development project in Japan". Paper no. 93-GT-188, ASME, NY, NY.
[15] Wilson, D. G. (1966). "A method of design for heat-exchanger inlet headers." Paper no. 66-WA/HT-41, ASME, NY, NY.
[16] Beck, D. S. (1993). "Regenerators for gas-turbine applications." Ph.D. thesis, mechanical engineering, MIT, Cambridge, MA.
[17] Long, Alexander E. (1992). "Discontinuous rotary-regenerator dynamic analysis". UROP report, MIT, Cambridge MA.
[18] Pfahnl, Andreas Carl, and David Gordon Wilson (1996). "A low-leakage discontinuously rotating regenerator designed for a motor-vehicle gas turbine." Paper no. 96-GT-452, ASME, NY, NY.
[19] Kays, W. M., and A. L. London (1984). "Compact heat exchangers." Third edition, McGraw-Hill, NY, NY.
[20] Beck, Douglas Stephen, and David Gordon Wilson (1996). "Gas-turbine regenerators." Chapman & Hall, NY, NY.
[21] Wilson, D. G. (1993). "Heat exchanger containing a component capable of discontinuous movement". US patent no. 5,259,444, assigned to MIT, Cambridge, MA.

970242

The Flywheel Battery Containment Problem

M. A. Pichot
The University of Texas at Austin Center for Electromechanics

J. M. Kramer
Argonne National Lab.

R. C. Thompson, R. J. Hayes, and J. H. Beno
The University of Texas at Austin Center for Electromechanics

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Flywheel energy storage systems (FESSs), commonly known as flywheel batteries, are being developed for a number of power averaging applications in the transportation industry. Most FESS designs incorporate rotors made from composite materials for which operating experience is limited and little failure data is available. Given the high energy levels inherent in these devices, a major task for the FESS designer is to provide adequate containment structures to insure safe operation in the event of flywheel failure. This paper discusses ongoing work intended to provide a better understanding of composite flywheel failure mechanisms and development of a methodology for safe containment design.

INTRODUCTION

The University of Texas at Austin Center for Electromechanics (UT-CEM) is currently under contract to develop composite FESSs for power averaging on hybrid electric transit buses and high speed locomotives. In addition, UT-CEM is participating in a Defense Advanced Research Projects Agency (DARPA) sponsored Flywheel Safety Project that seeks to investigate composite flywheel failure mechanisms and to develop systems to mitigate effects of those failures.

This paper discusses FESS principles, applications, and operating parameters. Likely failure scenarios for composite flywheels are shown, the need for containment structures is discussed, and additional information is given regarding the DARPA Flywheel Safety Project. A model to predict containment loads due to composite flywheel failures was developed by Argonne National Laboratory under contract to UT-CEM; the assumptions behind this model are discussed, and some sample calculations are presented. Results from flywheel failure testing are shown, and plans for future tests are discussed. Finally, a discussion of possible approaches to containment design is presented.

FLYWHEEL BATTERY PRINCIPLES

As the world consumption of fossil fuels has increased, so have the concerns about the environmental effects associated with our dependence on this nonrenewable resource. One result of this concern is a growing interest in electric and hybrid electric vehicles. All-electric vehicles typically use a large bank of electrochemical batteries to provide both power and energy storage and are range limited, based on the on board energy storage. Hybrid electric vehicles address the range issue by utilizing a small internal combustion prime power unit to drive an electric generator to meet the average power requirements. To address peak power requirements, an auxiliary energy storage and power conversion device (typically a FESS or a bank of electrochemical batteries) are required. In this type of hybrid system, or when used in conjunction with a bank of batteries in an all-electric vehicle, a FESS is used for power averaging.

Power averaging is accomplished primarily by storing energy generated during regenerative braking then releasing it during acceleration of the vehicle. The generation of braking energy on a vehicle is accomplished through the use of electric traction motors which operate as generators during braking. During the braking period, a motor/generator in the FESS is used to convert the electrical braking energy from the traction motors into mechanical energy in the spinning flywheel. When the vehicle accelerates, the FESS motor/generator converts energy stored in the flywheel back into electrical energy to power the drive wheels, completing the storage and recovery cycle. Because regenerative braking reduces losses, the fuel efficiency of the vehicle can be increased significantly. The peak power required of the prime power unit is reduced as well, which reduces both initial system costs and operating costs.

While both power averaging and regenerative braking have been demonstrated with electrochemical batteries and FESSs [1][2], there are significant benefits to using a FESS. Although electrochemical batteries can have high energy densities, their power densities are typically lower than those for a FESS [3][4]. Furthermore, present electrochemical battery designs require design tradeoffs between power density, energy density and system life. Because FESSs use two distinct components, the flywheel and the motor/generator, to perform the energy storage function and the power conversion function respectively, both of these key parameters can be optimized for

a specific application. Additionally, because a FESS requires little maintenance and can have a service life equivalent to that of a commercial vehicle (compared to the high maintenance requirements for a two to four year life cycle on a conventional electrochemical battery), the life cycle cost for a FESS should be less than that for a comparable electrochemical battery.

UT-CEM and its industrial partners (AVCON Inc. and AlliedSignal Aerospace Equipment Systems) are developing two FESSs for power averaging in hybrid electric drive trains. The first of these is a 2 kWh, 150 kW, 40,000 rpm device for a 13,600 kg transit bus; this device is illustrated in figure 1. In this system, the primary benefits of the FESS are to facilitate the use of a smaller prime power unit, reduce fuel consumption, and reduce emissions. The second system is a 165 kWh, 3 MW, 15,000 rpm system for a high speed locomotive where a load leveling flywheel is required for faster acceleration and reduced emissions. The combination of a flywheel with a hybrid drive train provides the locomotive with the additional benefits of regenerative braking and improved performance without the additional infrastructure expense normally required for all-electric trains [5]. At present, for these two transit applications, where the vehicle is frequently stopping and starting and where fuel costs are a large part of the operating costs, a flywheel battery appears to have a reasonable payback period. Other potential near term uses for flywheel systems are in passenger cars and as uninterruptable power supplies in stationary and mobile applications.

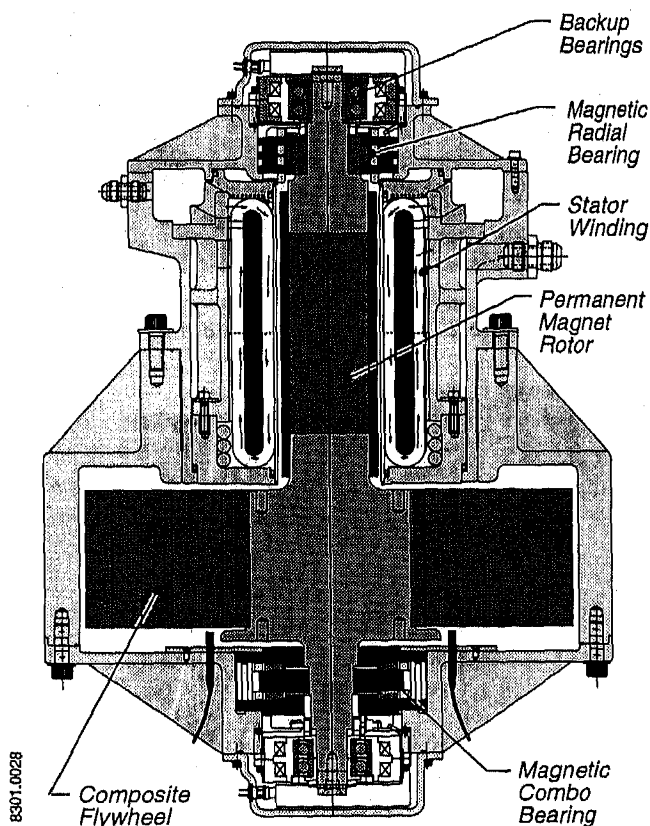


Figure 1. Transit bus flywheel battery

THE NEED FOR CONTAINMENT

A flywheel's energy storage potential is proportional to its mass moment of inertia and the square of the rotational speed. For a specific configuration, speed is limited by the material strength-to-density ratio. In terms of performance, high strength composite materials are currently the preferred flywheel material [6]. Furthermore, while metal wheels may provide a lower cost, albeit heavier, alternative for small stationary applications, composite flywheels will be required to meet the weight requirements for mobile applications. It is also apparent that to maximize energy storage in composite flywheels, they must operate with stresses near their yield strength.

Based on a long history of operating experience, a great deal is known about the failure modes of metallic flywheels and the public has grown accustomed to using them in many everyday applications. This includes their use as power smoothing devices in manual transmission passenger cars, where the flywheel resides in close proximity to both the driver and passenger. This is of no concern to the average driver, and because metallic wheel failure mechanisms are well characterized, there is little reason for concern. Very little data, however, has been published about the failure modes of composite flywheels. What is known is that these failure modes (described later in this paper) are very different from the well characterized behavior of metallic wheels. It is also known that they are both design and material dependent, so accepted design rules have not yet been established. Additionally, catastrophic burst failures during overspeed of certain flywheel designs have been demonstrated during failure testing [7][8][9][10]. Based on the potential for a catastrophic failure and the lack of operating experience with composite flywheels, it is prudent to incorporate containment structures to mitigate the effects of a failure in near-term vehicular FESSs. While this is the approach taken by the U.S. FESS industry, it should be noted that low stress vehicular flywheels have been successfully operated without dedicated containment systems in Germany for many years, with no serious accidents reported.

DARPA FLYWHEEL SAFETY PROJECT

To address the aforementioned concerns, DARPA has funded a Flywheel Safety Project to both determine the effects of typical composite flywheel failures and develop systems to mitigate those failures. The objectives of this project are to develop burst test technology; conduct a combined theoretical and experimental program to gain a fundamental phenomenological understanding of composite flywheel failures; and devise a basic methodology for design of containment structures and systems. The program is fully integrated with Department of Energy efforts in composite flywheel containment and emphasizes technology sharing within the U.S. flywheel battery community. A major outcome of the project will be a final report documenting test results and development of containment design procedures for the U.S. flywheel battery community.

The project couples the major commercial U.S. flywheel spin-test facility, Test Devices, Inc., with a powerful group of competing flywheel battery developers: UT-CEM, Trinity

Flywheels Inc., and U.S. Flywheels Inc. The genesis of this project and the enthusiasm for mutual cooperation stems from the belief that flywheel safety is a problem of such importance and difficulty that it transcends normal competitions. The group membership was developed to pool resources and assemble a combined expertise, representing a suitable variety of flywheel battery design approaches, so that a relatively general approach for design of composite flywheel containment systems could be developed. The commitment to share intermediate test results, interim progress, and final project results stems from altruistic concerns about personnel safety as well as concerns about public perception and the overall health of the flywheel battery market potential. Although the project is relatively new, some results (specifically, results of a detailed literature search) are already available on the internet at the Southern Coalition for Advanced Transportation site, www.advtrans.org.

TECHNICAL BACKGROUND

COMPOSITE FLYWHEEL FAILURE SCENARIOS - Different composite flywheel designs possess a propensity for different types of failures. However, the reader should also be aware that similar designs can experience different failure modes if operating conditions or fabrication techniques are varied. If failure does occur, the more likely scenarios are:

- delamination (i.e., radial separation of layers of composite windings),
- burst, and
- transverse epoxy matrix failure.

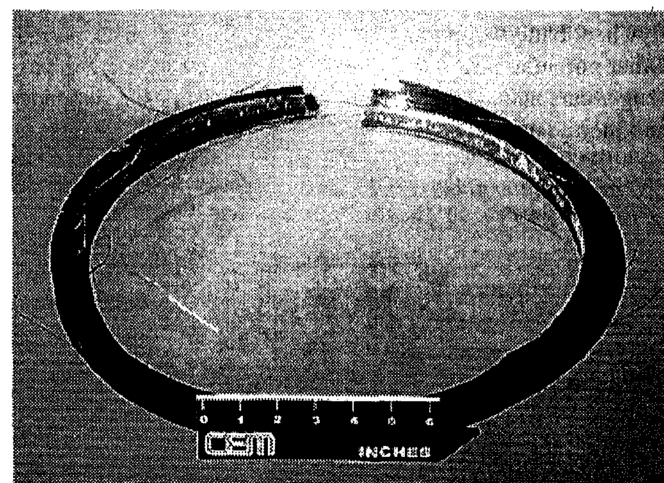
These scenarios with causes and effects are listed in table 1. Order of presentation in the table is in the "plane of action" of the event based upon a cylindrical coordinate system.

Table 1. Summary of failure scenarios

Failure Scenario	Cause	Effects of Failure
In the	*Radial (R)*	*Plane*
Delamination: Ring radial separation	• Over speed • Epoxy matrix degradation and/or shifting of imbedded fibers (creep) • Elevated temperature	Maybe benign for some designs: Produces imbalance due to mass shift requiring flywheel shutdown
In the	*Theta (θ)*	*Plane*
Burst: Hoop fiber crack	• Over Speed • Fatigue	Catastrophic with significant energy release
In the	*Axial (Z)*	*Plane*
Matrix crack transverse to hoop fibers	• Induced by high axial tensile stress from Poisson effects	Benign in limited occurrences. May produce imbalance

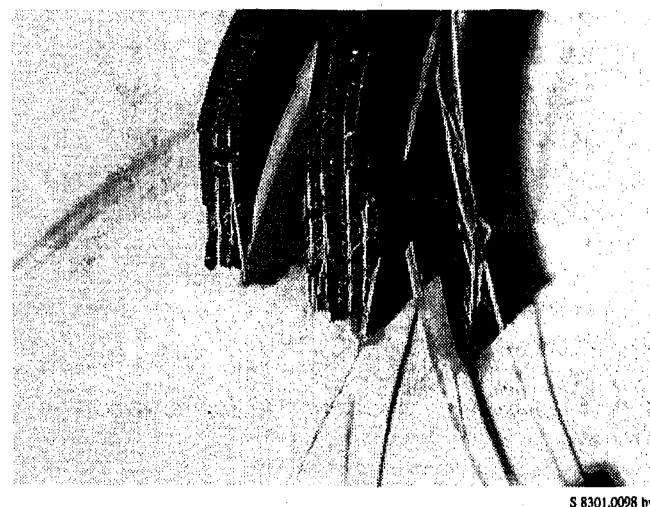
Mass loaded flywheels, i.e. flywheels with radially increasing specific modulus, generate radial compression with increasing speed. Generally, this design can be more susceptible to burst, especially if loss of the outer ring results in hoop failure of all the inner rings. A burst is a failure of fibers in the hoop direction. Figures 2 and 3 show different views of a graphite composite ring which underwent a burst type failure in static tests. Bursts can be caused by overspeed, fatigue, or fabrication flaws, which either exceed or compromise fiber hoop strength resulting in significant kinetic energy release.

Preloaded flywheel designs, by contrast, exhibit constant or near constant specific modulus. They are characterized by a tendency to delaminate before experiencing a burst, especially when employing lower modulus graphite composites. Delamination is a physical separation of a localized portion of the flywheel due to loss of radial compression, leading to a debond between hoop fiber layers. Causes include high spin loads due to overspeed. Debonds can also be initiated by creep of the epoxy matrix or matrix degradation due to elevated temperature. Polar woven flywheel designs are resistant to delaminations due to the inclusion of radially oriented fibers.



S 8301.0093 bw

Figure 2. Burst of a composite ring



S 8301.0098 bw

Figure 3. End view of the burst of a composite ring

Figures 4 and 5 show views of a delaminated composite ring. Delamination of this ring was caused by a fabrication anomaly. However, the pictured radial separation is representative of a delamination during flywheel operation.

Spin tests conducted at Oak Ridge National Laboratory [11] demonstrated a particular preloaded flywheel design which delaminates during overspeed conditions. This flywheel design failed in a benign mode referred to as a mechanical fuse. At a specific speed, significantly less than burst speed, the delamination event produced a slight mass imbalance which was detected by shaft displacement probes. The flywheel was safely shut down, avoiding a large energy release associated with a flywheel burst.

It should be emphasized that both mass loaded and preloaded flywheels can be designed to delaminate. In addition, not all delaminating flywheel designs possess a benign failure mode. An abrupt, fast acting delamination can result in significant imbalance forces quickly overcoming bearing load capacity. The result is a loss of control of the flywheel leading to flywheel failure.

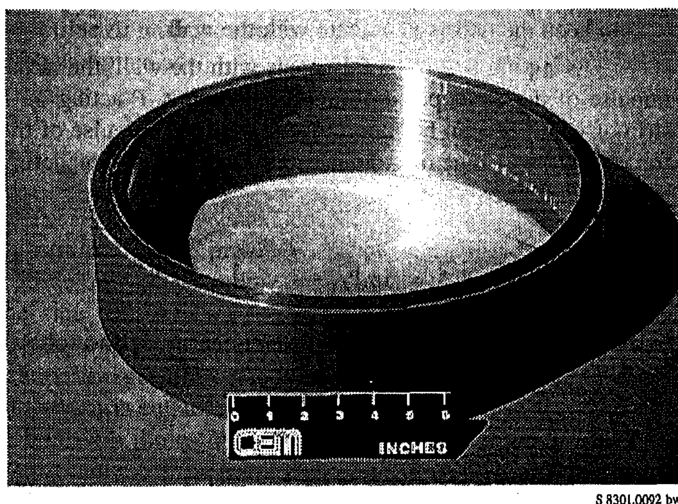


Figure 4. Delaminated composite ring

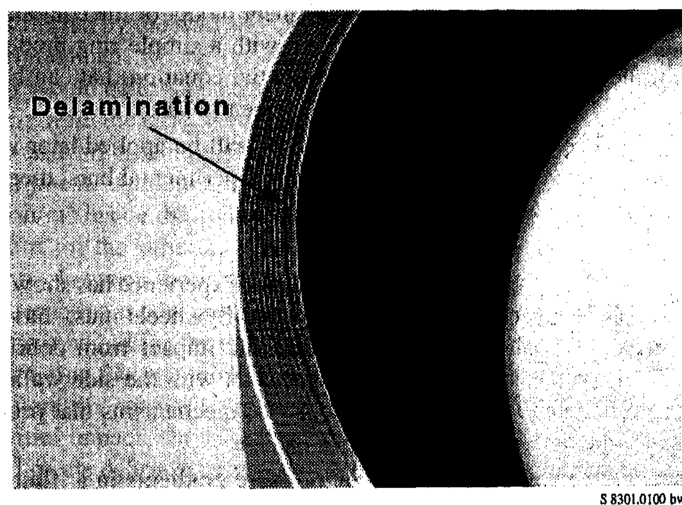


Figure 5. End view of delaminated composite ring

Another noted failure scenario [11] is cracking of the epoxy matrix transverse to hoop fibers. This is usually induced by high axial stresses from Poisson effects due to radial growth. According to the literature, this failure usually produces mild imbalances permitting safe flywheel shutdown.

CONTAINMENT LOADING ESTIMATES - The initial containment loading estimates have focused on the flywheel burst mode of failure, because it is generally agreed that this type of failure would represent the most extreme challenge to the containment. The burst mode of failure is characterized as a sudden failure of the rotor fibers in the hoop direction. For a partial burst, it is assumed that only the most highly-stressed outer rings of the rotor fail. For a complete burst, the entire rotor is assumed to fail instantaneously.

The size and shape of the fragments that are ejected from a flywheel burst are currently being investigated experimentally. If the fragments are large, they can have a sizable component of rotational kinetic energy in addition to their translational kinetic energy. Conservation of momentum can be used to show that the fragment containing the greatest amount of energy is a tri-hub sector that encompasses one third of the rotor. Although such fragments are often assumed for the analysis of the containment of metallic rotors, experience has shown that composite rotors tend to fracture into much smaller pieces.

The seminal work in the analysis of composite flywheel breakup was performed by Coppa [12]. Coppa's model assumes an axially symmetric distribution of finite-sized fragments that impact the containment wall along a tangential trajectory. The momentum transferred to the wall is calculated assuming that the fragments hit the wall and then lose momentum progressively as the fragments crush. In the analysis described below, we have taken a somewhat simpler view that the flywheel burst itself produces sufficiently finely divided fragments such that the debris can be modeled as a continuum during its impact with the containment wall. Experimental results support this assumption.

Radial Wall Loading - The simplified geometry used for the debris loading calculations is shown in figure 6. The inner radius of the rotor is "a" and the outer radius is "b". For complete burst of the flywheel, "a" corresponds to the radius of the arbor, whereas for a partial burst, "a" corresponds to the inner radius of the outer ring. The inner radius of the containment is "c" and its thickness is "h". At time t=0, the rotor, spinning with an angular velocity ω, is assumed to instantaneously break into a large number of fragments of characteristic size Δm. Each fragment is assumed to travel an unimpeded trajectory "s" until it hits the containment wall.

The containment loading for finely divided debris can be determined by modeling the particles as a continuum rather than as discrete fragments. That is, the particles that originally occupied the rotor annulus between r and $r + \Delta r$ are assumed to impact the containment wall between time t and $t + \Delta t$, where t=0 is the time that the rotor disintegrates. The time t at which the particles from radius r reach the wall is related to the radius and the path length s by $s = r\omega t$. Because the particles

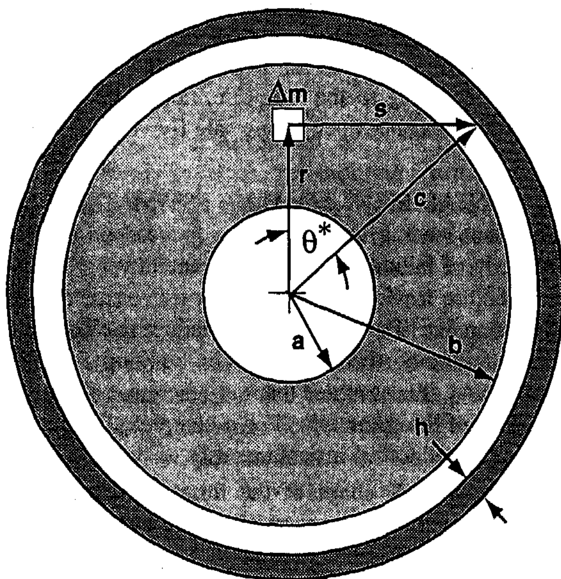


Figure 6. Load calculation geometry

traverse a tangential trajectory, the path length is related to the containment radius c and the original particle radius by the relationship

$$s^2 = c^2 - r^2 = r^2\omega^2 t^2 \qquad (1)$$

The first particles to arrive at the wall come from the outside radius r = b. The last particles to arrive at the wall come from the inside radius of the rotor r = a. The time span over which the impact event occurs ranges from t_0 to t_f, where, from equation (1)

$$t_0 = \frac{\sqrt{c^2 - b^2}}{b\omega} \quad \text{and} \quad t_f = \frac{\sqrt{c^2 - a^2}}{a\omega} \qquad (2)$$

The time increment Δt over which the fragments from the flywheel annulus between r and r + Δr arrive at the wall can be found by taking the differential of equation (1). In differentiating this equation, allowance is made for the fact that the containment radius is increasing during the impact period by a radial displacement w(t). At any time t the containment radius is given by $c(t) = c_0 + w(t)$, where c_0 is the original radius at t = 0. Although w/c_0 is expected to be small, dw/dt can be significant compared to the impact velocity of the fragments. Differentiation of equation (1) gives

$$\frac{\Delta r}{\Delta t} = -\frac{ct\omega^2}{\left[1+\omega^2 t^2\right]^{3/2}} + \frac{dw/dt}{\left[1+\omega^2 t^2\right]^{1/2}} \qquad (3)$$

The mass of the material Δm that came from the radius between r and r + Δr is given by Δm = ρLrΔθΔr, where ρ is the flywheel density and L is the flywheel length. Substituting this result into equation (3) yields the quantity of mass that impacts the wall during time Δt.

The flywheel fragments impact the containment wall at an oblique angle θ* (fig. 6) which depends on the radius from which they originated. The component of velocity normal to the wall is just $v_{n1} = r\omega \sin\theta^*$, or, from the geometry of figure 6,

$$v_{n1} = \frac{c\omega^2 t}{\left[1+\omega^2 t^2\right]} \qquad (4)$$

It is assumed here that the tangential component of the fragment momentum is conserved during impact and the coefficient of restitution for the normal component is equal to zero. The normal velocity v_{n2} of the particles after impact is, therefore, equal to the wall velocity dw/dt. This implies that the debris follows the radial motion of the wall during impact, and is not reflected back toward the axis of the flywheel, potentially doubling the load. The assumption of plastic impact can be justified by the observation that the impact is likely to result in further fragmentation of the debris, and the centrifugal force from the rotational component of the fragment velocity will tend to keep the debris in contact with the wall.

During impact of the fragments with the wall, the radial impulse of the force per unit area, or pressure, P acting on a strip of width Δθ can be found by equating the impulse of the force with the change of momentum of the particles impacting that strip. That is,

$$-Pc\Delta\theta L\Delta t = \Delta m\left[v_{n2} - v_{n1}\right] \qquad (5)$$

Substituting equations (3) and (4) into equation (5) and solving for the pressure gives

$$P = \frac{\rho c^2\omega^4 t^2}{\left[1+\omega^2 t^2\right]^3}\left[1 - \frac{dw}{dt}\frac{\left(1+\omega^2 t^2\right)}{c\omega^2 t}\right]^2 \qquad (6)$$

for $t_0 \le t \le t_f$. The impact loading function given by equation (6) can be used to drive a finite element model of the containment response, or can be combined with a simple ring model of the containment to yield a differential equation that can be solved directly for the radial displacement w(t).

The containment loading model will be applied later in this paper to predict the loading on an experimental burst measurement device (fig. 12 and table 3).

Axial Loading on the End Plates - Experience has shown that a viable containment for composite flywheels must have adequate axial strength to withstand the impact from debris that is deflected axially following impact with the side walls. Unfortunately, little is known about the mechanisms that produce the deflection. Possibilities include

1) the development of axial pressure gradients in a fluid-like bed of debris that builds up on the containment wall,

2) scattering of fragments at the wall from the impact of incoming fragments, or

3) fracture of fragments along oblique planes as they hit the wall, creating a wedge that drives the two pieces toward either end of the containment.

Whatever the cause, equal axial momentum must be directed toward each end plate. An upper bound on this momentum can be found by assuming that the kinetic energy lost during impact with the wall is converted to axial motion of the fragments. Although the conversion efficiency has been assumed here to be 100%, a more realistic value could be determined experimentally, thus avoiding the need to model the complex process of debris deflection. It is further assumed that the debris does not accumulate at the wall, but splits into two streams with the same axial velocity. Since the radial component of the fragment velocity v_{n1} is known from equation (4), and the radial velocity of the fragments and the wall following impact is dw/dt, conservation of energy dictates that the axial velocity v_z of the fragments following impact is $v_z = v_{n1} - dw/dt$. If debris does not accumulate at the wall, it impacts the end plates with the same mass flow rate as it arrives at the wall. This rate is given by equation (3) and the relationship between Δm and Δr. Equating the change in momentum of this mass over time Δt with the axial impulse gives the axial force F(t) on each end plate as

$$F(t) = \frac{1}{2}\frac{dm}{dt} v_z \tag{7}$$

Substitution of the previous results for dm/dt and v_z produces a force on each end plate that is just 1/2 the force exerted on the side wall by P(t). Although the assumptions made here appear to be extreme, it is noted that the model typically yields kinetic energies imparted to the end plates and the axial motion of the debris equal to 3 to 5% of the rotor energy, which is consistent with the estimate from at least one rotor test that produced failure of the lid of the containment vessel [13].

Torque- The calculations given above for the normal loading on the containment wall and the end plates assumed that debris does not accumulate at the wall, but rather is deflected by the wall and the end plates toward the center of the containment where, presumably, much of the remaining energy would be dissipated by collisions between the fragments. If this were the case, the swirling motion of the debris would not exert a significant centrifugal or torsional loading on the containment. Although this picture may be reasonable if the free volume inside the containment is large, it would not be reasonable for the other extreme. Here a debris bed would likely form between the intact portion of the rotor and the wall. An alternative model has therefore been developed to estimate the upper bound for the torsional loading on the containment, recognizing that the model is not consistent with that used to calculate the normal force on the wall and the end plates during initial impact. In calculating the maximum torsional loading, all of the debris is assumed to accumulate next to the wall as it swirls inside of the containment.

Figure 6 can also be used to model the debris bed geometry. Here, each fragment is stopped at the inner radius of the debris bed, instead of reaching the wall. Again, the flywheel is assumed to break up instantaneously. The fragments from the outside of the flywheel impact the wall at the same time t_0 given previously in equation (2). These particles now remain against the wall. Particles from flywheel radius r impact the inner radius $r_2(t)$ of the debris bed at a later time. The inner radius of the debris bed can be found from the conservation of mass and the porosity of the bed γ. Furthermore, the time t at which the particles impact the inner radius is related to their tangential velocity at the time of breakup. The geometric relationships indicated in figure 6 are also valid here if the radius c is replaced by r_2. In a manner similar to the derivation given previously, these relationships can be used to calculate the rate at which mass is added to the debris bed and the radial and tangential components of the velocity of the fragments as they join the bed. The change in angular momentum of the debris at the wall consists of two components. One is due to the addition of new debris from the disintegrating flywheel, and one is due to the slowing down of the debris from the torsional forces at the wall. Assuming that the layers of the debris bed are locked together by the random packing of the fragments, the common angular velocity of the fragments ω_d and the torque T_p on the containment wall can be calculated by equating the torsional impulse with the change in angular momentum.

The pressure exerted on the containment wall from the centrifugal force of the swirling debris can be calculated by considering the equation of motion of a differential element of the debris. The radial component balances the centrifugal force and the differential pressure on the radial faces of the element. However, unlike an intact flywheel rotor, the debris cannot support a tensile load in the hoop direction. The equation of motion therefore reduces to

$$\frac{d}{dr}(r\sigma_r) = -\rho(1-\gamma)r^2\omega_d^2 \tag{8}$$

Equation (8) can be integrated from the inner radius of the debris bed r_2 to the outer radius c. The pressure at the inside radius is small because the normal velocity of the particles impacting the bed is small. Integration of equation (8) then gives the pressure at the wall N_p as

$$N_p(t) = \rho(1-\gamma)\omega_d^2 \frac{\left(c^3 - r_2^3\right)}{3c} \tag{9}$$

The torque on the containment ring T_p is just

$$T_p(t) = 2\pi c^2 L\mu N_p(t) \tag{10}$$

where μ is the coefficient of friction between the fragments and the wall. In general μ has to be determined experimentally. However, if torque is limited by the interlaminar shear strength of the composite fragments (100 to 120 MPa), the effective coefficient of friction could not exceed about 0.3. This is in reasonable agreement with the value of μ (0.36)

needed to match the model results with the torque and spin-down times measured by Steele and Babely [14] during failure of a 0.695 MJ composite rotor.

The torque predicted by equation (10) can be significant and in some applications may be excessive (see predicted torque for a partial rotor burst presented in table 3). For vehicle applications, for example, it is likely that some torque mitigation features must be incorporated into the containment. A discussion of containment concepts which can reduce torque loads is presented later in this paper.

TESTING AND CONTAINMENT DESIGNS

TEST RESULTS - To aid in the design of candidate containment devices, a number of tests will be performed. These include spinning composite flywheels to failure to understand and identify failure modes. Flywheels will also be spun to failure in heavily instrumented burst measurement structures to measure loads. Measuring and understanding flywheel failure loads is key to successful containment designs. Finally, flywheels will be spun to failure in real containment structures to proof test these designs.

Recently, UT-CEM successfully burst a multi-ring, preloaded composite flywheel. This flywheel design was an altered operational design. The operational design, which delaminates before bursting, was altered by substituting weakened graphite material in the fabrication of the outer ring.

Main objectives of the spin test were to demonstrate a controlled composite flywheel burst and then to attempt "soft capture" of the flywheel fragments using a Kevlar catch ring. As designed, only the outer flywheel ring failed at burst. An overview of test results are presented in table 2.

The test flywheel stored 5.8 MJ of energy at burst speed and released 1.1 MJ during the burst event. The flywheel prior to testing is shown in figure 7 and is shown again in figure 8 after testing. Other than failure of the outer composite ring as intended, little damage occurred. Damage was limited to surface wear as shown in figure 9.

Table 2. Overview of test results

•	Outer flywheel ring successfully burst at 35,200 rpm (830 m/s) within the predicted burst speed range of 33,400 rpm to 37,600 rpm
•	Outer ring burst did not cause the remaining flywheel rings to fail
•	57% of the failed ring material was recovered following the test
•	Significant mass of flywheel fragments (80% of the recovered material) was captured by the Kevlar catch ring
•	During multiple low speed runs with dwells and the final run to burst, the flywheel maintained excellent mass balance
•	After outer ring burst at 35,200 rpm and separation from the spindle, the remaining flywheel fell to the bottom of the pit and "spun down". Afterward, inspection of the flywheel showed no damage other than surface wear

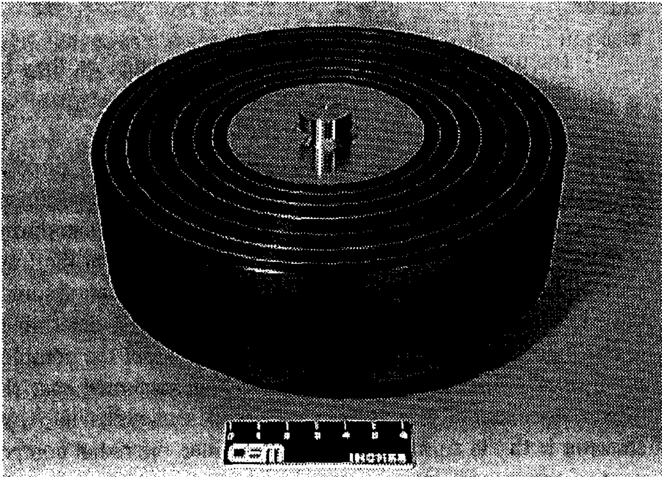


Figure 7. Composite flywheel prior to burst

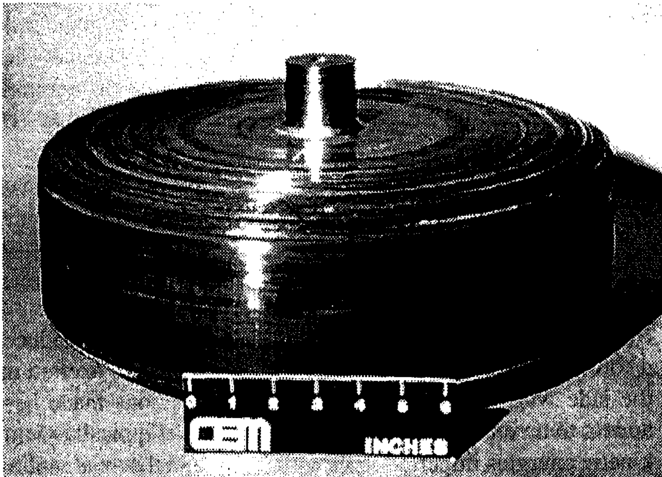


Figure 8. Composite flywheel after burst



Figure 9. Surface detail of composite flywheel after burst

Other flywheel developers have attempted to induce flywheel bursts with mixed results. This test delivered a remarkably clean burst event. Advantages of achieving a successful burst are twofold.

First, bursting a composite flywheel ring clearly establishes both induced and allowable load limits of the composite laminate structure. This correlates analytical tools for designing safe flywheels.

Second, a flywheel design that can be tailored to deliver specific burst energy levels to candidate containment designs presents a valuable tool to assist the DARPA Flywheel Safety Project. Attempts to catch burst fragments are relevant since evaluation of their shape and size may lead to additional insight for improved containment designs.

Flywheel fragments recovered from the Kevlar catch ring varied in size with a high percentage (30%) of significant size. An example of some of the larger flywheel fragments recovered from the catch ring is shown in figure 10.

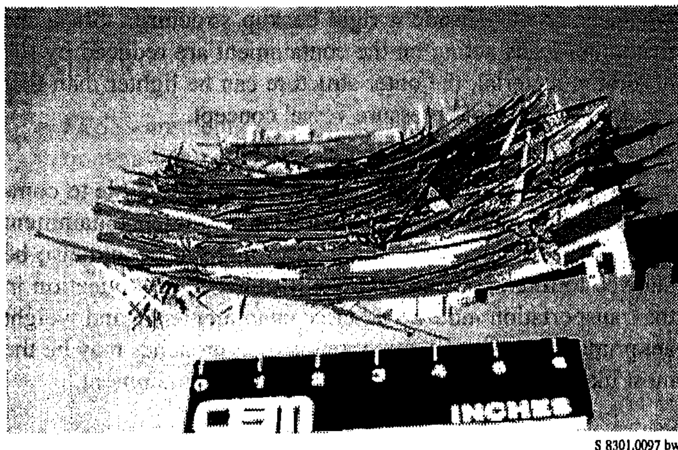


Figure 10. Captured composite flywheel fragments

UPCOMING TESTS - Upcoming tests include additional flywheel bursts early this year. The bursts will occur inside an instrumented steel burst measurement device with the objective of measuring radial, axial, and torque loads.

While the tests described above will use rings designed to burst at overspeed, another test will use a preloaded flywheel designed to delaminate. The purpose of this test is to explore this failure mechanism and its consequences.

Using the loading estimate model previously described, the burst measurement device is being designed to survive burst testing without permanent deformation, allowing it to be reused for repeated tests; an illustration of the device appears in figure 11. Loading calculations for this device appear in table 3, and a curve for the predicted radial impact loading is shown in figure 12.

The burst measurement device will be instrumented so that loads predicted by the loading estimate model can be verified by experiment. In addition, the device is designed so that energy-absorbing materials can be inserted inside the vessel, allowing testing of some of the concepts presented later in this paper.

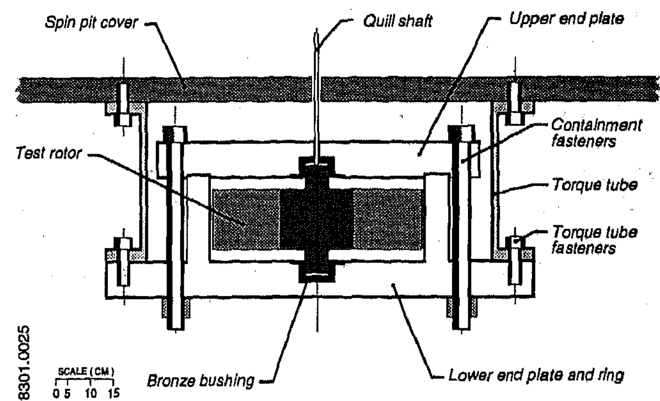


Figure 11. Burst measurement device

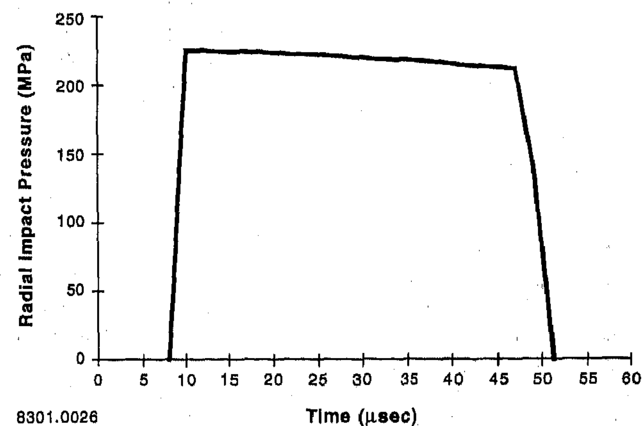


Figure 12. Predicted radial impact loading

DESIGN CONCEPTS FOR CONTAINMENT - Because of the lack of operating experience and test data with composite flywheels, design of adequate containment presents a formidable challenge to the designer. There is a substantial body of experience dealing with metallic flywheel failures, but as discussed earlier, it is unclear how much of that experience can be applied to composite flywheels. In spite of this fact, a number of containment strategies are emerging as potential approaches to the problem. Each has its own set of advantages and disadvantages and each must undergo intensive testing to prove its suitability as a viable containment strategy. In the following discussion, it is assumed that the containment structure must contain a true flywheel burst, which is thought to present the highest forces and energy levels to the containment structure. In all the design concepts discussed below, the containment will be considered successful and effective if fragments from a burst flywheel are completely contained within the containment structure.

Rigid Pressure Vessel - In this approach, a heavy-walled pressure vessel envelops the flywheel, as illustrated conceptually in figure 13. This is a very direct approach to the problem:

the vessel has sufficient strength to contain burst forces without the use of any dedicated energy absorbing materials, although it is acceptable for the vessel itself to sustain permanent deformation during the burst event. In fact, it is probably desirable for the vessel to deform plastically, absorbing a substantial portion of the flywheel energy in the process. Advantages to this approach include simplicity of design and the possibility of substantial safety margins. Disadvantages include difficulties in limiting torque loads on the mounting system and high containment system weight.

Table 3. Burst measurement device load conditions

Load conditions for a flywheel outer ring burst at 35,000 rpm inside burst measurement device

Rotor Information:		
Outer Ring Outer Diameter	0.445 m	17.5 in.
Outer Ring Inner Diameter	0.414 m	16.3 in.
Length	0.127 m	5 in.
Speed at Burst	35,000 rpm	35,000 rpm
Mass	4.46 kg	9.84 lb
Tip Speed	815 m/s	2660 ft/s
Stored Kinetic Energy in Outer Ring	1.38 MJ	1.02×10^6 ft-lb
Measurement Device Shroud Ring Information:		
Radial Thickness	0.051 m	2 in.
Inner Diameter	0.521 m	20.5 in.
Material	4340 Steel	4340 Steel
Predicted Measurement Device Loads:		
Radial Impact Load Duration	42 µs	42 µs
Max. Radial Impact Pressure	225 MPa	32,600 psi
Max. End Force	2.33×10^7 N	5.24×10^6 lb
Approx. Max. Hoop Stress in Shroud Ring	883 MPa	128,000 psi
Shroud Ring Max. Radial. Displacement.	1.12 mm	0.044 in.
Shroud Ring Max. Strain	0.4%	0.4%
Shroud Ring Max. Torque	1.00×10^6 N-m	7.39×10^5 ft-lb

Rotating Liner - The strategy incorporated in this design approach reduces the torque that must be transmitted from the flywheel battery to its mounting system (fig. 14). A rotatable liner inside the containment vessel is impacted by the swirling debris from a rotor burst, absorbing debris rotational energy. The overall effect of the liner rotation is that tangential loading forces acting on the containment are spread over a longer period, thus reducing the maximum torque loads on the containment vessel. In addition, a liner braking mechanism is provided to impart a braking torque to the rotating ring, limiting its maximum angular velocity and dissipating rotational energy. The primary advantage to this approach is that the torque capacity required in the FESS mounting system is reduced. This is achieved at the cost of increased containment system complexity.

Energy Absorbing Liner - This approach includes an energy-absorbing material within the containment, designed to absorb a substantial portion of the debris kinetic energy (fig. 15). As presently conceived, the energy-absorbing material would be located inside a rigid backup structure. Since the maximum loads acting on the containment are reduced by the absorbing material, the outer structure can be lighter than that required for the rigid pressure vessel concept.

Combination of Approaches - It is also possible to combine the approaches listed above into a single containment device. For example, an energy-absorbing liner could also be allowed to rotate, reducing torque loading. For application in the transportation industry where containment size and weight are primary concerns, a combination of approaches may be the most likely solution to providing adequate containment.

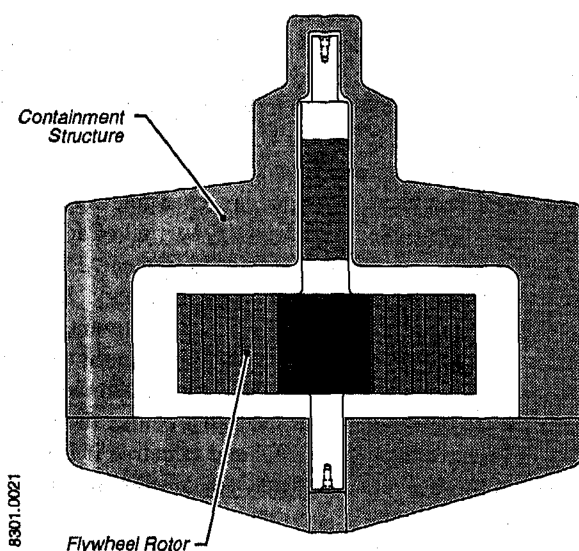


Figure 13. Rigid containment structure

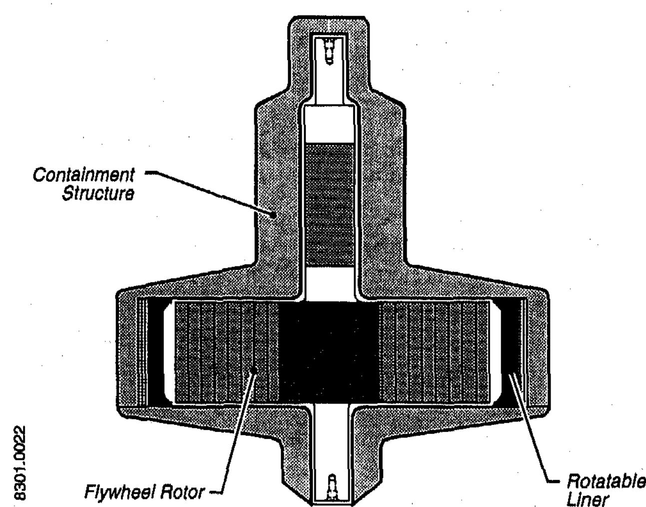


Figure 14. Containment with rotating liner

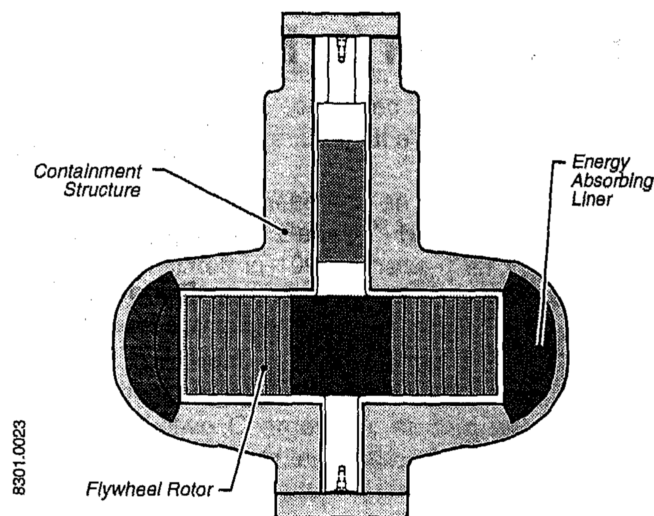


Figure 15. Containment with energy-absorbing liner

SUMMARY

- FESSs are being developed for power averaging applications in the transportation industry.
- A potential for flywheel bursts exists with FESSs that incorporate composite flywheels.
- Failure mechanisms of composite flywheels historically have been poorly understood due to lack of operating experience and experimental data. Test programs are underway which are improving the understanding of these systems.
- Failure modes of composite flywheels are design and material dependent; some designs may prove to fail in a benign manner.
- In the near-term, FESSs using highly stressed composite flywheels should incorporate containment as protection against flywheel bursts.
- A DARPA-sponsored containment development program brings together prominent flywheel developers in an effort to identify a methodology for safe composite flywheel and containment design.
- A theoretical model has been developed to predict containment loads during burst conditions; a testing program has begun to confirm the theoretical predictions.
- In the long-term, operating experience may show that composite flywheels can be designed to fail in a benign manner, so that containment is unnecessary. Alternatively, a better understanding of flywheel failure behavior may lead to lightweight containment designs that take advantage of energy-absorbing features.

ACKNOWLEDGMENTS

The authors wish to thank Robert Rosenfeld of DARPA and Alan J. Sailer of Houston Metro for their support of this work. In addition, we wish to thank Lori Moore, Daniel Gleesing, and Roy Pena of UT-CEM for their assistance in the publication of this paper.

REFERENCES:

[1] Reiner, G., "Current State in the Application of Magnetodynamic (Flywheel) Storage Systems in Urban Transport Buses," presented at Flywheel Energy Storage Technology Workshop, Oak Ridge, Tenn., November 1995.

[2] Drawe, J., Bluebird Body Company, Fort Valley, GA, personal communication, December 1996.

[3] Reiner, G., et. al., "Experiences with the Magnetodynamic (Flywheel) Storage System (MDS) in Diesel Electric and Trolley Busses in Public Transport Service", presented at Flywheel Energy Storage Technology Workshop, Oak Ridge Tenn., Nov. 1993.

[4] Carey, J. "Where's the Juice?", Business Week, pg. 110, May 30, 1994.

[5] Tsai, Tom and Price, John, "Advanced Non-Electric Propulsion System for High Speed Rail," 1995 International Mechanical Engineering Congress and Exposition, San Francisco, CA, November 1995.

[6] Stull, S. and Post, R.F., "A New Look at an Old Idea" Science and Technology Review, pp.12-19, April 1996.

[7] "Spin Test Tragedy" Performance Materials, pp1-3, Volume 11, Number 4, Nov. 20, 1995.

[8] Rabenhorst, D.W. and Small, T.R., "Spin Test of the University of California Lawrence Livermore Composite Flywheel at the Applied Physics Laboratory of the Johns Hopkins University", UCRL 13910, pp 1-14, July 18, 1978.

[9] Coppa, A.P., "Composite Flywheel Containment Housing Design and Testing", Section V, pp 1-12, UCRL 15448-Vol.2, DE85 001042.

[10] Kass, M.D. et al,"Evaluation of Demo 1C Composite Flywheel Rotor Burst Test and Containment Design" pp. 14-20, Oak Ridge National Laboratory TM-1315, March, 1996.

[11] "Project Summary of Flywheel Testing and Evaluation," R.S. Steele, Oak Ridge National Laboratory, CONF-820827--7,1982.

[12] Coppa, A. P., Flywheel Containment and Safety Considerations, Proceedings of the NASA Flywheel Inertial Energy Storage Workshop, NASA Conference Publication 2346, pp. 243-253, Huntsville, Alabama, Feb. 7-9, 1984.

[13] Sonnichsen, H. E., Ensuring Spin Test Safety, Mechanical Engineering, Vol. 115, pp. 72-77, American Society of Mechanical Engineers, New York, December, 1993.

[14] Steele, R. S., Jr., and E. F. Babelay, Data Analysis Techniques Used at the Oak Ridge Y-12 Plant Flywheel Evaluation Laboratory, Proceedings of the 1980 Flywheel Technology Symposium, pp. 423-433, CONF-801022, Scottsdale, Arizona, 1980.

970243

EV Battery Life Extension in Field Testing

Raymond S. Hobbs
Arizona Public Service

Frank A. Fleming
Hawker Energy Products

Donald B. Karner
Electric Transportation Applications

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Results of the Charger Test Project conducted by Arizona Public Service (APS) indicated the importance of the charger/charge algorithm/battery interface with respect to battery capacity. It also indicated that EVs rapidly charged demonstrated longer cycle life in field testing without reduction in capacity. Additional testing conducted by APS has shown that cycle life of the batteries is strongly related to the charging scheme utilized. Application of results of laboratory charging to the charge scheme utilized in this testing has been hampered by extremely long test periods in field tests designed to cycle batteries to the end of life. Development of a predictive method of cycle life extrapolation based upon microscopic analysis promises to reduce the time required for battery cycle life testing.

INTRODUCTION

Three phases of testing will be addressed. Phase I testing was performed on EVs in the field to evaluate the performance of Hawker Genesis® batteries with a variety of chargers. Based on the results of the Phase I testing, Phase II testing was initiated to examine battery performance in the field at various rapid charge rates. Phase III addresses microscopic analysis of batteries that have been cycled at different rates to determine if the end of useful life can be predicted.

PHASE I - *EV AMERICA* FIELD TESTING

The first round of *EV America* Performance Testing concluded in November 1994. Among the numerous bench mark performance data collected was EV range at constant speed 72.4 km/h (SAE 227A), constant speed 96.5 km/h (SAE 227A), and per fuel efficiency drive cycle (SAE J1634). Over 225 EVs were purchased in the Phase I Program (Solectria E-10 pickups and Force sedans, and US Electricar [USE] ES10 pickups and sedans) by utilities and government agencies across the United States. The *EV America* Phase I range test results for the USE ES10 and the Solectria E-10 pickups are provided in Table 1. All of these vehicles utilized Hawker Genesis® 38 amphour VRLA AGM batteries, model G190W.

Table 1. *EV America* Performance Testing, Phase I Range Testing

EV Manuf	km at 72.4 km/h	km at 96.5 km/h	km at SAE J1634
USE ES10	113.8	76.1	110.7
E-10	117.1	63.6	93.2

In use, the Phase I EVs reported a significant loss in range with low battery cycles. Proving ground range testing (SAE 227A @ 72.4 km/h) in Phoenix confirmed an approximate 50% range reduction in a USE ES10, and a 37% range reduction in a Solectria E-10.

An evaluation by Hawker Energy Products (HEPi) concluded that an undercharge condition existed in the batteries. The USE ES10 was originally equipped with a 2 kW integral charger which charged a 52 battery module, useable 21 kilowatt hour pack. For this vehicle,

HEPi concluded that a minimum 12 kW charging rate was required to properly charge the battery pack. The Solectria E-10 was originally equipped with a 2 kW charger which charged a 36 battery module, useable 16 kilowatt hour pack. HEPi concluded, for the E-10, that an 8 kW charging rate was required at a minimum.

Based on the need for an acceptable charge protocol for the Hawker Genesis® batteries, a field test program was initiated in 1995 (Phase I) for the purpose of evaluating the performance of Hawker Genesis® batteries with a variety of chargers. The following chargers were evaluated; 150 kW Norvik Minit Charger, 2 kW Solectria Charger, 3 kW Solectria Charger, and 6.6 kW Delco Inductive Charger. Six electric pickup trucks were selected for the field test. None of the Phase I EVs were equipped with a battery management system to monitor modules during charge and discharge. Thermal management of the EV packs was limited to a few small muffin fans in half of the vehicles, while the other half had no devices. This test program began in May 1995 and terminated in January 1996. The EVs were driven on the streets and freeways in Phoenix where summer daytime temperatures can reach 49° C.

Table 2 describes the chargers for the Phase I vehicles. Fast charging demonstrated its value early in the test, by restoring the capacity of a battery pack which was believed to have been irreversibly damaged through long term undercharging. The Norvik Minit Charger was used to fast charge two of the EVs at an arbitrarily chosen rate of 3C, i.e. three times the one hour discharge rating of the Hawker battery. EVs charged by the Norvik charger had to undergo a periodic equalization charge. EV138 was equalized periodically with the Norvik charger. Other vehicles were equalized by using on or off board low rate overnight chargers.

Norvik charging used 2.45 volts per cell for the Genesis® battery and equalization was based upon 2.65 volts per cell. EV137 was equalized with the Delco inductive charger. The Delco charger conducted the main charge routine with a 2.45 volts per cell basis and the equalization routine was based upon 2.56 volts per cell. The Solectria chargers used a 2.45 volts per cell and cell equalization was accomplished with extensive overcharging at this voltage.

Table 2. Phase I Field Test EV Chargers and Charge Cycles

EV	Fast Charge Rate	Overnight Ind Rate	Overnight Cond Rate	Equalization Charge
133	N/A	N/A	3kW	Cond
134	N/A	N/A	2kW	Cond
135	2C	N/A	2kW	Cond
137	3C	6.6kW	N/A	Ind
138	3C	N/A	N/A	Fast
139	N/A	6.6kW	N/A	Ind

Ind-inductive, Cond-conductive

Table 3 provides a summary of battery pack performance in the EVs and the cause of battery module failure. Although the batteries were the same in each EV, the consequences of chargers and charge algorithms is dramatically demonstrated in cycle life. In considering the battery performance in this test it was evident that the Hawker Genesis® VRLA AGM battery could be easily damaged by low power charging, and the battery performed better as the charge power increased. It was also surprising that the summer heat in Phoenix did not impact battery performance in high rate charging. One pack was destroyed by a fire, apparently caused by loose battery interconnections. This event reinforced the importance of good battery pack assembly procedures.

Table 3. Phase I Battery Pack End of Life Cycles, Over the Road km, and Module Failure

EV	Charge Cycles	Pack km	Cause of Module Failure
133	197	12,067	grid corrosion, 12 modules
134	52	1,931	thermal run away, 12 modules
135	113	5,749	thermal run away, 36 modules
137	360	26,549	grid corrosion, 8 modules
138	140	6,838	pack fire due to loose connection, 52 modules
139	120	6,240	grid corrosion, 17 modules

From the results of the Phase I field test and autopsy of modules from the EVs, it was concluded that the Norvik 3C charging provided the best vehicle performance and longest battery cycle life. The abrupt failure of 17 modules in EV139 (Delco inductive charger), caused by grid corrosion, was hypothesized to be the result of equalization or overcharging at the elevated voltage portion of the implemented algorithm. (Reference A) Coincidentally, the same number of equalization cycles (120) on EV137, charged by the Norvik charger and equalized by the Delco charger, resulted in a similar failure mechanism in 6 modules. The Solectria 2 kW charger produced battery failure due to thermal run away

early in battery life. The Solectria 3 kW charger produced failure via grid corrosion.

PHASE II - FIELD TESTING

For years, laboratory testing has been performed to determine the effects of charging methods on battery performance in EV applications. Parameters such as battery capacity, life, and grid corrosion have been examined. Results of laboratory testing, to date, have indicated that the charging scheme utilized has a direct effect on battery performance. However, testing within laboratories, even when simulated drive cycles are used for discharge, does not accurately represent the environmental rigors of fleet use. Fleet vehicles do not operate under a controlled set of conditions. Traffic conditions, ambient temperature, and operator usage are just a few of the variant conditions with which the battery must contend.

Significant cycle life extension has been achieved using high charging rates in laboratory testing. Chang and Valeriote *et al* (References B-H) have demonstrated that sealed valve regulated lead acid (VRLA) batteries can be rapidly charged with higher energy efficiency than traditional charging without detrimental effect to the battery and with significantly extended cycle life. VRLA batteries in laboratory testing were given five to ten rapid partial charge cycles and fully recovered their capacity in an extended charge.

Therefore, based on the Phase I and laboratory test results, a new program (Phase II), was commissioned. Phase II focused on: 1) partial charging at 3C, 5C, and 9C rates; 2) determination of whether periodic fast charging would reverse the damage from traditional low rate recharge; 3) evaluation of the cycle impact caused by reduction in the frequency of pack equalization; 4) evaluation of laboratory test results to improve field results; and 5) evaluation of the performance of NiCad and Nickel Metal Hydride batteries in the field environment.

Table 4 describes the field test fleet used in the Phase II program. EVs 114, 130, 136, 137, and 333 were given multiple fast charges per day and operated by a dedicated driver. This method of accelerated aging of the EV compressed the testing time. All accelerated aging vehicles were operated and maintained in accordance with test procedures. The test procedures describe the drive route, the depth of discharge strategy, and the recharge strategy for each EV in the program.

The remaining EVs were driven on arbitrary routes. In fleet testing, the EVs were driven by Arizona Public Service (APS) employees performing a certain work mission. EVs 131 and 134 were driven by meter readers. At the end of the work shift, EVs 131 and 134 were charged by the Delco charger. Twice per month these EVs were given one 3C charge on the Norvik charger.

The EVs used in Phase II Field Testing did not use battery management or battery module monitoring equipment. The thermal management equipment utilized varied between vehicles. The two EVs charged at the 9C rates (114 and 333) were equipped with module cooling systems. Specifically, EV333 was equipped with high air velocity, manually initiated, module cooling. EV114 was equipped with a phase change material provided by Shape. EVs 124, 130, 131, 133, 134, and 135 used small muffin fans, activated by battery temperature, to ventilate ambient air across the pack.

EV116 (Nickel Cadmium), used in the APS Property Construction Department, received almost 100% fast charges, with a minimal number of equalizations. The Nickel Metal Hydride vehicle (EV124) was charged by a Solectria 3 kW charger and given frequent opportunity charges in its fleet mission.

All of the EVs in Phase II were tested during the Phoenix summer, where day time ambient temperatures can be as high as 49°C. On six to eight week intervals, vehicles were given a 96.5 km/h range test (SAE 227A) to evaluate battery capacity.

Table 4. Phase II Field Test Vehicles

EV	EV Model	Modules in Pack	Battery Manuf	Fast Charge Rate	Equalization Charger
114	S10	42	Hawker	9C	Martin M
333	S10	26	Hawker	9C	Norvik
137	S10	52	Hawker	5C	Delco
136	S10	24	GNB	3C	Norvik
136	S10	24	Optima	5C	Martin M
130	S10	36	Hawker	3C	Solectria
131	S10	36	Hawker	3C	Delco
133	S10	36	Hawker	3C	Solectria
134	S10	36	Hawker	3C	Delco
135	S10	36	Hawker	3C	Solectria
124	Geo	15	GMO	N/A	Solectria
116	TEVan	30	Saft	3C	Norvik

The accelerated aging group required extensive use of fast charging. Fast charging occurred at the 3C, 5C, and 9C rates. Daily fast charge cycles varied from one to five. Battery equalizations were varied from every third to every tenth cycle, depending upon the specific EV. Therefore, the majority of EV charging and operation was with partial charge cycles including the overnight

charging. Equalization charge cycles were minimized because of the module failures due to grid corrosion identified in Phase I.

Figure 1 depicts typical rapid charges delivered at the 3C, 6C, and 9C charge rates by the Norvik charger, Model 150kW, manufactured by Norvik Technologies Inc., Mississauga, Ontario, Canada. This charger monitors the resistance free voltage by interrupting about five times per second and then adjusting the applied voltage to match the charge acceptance ability of the battery. An identifier is installed on each EV which permits the charger to uniquely identify its battery type and maximum charge rate. The Norvik charger has an equalization feature which is manually initiated. With the Hawker Genesis® battery in Phase II, charging is accomplished at 2.45 volts per cell. In addition, the battery is temperature compensated and the equalization stage is based upon 2.50 volts per cell.

Figure 1. Norvik Charger Profile

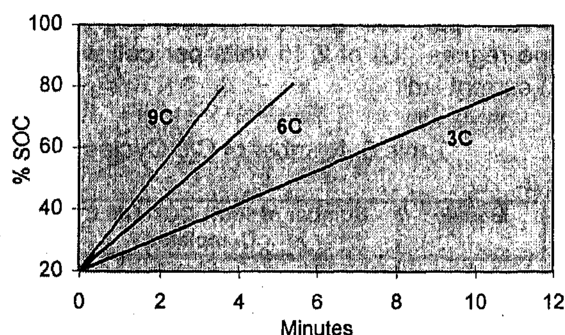


Table 5 provides the results of battery charger performance during the Phase II testing through December 1, 1996.

EVs 333 and 114 were primarily charged at the 9C rate. EV333 was equalized every sixth cycle by the Norvik charger. EV114 was charged overnight by the Martin Marietta 10 kW charger and equalized every fifth charge cycle by the Martin Marietta. Equalization voltage on the Martin Marietta charger was 2.50 volts per cell for the Hawker Genesis® batteries. During the overnight cycle on EV114, the Shape phase change material used for pack thermal stability, was frozen using a standard window air-conditioning unit to circulate air through the battery pack. Although a relatively large number of battery modules were lost by these high rate charges, the results are very promising due to the large number of modules which survived without the benefit of individual module charge control. Typical 9C charge rate module failures were observed to have a hot high rate venting of a module cell, which was detected by the Norvik charger and the charge was terminated.

Table 5. Battery Charger Performance Results Through December 1, 1996

EV	Days of Operation	Pack km	Pack Cycles	End of Life	Module Failures
114	78	2,851	108	no	2
333	165	7,413	200	no	9
137	247	16,064	251	yes	1
136/ GNB	279	6,318	97	yes	0
136/ Optima	175	4,277	137	no	4
130	367	18,008	399	no	1
131	414	6,024	140	no	0
133	610	18,270	427	yes	9
134	368	14,431	344	no	12
135	321	9,855	UNK	no	0
124	473	19,538	379	no	2
116	1,120	19,207	367	no	0

EVs 137 and 136 were charged at the 5C rate. EV137, which utilized Hawker Genesis® batteries, was charged overnight by the Delco inductive charger and equalized at 2.56 volts per cell on every third Delco charge. EV136, using Optima batteries, was equalized every sixth cycle by the Norvik charger. EV137 operated extremely well, incurring over 160.9 km (two partial 5C charges daily) in daily operation. At just under 16,090 km with no battery maintenance whatsoever, one module failed. At this point the test on EV137 was terminated. Problems with battery module interconnections in EV136 caused arcing and battery damage.

EV130 had very successful performance. The vehicle was predominately charged at the 3C rate. Overnight charges were accomplished with the Solectria 3 kW charger. EV130 showed no significant deterioration in range up to the time of its first module failure (17,915 km). The test protocol for EV130 was based on the protocol used for EV137 in Phase I. EV137, during Phase I, operated very successfully at the 3C charge rate using Delco overnight charges. While using this charge scheme, EV137 achieved about 26,549 km. In both cases, these EVs operated extremely well without battery management systems. If these systems were added to these EVs, their performance should improve. With this charge configuration, EV130 accumulated a total of 399 charge cycles and 18,008 km before incurring its first module failure. All of the fast charge cycles delivered achieved 80% State of Charge (SOC) in approximately 15 minutes.

EV131 and EV134 were driven daily in meter reading duty cycles (about 32 to 40 km per day) in Phase

II. EV134 was part of the Phase I testing where, using the Solectria 2 kW charger, 12 battery modules were lost due to thermal run away. These 12 modules were replaced and the pack charged using numerous fast charge cycles at the 3C rate. The vehicle was then moved into the Phase II program. Both EV131 and EV134, at the end of the work day, were charged with the Delco charger with equalizations occurring every fifth cycle. These EVs were fast charged once every other weekend with the Norvik charger at the 3C rate. Over the test period their range continued to deteriorate, until in the early part of November 1996 their range had dropped to about 37 km at constant speed 96.5 km/h. EV133 also continued from Phase I into Phase II field testing. In Phase I, EV133 accumulated about 12,550 km on its pack while being charged by the Solectria 3 kW charger, at which point the range had fallen to about 40 km. At this time the charging scheme for this EV was changed to being charged solely by the Norvik charger at the 3C rate. The range of the vehicle improved significantly and was then returned to fleet duty in construction where it continued to operate until abrupt pack failure at 18,270 km.

Phase II testing also included the evaluation of a Nickel Metal Hydride (NMH) battery and the Saft Nickel Cadmium (NiCad) battery provided with the Chrysler *TEVan*. The NMH is a prototype battery and is evaluated as being very promising. The NiCad battery has proven to be very durable. It has undergone a long period of partial fast charging at about the 3C rate. It has been in operation for three years. Fast charging dramatically reduced the amount of watering required by the battery. It has had extremely good reliability in the thirty months preceding December 1, 1996. There have been no battery modules replaced on this EV.

Preliminary results of the Phase II testing indicate that EVs charged at 3C and 5C charge rates perform extremely well, specifically EVs 130, 133, and 137. These EVs, with no battery management systems installed, achieved over 15,000 km on their battery packs without experiencing end of life of the batteries. In addition, charging at the 9C rate appears possible. However, further testing at the 9C charge rate is required using a battery management system.

Testing in Phase I and Phase II has extended battery life beyond the point which is reasonable to fully test in the field. The time and expense of operating vehicles in excess of 16,000 km dictates that a method be developed to predict remaining battery life such that testing to end of life is no longer necessary to evaluate the effectiveness of a particular charge protocol. In order to determine a means to predict remaining battery life, batteries with varying charge/discharge cycles were subjected to destructive analysis in Phase III of the test program.

PHASE III - ANALYSIS OF CYCLED BATTERIES

When a VRLA battery is cycled, i.e. repeatedly discharged then recharged, a host of largely irreversible processes occur internally. These processes impact the: a) morphology and crystallography of both the positive and negative active materials; b) corrosion and growth of the positive grid material; and c) electrolyte concentration and saturation level as both water loss and consumption of oxygen in the grid corrosion process occurs. Eventually, one or a combination of the above will result in unacceptable battery performance, i.e. end of useful life.

In order to correlate internal aging of the battery with use, a group of Hawker Genesis® G190W batteries were subjected to a number of cycles, as seen in Table 6, and subsequently analyzed. The batteries were cycled at the C/5 rate of discharge to an end point voltage of 1.67 volts per cell and recharged using a constant voltage regime (IU) of 2.45 volts per cell with a 0.4 C/5 (15A) current limit.

Table 6. Number of C/5 Cycles

Module	Number of 100% DOD C/5 Cycles Completed
A	0
B	5
C	40
D	62
E	157
F	286

Positive and negative plates from the above cycled modules were removed, washed so as to remove electrolyte, then dried and stored under inert conditions in order to prevent further changes prior to analysis. Active material samples were then carefully extracted from along the plate diagonal.

The analysis techniques employed for this study were BET Surface Area (BET SA), Particle Size Analysis, X-ray Diffraction (XRD) and Scanning Electron Microscopy (SEM).

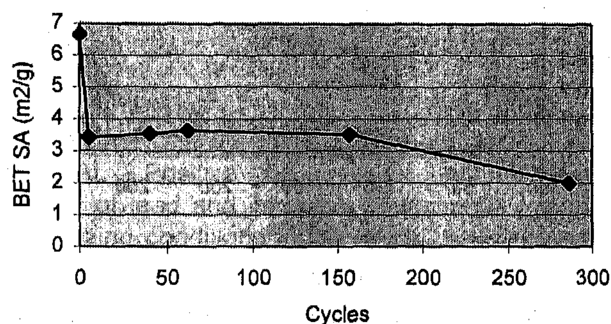
CYCLE LIFE CHANGES TO THE MORPHOLOGY AND CRYSTALLOGRAPHY OF THE POSITIVE ACTIVE MATERIAL - Table 7 summarizes the results for the positive active material (PAM).

Table 7. Morphological and Crystallographic Changes of the PAM During Life Cycling

Module	C/5 cycles	BET surface area (m^2/g)	Median pore diam (nm)	XRD % α-PbO_2	XRD % β-PbO_2	XRD % $PbSO_4$
A	0	6.66	24.9	12	81	7
B	5	3.42	35.8	8	88	4
C	40	3.55	28.8	2	96	1
D	62	3.64	49.1	1	97	2
E	157	3.52	4.9	0	100	0
F	286	2.00	4.2	1	99	0

As shown in Table 7, the first measured quantitative change to the PAM, throughout its cycle life, is the conversion from an initially high surface area crystalline material, immediately following formation, to a progressively lower surface area coralloid structure, as seen in Figure 2. (Reference I)

Figure 2. PAM BET SA vs Cycles



This transition may clearly be seen in the following SEM pictures, Figures 3 through 5.

Figure 3. Bulk PAM from Module A, 0 Cycles, x2500

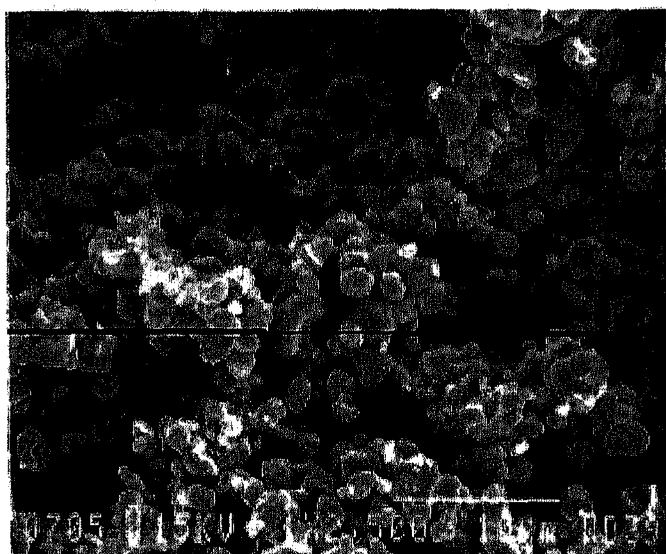


Figure 4. Bulk PAM from Module D, 62 Cycles, x2500

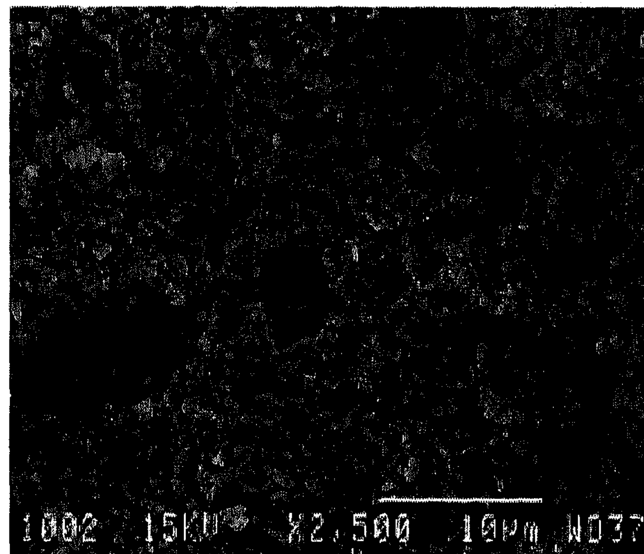
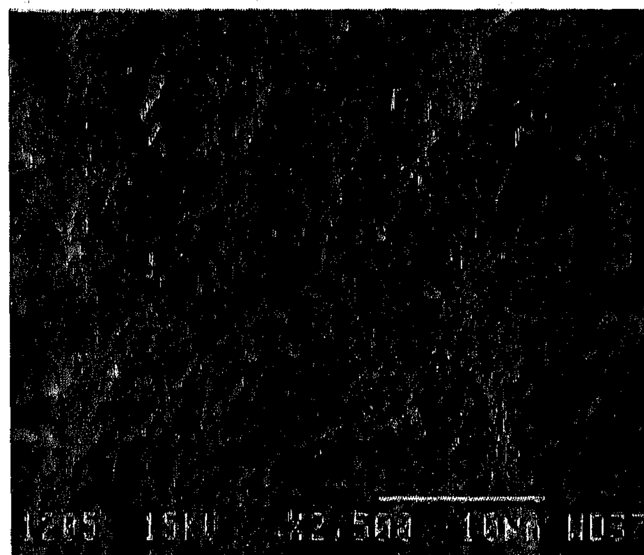


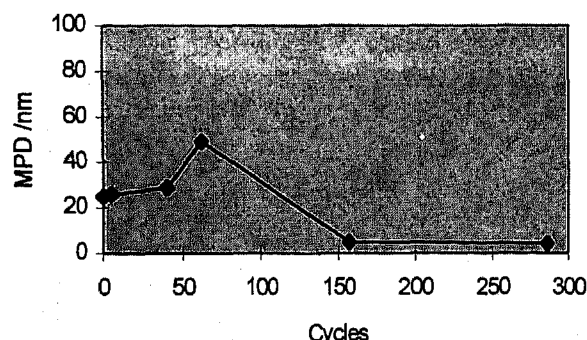
Figure 5. Bulk PAM from Module F, 286 Cycles, x2500



Healthy PAM is typically considered to possess a surface area (SA) of between 3-4 m^2/g. Toward end of life the contraction of the active mass results in a reduced SA approaching 2 m^2/g.

The second measured transition is the reduction in the median pore diameter (MPD) of the PAM, as shown in Figure 6.

Figure 6. PAM Median Pore Diameter

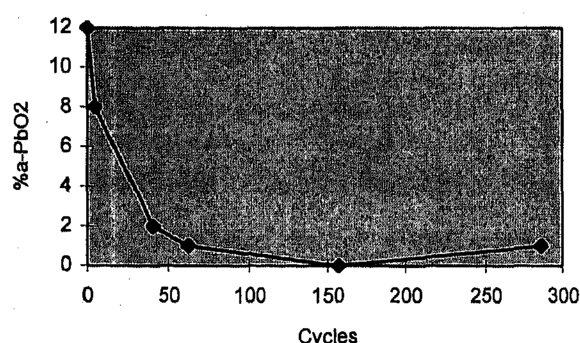


The median pore diameter referred to here is for PAM removed from the grid and gently broken-up, not left in-situ on the grid. It is thought that the in-situ PAM will have both very large pores left in the pellet, as the active mass consolidates, and small pores within the coralloid structure itself.

The MPD is a measurement of the degree of contraction of the PAM. In the above experiment Module E (157 cycles) exhibits a similar MPD to Module F (end-of-life) but still has 50% cycle life remaining. This anomalous behavior may be attributable to the high pack compression associated with the Genesis® technology. It is now recognized that high pack compression alleviates many of the shortfalls associated with cycling the positive plate by physically constraining the PAM. (Reference J)

The third measured quantitative change within the PAM is the transition of the α-PbO_2 component into β-PbO_2. This is shown in Figure 7.

Figure 7. Crystallography of PAM vs Cycles



This transition is generally a function of the battery design, specifically the degree of PAM utilization. For the Genesis® technology, this transition occurs early in life and, consequently, has no correlation with remaining useful cycle life.

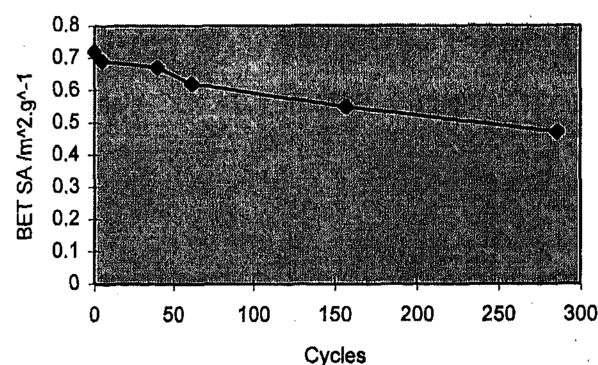
CYCLE LIFE CHANGES TO THE MORPHOLOGY AND CRYSTALLOGRAPHY OF THE NEGATIVE ACTIVE MATERIAL - Table 8 summarizes the results for the negative active material (NAM).

Table 8. Morphological and Crystallographic Changes of the NAM During Life Cycling

Module	**C/5 cycles**	**BET surface area (m^2/g)**	**MPD (nm)**	**XRD % Pb**	**XRD % PbO**	**XRD % $PbSO_4$**
A	0	0.72	4.8	99	0	1
B	5	0.69	4.7	99	1	0
C	40	0.67	4.5	99	0	1
D	62	0.62	4.2	99	0	1
E	157	0.55	4.5	96	3	0
F	286	0.47	4.9	97	2	1

The only measured change found to the NAM during cycling is a gradual reduction in surface area, possibly as a result of the negative expanders losing activity. This reduction is shown in Figure 8.

Figure 8. NAM BET SA vs Cycles



SEM analysis of the negative plates, at the beginning of life and after 286 cycles, reveals little visual change, as seen in Figures 9 and 10.

Figure 9. Bulk NAM from Module A, 0 Cycles, x2500

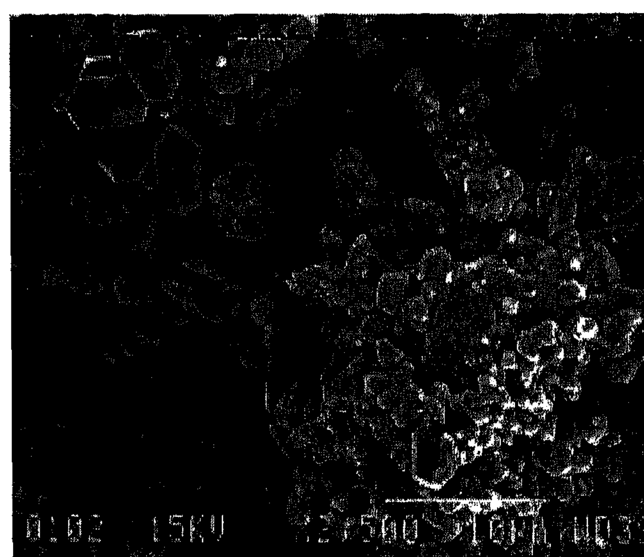
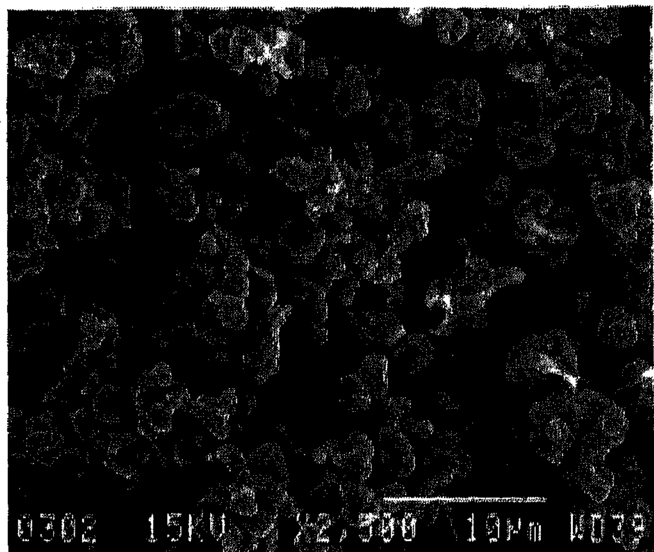


Figure 10.Bulk NAM from Module F, 286 Cycles,x2500

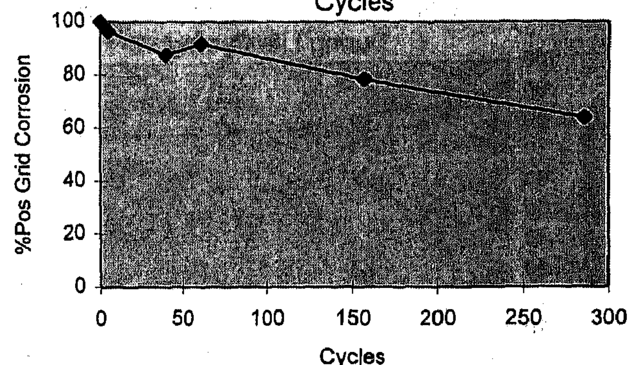


Although the negative electrode was never found to be capacity limiting, even at the end of life, there does appear to be a good correlation with BET SA and cycle number. Recent work indicates that the decline in negative surface area is alleviated by rapid recharging. (Reference I)

CORROSION OF THE POSITIVE GRID MATERIAL - Positive grid corrosion occurs due to the instability of lead and its alloys at the anodic potentials and oxidizing conditions found at the positive electrode. Corrosion is kinetically inhibited by the growth of a passivating layer of PbO_2 on the grid surface. However, corrosion will eventually limit the capacity of the positive electrode by a combination of growth, caused by volumetric changes of the corrosion product at the grid surface and at the grain boundaries, and reduction of electronic continuity throughout the grid. Furthermore, it aggravates the situation by consuming much-needed water from the electrolyte, which in an acid starved system can be a major contributor to end of useful life.

Positive grid corrosion was estimated by mechanically removing the PAM, chemically stripping any residual PAM and corrosion product from the grid, and then finally weighing the residual lead grid. The difference in weight for the sample grid and the grid from Module A (0 cycles) yielded the amount of lead oxidized during cycle life. The results are shown in Figure 11.

Figure 11. %Positive Grid Corrosion vs Cycles



For the Genesis® technology, under these cycling conditions, there is a very strong correlation between % grid corrosion and remaining cycle life.

ELECTROLYTE CHANGES - Water loss is an inevitable consequence of cycling batteries. Water is electrochemically removed from the system by: a) charging inefficiencies resulting in electrolytic decomposition; and b) corrosion of the positive grid consuming oxygen from the electrolyte, according to:



$$Pb + 2H_2O \longrightarrow PbO_2 + 4H^+ + 4e^-$$

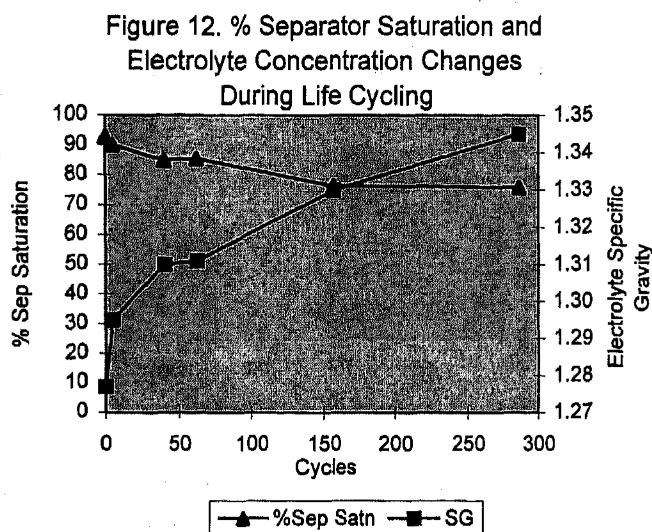
The result is a progressive increase in electrolyte concentration and a reduction in saturation level.

These changes were evaluated by: a) measuring the total module weight loss over cycle life, b) removing and weighing the wet glass mat separator from one cell of each of the modules, and c) extracting the acid from the separator and measuring its concentration. These results are summarized in Table 9.

Table 9. Electrolyte Changes During Life Cycling

	C/5 cycles	Module wt loss (g)	% wet sep wt comp to Module A	Electrolyte conc (% H_2SO_4)	Electrolyte SG @ 15°C (kg/dm^{-3})	Calc sep % satn
A	0	0	100	37.2	1.277	92.9
B	5	5.7	95	39.2	1.295	90.2
C	40	12.2	87	40.9	1.310	85.0
D	62	15.8	87	40.9	1.311	85.5
E	157	67.9	82	43.1	1.330	76.5
F	286	88.4	70	44.8	1.345	75.9

Figure 12 shows the variation in % separator saturation and electrolyte concentration with cycle life.



For Genesis® technology, a strong correlation between water consumption exists (measured as % separator saturation), and subsequently, electrolyte concentration and cycle life.

The results of the Phase III testing indicate that positive active material morphology or crystallography does not yield a good correlation to remaining cycle life. It is now recognized that high separator compression alleviates some of the problems associated with cycling the positive active mass. Therefore, high compression systems may continue to exhibit good cycling capability long after the PAM has transformed into a structure traditionally considered to be life limiting. In addition, positive grid corrosion and separator saturation (a measure of grid corrosion and dry out) exhibit a strong correlation to cycle life.

Therefore, based upon work completed to date, grid corrosion and/or separator saturation can be used to evaluate remaining life in Hawker Genesis® batteries. Additionally, measures to reduce grid corrosion/dry out will improve battery cycle life. This includes charge algorithms, particularly the amount of time the battery spends at high charging voltages.

CONCLUSIONS

Battery performance, in Phase I testing, indicates that Hawker Genesis® batteries can be damaged by low power charging. Both Phase I and Phase II testing results suggest that the utilization of a 3C fast charge rate, in vehicles not using battery management systems, provides good vehicle performance and extended battery life. In addition, charging of the Hawker Genesis® batteries at the 9C rate is possible, but a battery management system is required to reduce module failures. In Phase III, it was shown that grid corrosion and separator dry out may be used to predict remaining life in VRLA Hawker Genesis® batteries.

Through Phases I and II of the APS Charger Test Project substantial progress has been made in extending the life of VRLA batteries in fleet service and dramatically reducing the time required for charging. Additional work is required in Phase II testing to determine optimum charge protocols. However, the method of remaining battery life estimation pursued in Phase III is necessary to allow test work to be completed in a reasonable time period.

Future work will include continued Phase III evaluation of batteries cycled in Phase II to solidify the method of remaining life determination. Testing should then commence on batteries, charged between a 3C and 6C charge rate, using battery management systems to validate the preliminary results of Phase II. Testing should also continue using a 9C fast charge rate on batteries with battery management systems. In addition to this testing, other VRLA batteries should be tested to determine if similar cycle life predictive methods can be used.

ACKNOWLEDGMENTS

The authors would like to thank Sean O'Keefe and Scott Lichte, HEPi, for all sample preparation and testing.

REFERENCES

A. R.S.Hobbs: "Electric Vehicle Charger Test Project at Arizona Public Service - Final Report", March 1996.

B. E.M. Valeriote, J.K. Nor, and V.A. Ettel: *Proceedings of the 5th International Lead/Acid Battery Seminar,* International Lead Zinc Research Organization, Inc., Research Triangle Park, NC 1991, p. 93.

C. E.M. Valeriote and D.M. Jochim: Journal of Power Sources, 40 (1992) 93.

D. T.G. Chang, E.M. Valeriote and D.M. Jochim: *Proc. 26th International Symposium on Automotive Technology*, Aachen, Germany, September 13-17, 1993, pp. 405-413.

E. T.G. Chang, E.M. Valeriote and D.M. Jochim: Journal of Power Sources, 48 (1994) 163.

F. E.M. Valeriote, T.G. Chang and D.M. Jochim: "Fast Charging of Lead Acid Batteries", *9th Annual Battery Conference on Applications and Advances,* California State University, Long Beach, January 11, 1994.

G. E.M. Valeriote: "Rapid Charging of Electric Vehicle Batteries", *2nd Annual Conference of the Advanced Lead Acid Battery Consortium,* Sacramento, California, February 1996.

H. T.G. Chang and D.M. Jochim: Journal of Power Sources, *Rapid Partial Charging of Lead Acid Batteries,* 1996.

I. T.G. Chang and D.M. Jochim: *Proceedings of the 5th European Lead Battery Conference,* 1996.

J. A.F. Hollenkamp, et. al.: ALABC Project No. AMC-009, *Advanced Lead-Acid Batteries For Electric Vehicles:Strategies for Raising Positive Plate Performance*, June 1996.

ADDITIONAL SOURCES

A. Winsel, E. Voss, U. Hullmeine: Journal of Power Sources, *The Aggregate of Spheres ("Kugelhaufen") Model of the $PbO_2/PbSO_4$ Electrode,* 30 (1990) 209-226.

B. J.K. Nor and V. Pavlovic: "Ultra Rapid Battery Charging: Charge Time, Efficiency, and Battery Life Considerations", *EVS 13,* Oshsha Japan, 1996.

C. J.K. Nor: "Improvement of EV Battery Performance Resulting from Ultra Rapid Charging", *Zero Emission Vehicles BRNO 96 Conference*, Brno Czech Republic, June 1996.

DEFINITIONS, ACRONYMS, ABBREVIATIONS

AGM - Absorbed Glass Mat

ALABC - Advanced Lead Acid Battery Consortium

APS - Arizona Public Service

Battery management system - A system designed to manage the charge of each individual module.

C charge rate - The C rate of charging is defined to be the rated energy of a battery delivered over a one hour period. Fast charge rates are linear multiples of that one hour rating.

Comp - compared

Conc - concentration

Cond - conductive

Equalization Charge - Charging of the battery to 100% SOC.

Ind - inductive

Manuf - manufacturer

MPD - Median Pore Diameter

NAM - Negative Active Material

NiCad - Nickel Cadmium

NMH - Nickel Metal Hydride

PAM - Positive Active Material

Partial Charge - Charging of the battery to less than 100% SOC.

SA - surface area

Satn - saturation

SEM - Scanning Electron Microscopy

Sep - separator

SG - specific gravity

SOC - state of charge of the battery

UNK - unknown

VRLA - Valve Regulated Lead Acid

Wt - weight

XRD - X-ray diffraction

970253

Switchable Holograms for Automotive Applications

Bob L. Epling, Donna M. Brandelik, Lalgudi V. Natarajan, Richard L. Sutherland, Vincent P. Tondiglia, and Timothy J. Bunning
Science Applications International Corp.

ABSTRACT

This paper will discuss the relatively new technology of polymer dispersed liquid crystal (PDLC) based switchable holograms and its potential for application in the automotive industry. We have developed a novel photopolymer dispersed liquid crystal based material system, which allows fast, single-step recording of holograms with electrically controllable diffraction efficiency. The various characteristics of the holograms will be described as well as special considerations for various potential device configurations. Material characteristics to be described include transmission efficiency, diffraction efficiency, switching dynamic range, switching speed, recording sensitivity, temperature effects, voltage requirements, and system stability. Device applications to be considered include holographic optical elements, fiber switching, and displays.

INTRODUCTION

Diffractive and holographic optical elements (HOEs) offer the ability to manipulate light in ways not available to conventional reflective or refractive optics. Holography is being applied to an ever increasing range of applications which include displays, holographic interferometry, precision measurements, optical information processing, HOEs, optical switching, polarization control, narrowband notch filters, optical interconnects and couplers, lithography, and data storage.

We have been developing a holographic material system which allows for the single-step recording of

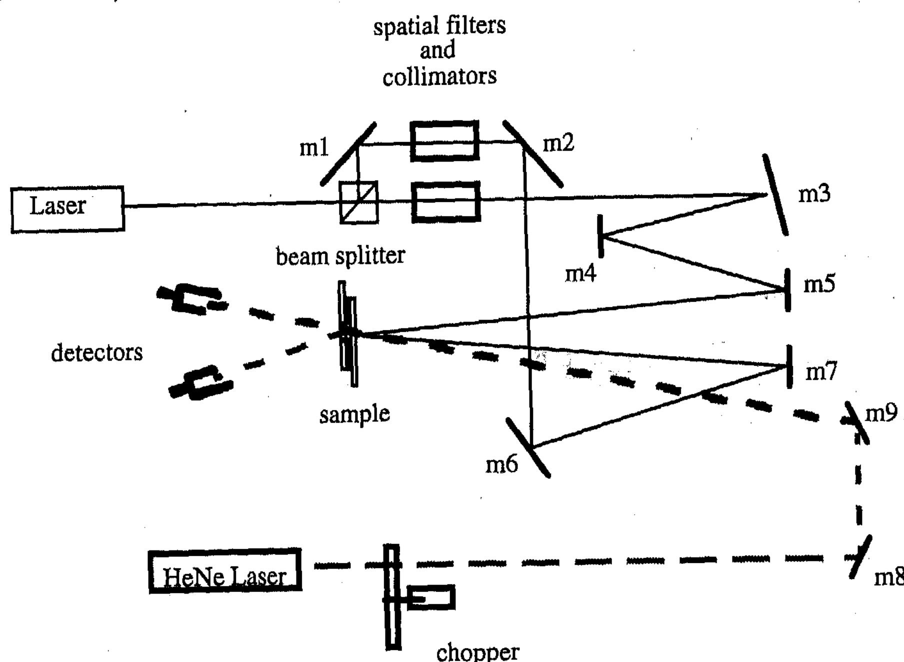


Figure 1 Experimental setup used to write PDLC gratings.

electrically controllable holograms.

Recording - Unlike most holographic recording processes, our PDLC based system allows for a single-step process. The syrup is placed between two indium tin oxide (ITO) coated slides which are separated with 15 - 100 μm spacers. Simple transmission gratings are recorded using the experimental configuration illustrated in Figure 1. In this configuration, two laser beams are crossed at the recording plane. In the regions of constructive interference, the light polymerizes the monomer. A diffusion of the liquid crystal to the regions of destructive interference occurs forming planes (fringes) of microdroplets. An SEM image of these microdroplet fringe plane is illustrated in Figure 2.

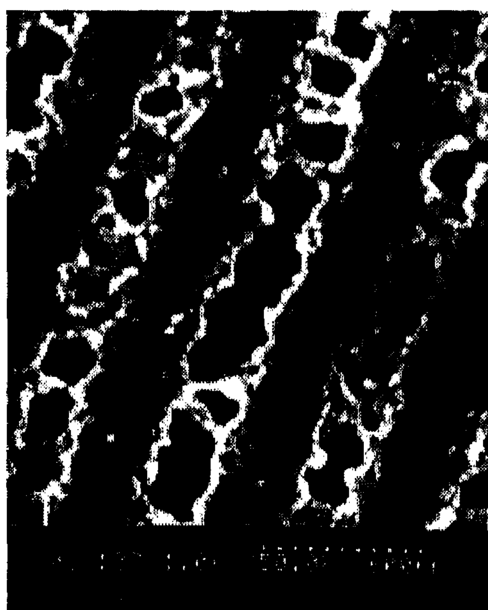
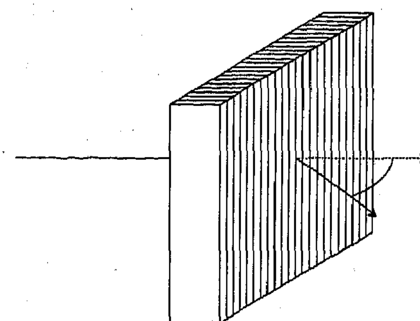


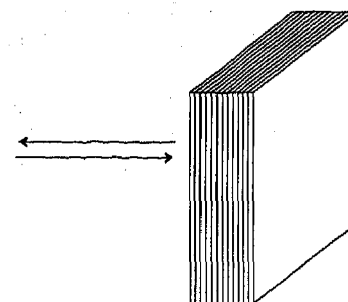
Figure 2 SEM surface view of PDLC grating.

Depending on the recording geometry, transmission gratings or reflection grating can be formed. In the case of transmission gratings, the fringe planes form perpendicular to the recording plane as illustrated in Figure 3. In reflection gratings, the fringe planes form parallel to the recording plane.

Material System - The material system is composed of a acrylate based photopolymer containing a nematic liquid crystal dispersion and has been described previously.[1,2,3,4] As the syrup is exposed, the holograms are photochemically induced by the writing beams in the regions of constructive interference. Polymerization results in a change in the chemical potential of the system and decreases the miscibility of the liquid crystal in the host. The liquid crystal phase separates from the host forming microdroplets of pure liquid crystal. The droplets can range 10 nm to 10 μm and form along the destructive interference fringe planes.



Transmission Grating



Reflection Grating

Figure 3 Fringe plane orientation for transmission and reflection gratings.

MAIN SECTION

SYSTEM PERFORMANCE - Using our PDLC material system, a broad range of performance is achieved. Many tradeoffs can be performed in the material system to optimize a particular performance parameter. The performance characteristics discussed in this paper have been characterized at 633 nm and room temperature, unless otherwise noted. This wavelength is acceptable for the SAE Red J578 (tail lights).[5]

Diffraction Efficiency - The diffraction efficiency for the switchable holograms is defined as the percentage of light diffracted into the Bragg angle relative to the incident beam. Diffraction efficiencies of 99% with narrow angular selectivity (1°) have been achieved. A goal of 40 dB is felt to be achievable. This would correspond to switching between 98% and 0.01% diffraction efficiency.

Dynamic Range - A dynamic range of >35 dB is currently achievable using the current materials and processing techniques. This corresponds to switching between 80% and 0.02% diffraction efficiency. Extending the dynamic range to 40 dB is a primary focus of our current research.

Switching Speed - The switching response times of the holograms have been measured and agree with simple models. Switching "ON" times were found to be approximately 25 μs where the "OFF" times were approximately 44 μs. It may be possible to further reduce the "ON-OFF" times by using a spiked voltage driving scheme.

Recording Wavelengths/Sensitivity - Switchable holograms can currently be recorded using wavelengths in the 400 - 800 nm range. We are currently exploring alternate photoinitiators to extend this sensitivity to the 900nm, 1.3 μm, and 1.55 μm range. Laser intensities of 1 - 3 mW/cm^2 are typically used to record the grating with exposure times of 20 seconds to 2 minutes being typical.

Driving Voltages - The holograms are typically driven using a 2 kHz square wave of approximately 5 V/μm. This corresponds to approximately 80 V_{p-p} for a 15 μm thick hologram. Figure 4 illustrates the diffraction efficiency as a function of driving voltage.

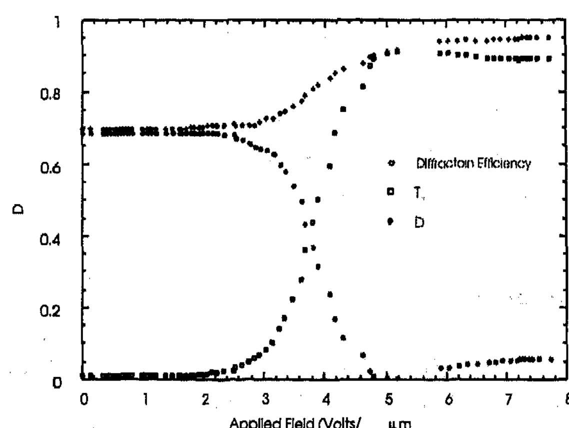


Figure 4 Electrical switching of a transmission grating.

Temperature Range - The currently used liquid crystal (E7) has a nematic range of -10 to 60.5 °C. A broad range liquid crystals are available and broader ranges of operation can be achieved e.g. NP 1291 TNC has a nematic range of -10 to 107°C.[6] As temperatures approach the upper end of the nematic range, the ordinary index of refraction increases while the extraordinary index decreases. For temperatures above the nematic range, the liquid crystal become isotropic and cannot be switched to a clear state. This temperature dependence can have a significant impact on the performance of the hologram. Figure 5 demonstrates the decrease in diffraction efficiency with increasing temperatures. The largest impact of temperature will ,however, be on the switching dynamic range. The decrease in switching dynamic range with temperature will be presented.

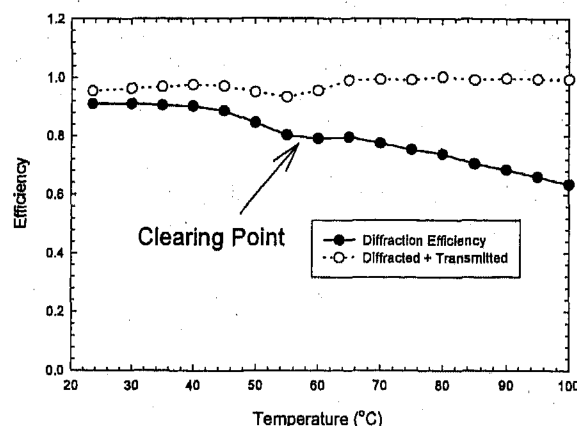


Figure 5 Diffraction efficiency of a PDLC hologram as a function of temperature for E7 liquid crystal.

DEVICE CONFIGURATIONS

Holographic Optical Elements - Switchable holographic lenses or general holographic optical elements can be produced using our material system. Figure 6 illustrates a holographic lens in the "ON" and "OFF" states as well as the chromatic dispersion anticipated with a diffractive lens. These lenses have been produced and can also be stacked to produce a switchable focus lens.[7] A significant advantage of our material system, is that the holograms are recorded in-situ in a single-step process. This feature can alleviate alignment errors of the HOEs.

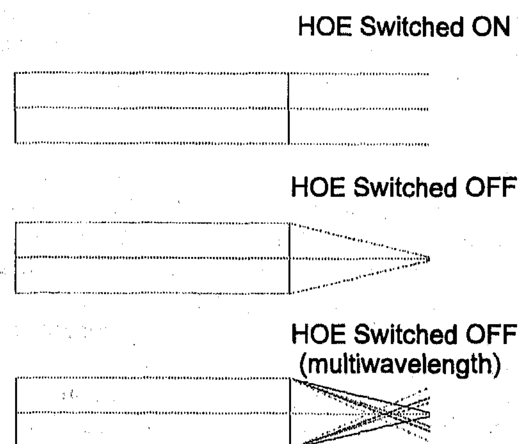


Figure 6 Illustration of switchable holographic lens.

Fiber Optic Switching - Our material system has been used by Foster-Miller to produce a 2x2 single-mode 1.3 μm fiber optic switch. The switch was capable of 25 dB switching.

Holographic Reflection Filter/Displays - Switchable reflection holograms can be formed using our PDLC material. Figure 7 illustrates the voltage modulation of a narrowband reflection filter. Switchable reflection hologram technology is less mature and current work is focused on improving the switching dynamic range. Switchable reflection grating have been explored for application in high contrast reflective color displays.[8] In this application, three switchable filters (RGB) are stacked to achieve a full color display.

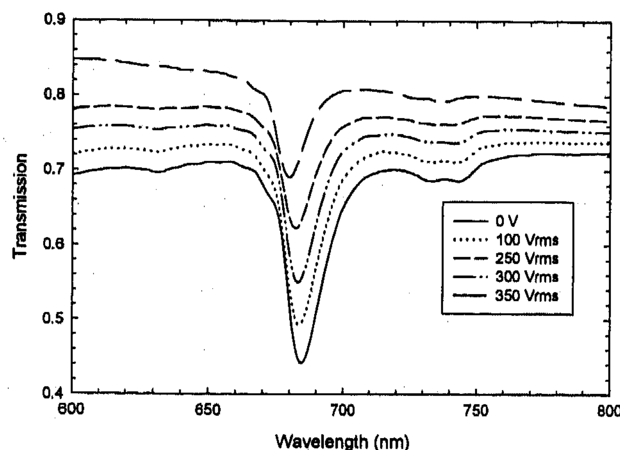


Figure 7 Voltage modulated reflection grating.

CONCLUSIONS

In this paper, we have discussed the current status of polymer dispersed liquid crystal based switchable hologram technology. Various material performance characteristics were summarized however it should be noted that the materials system can be modified to optimize the performance the various parameters. Various device configurations were also presented to illustrate the broad utility of this material system.

ACKNOWLEDGMENTS

This work was supported by Science Applications International Corporation and by the US Air Force through Contract F33615-95-C-5423.

REFERENCES

1. R. L. Sutherland, L.V. Natarajan, V.P. Tondiglia, T. J Bunning, Chem. Mater., **5** (10), pp.1533-1538, 1993.
2. R. L. Sutherland, V.P. Tondiglia, L.V. Natarajan, T. J Bunning, W.W. Adams, Appl. Phys. Lett., **64** (9), pp. 1074-1076, 1994.
3. R. L. Sutherland, L.V. Natarajan, V.P. Tondiglia, T. J Bunning, W.W. Adams, Proc. SPIE, Vol. 2404, pp. 132-143, 1995.
4. L.V. Natarajan, R. L. Sutherland, V.P. Tondiglia, T. J Bunning, W.W. Adams, J. Nonlin. Opt. Phys. Mater., **5** (1), 1996.
5. George Hulse, Personal discussion., Nov. 1996.
6. Handbook of Laser Science and Technology, Part 2, pp. 409-465, CRC Press, Inc., Boca Raton, 1986.
7. L.H. Domash, Ychen, B. Gomatam, C. Gozewski, R. Sutherland, L. Natarajan, V. Tondiglia, T. Bunning, W. Adams, Proc. SPIE, Vol. 2689, pp. 188-194, 1996.
8. G. Crawford, T. Fiske, L. Silverstein, Proc. SID Digest, Vol. 27, pp. 99-102, 1996.

970254

Comparison of Dual Focus Collector Schemes for Fiber Systems

W. J. Cassarly
Optical Research Associates

T. L. Davenport and R. L. Hansler
Lighting Innovations Institute

G. R. Allen
GE Lighting

ABSTRACT

The development of efficient, multi-output High Intensity Discharge (HID) light sources is important for driving fiber systems in automotive applications, in particular for headlamps. Such systems must collect a large fraction of the light from the HID source while efficiently coupling the light into large core light pipes.

An earlier paper [1] showed how multifocus optical systems can efficiently split the light evenly between separate output ports. Multifocus systems were shown to be more efficient than a mixing rod approach, especially for the case of two ports. In the case of dual output systems, many more cases need to be considered. To facilitate the investigation, a state-of-the-art optical analysis/design program was employed to demonstrate the tradeoffs between six different dual focus geometries. A novel dual ellipse approach is presented.

BACKGROUND

The technology required for automotive headlamp distributed lighting systems is approaching commercial viability. The instant start metal halide light sources [2, 3] required for automotive headlamps are now commercially available [4], and have been used to illustrate that single source headlamp systems are viable [5, 6]. Low cost fibers which can handle the automotive underhood environment are currently under development and fibers for less thermally demanding environments are readily available today. For headlamp systems, as well as other distributed lighting systems, a typical geometry is to use half the light for the driver's side headlamp and the other half for the passenger's side headlamp.

From a system perspective, an automotive fiber system needs to efficiently deliver the light to the output optics on either side of the automobile while minimizing the size of the fibers that are used to transport the light. This means that it is desirable to collect as much light as possible, focus the light into as small an area as possible, and maximize the transmission of the light through the fibers. High collection efficiency allows lower wattage bulbs to be used which minimizes the lamp package size, ballast cost, and thermal impact. Tight focusing of the light distribution means that the size of the fibers can be minimized. Small fibers are desirable to make the routing of the fibers easier and to lower fiber costs. Unfortunately, as the light is focused into a smaller area, the angular distribution gets wider. A wider angular distribution of light, in general, results in poorer transmission through the fiber. Minimizing the size of the light pipe and the extent of the angular distribution is the key to minimizing the size of the output optics.

OVERVIEW

This paper considers some of the tradeoffs in methods used to split the light evenly between two round fibers. To perform this splitting efficiently while focusing the light into a small area and minimizing the angular extent of the light, sources with high luminance and high efficacy are required. Fluorescent lamps provide high efficacy but possess low luminance. Xenon lamps provide high luminance but at only relatively poor efficacy. Metal Halide lamps with short-length arc gaps provide the desired balance between high luminance and high efficacy.

Modeling the light emanating from a discharge sources is complicated by the fact that what the source looks like from different view angles is determined by details such as the electrode geometry, the choice of fill

chemicals, the amount of fill chemicals, the wall temperature, the bulb wall shape, the burning orientation of the lamp, and the power applied to the lamp. Because of all these variables, and there are more, each discharge source will be unique when the luminance Vs view angle is evaluated. Since the intent of this paper is to compare different collector schemes and not perform an exhaustive comparison of the optical modeling of different types of discharge sources, a simple spherical volume source model is considered. This source averages the luminance of the core and the mantel regions of the plasma. The source radiates isotropically from within the sphere and is transparent to its own radiation. Much more complicated models can be developed, but they are outside the scope of this paper.

OPTICAL GEOMETRIES

Six dual focus collector geometries and one single focus geometry are considered using the same light source model. The light source emitting region is a 2.5 mm sphere that is centered inside of a quartz bulb wall that is modeled using an 8 mm outer diameter sphere with 1 mm thick walls. The legs of the lamp are 3.5 mm diameter cylinders that are nominally 20 mm long and are unioned with the spherical bulb wall. The ends of the lamp legs are made absorptive to remove light which might totally internally reflect (TIR) down the length of the legs. The diameter of the absorptive electrodes is 1 mm, and the gap between the two electrodes is 2.5 mm.

Figure 1 shows the seven different collection geometries that were modeled using LightTools®. The names and the details that will be used throughout the rest of the paper are described in the following paragraphs. A reflector reflectivity of 95% and coated glass surfaces with 95% transmission for two surfaces are used.

FIGURE 1A. STANDARD ELLIPSE - The lamp is placed at the first focus of an elliptical reflector with a glass cover plate attached near the rim of the reflector. The angle from the first focus to the rim of the reflector is 35 degrees. The distance from the first focus to the apex of the reflector is 10 mm. This single port collection system is used to provide a basis for comparison to the various dual focus systems.

FIGURE 1B. SPLIT ELLIPSE - The Standard Ellipse of Figure 1A is split in two, with the two halves rotated slightly at the first focal point about an axis 90° out of the paper. The second focus of each half is located 6 mm from the center line. A cover glass is added.

FIGURE 1C. V-MIRROR - The Standard Ellipse is used; however, in this configuration the light is split prior to the second focal point using two mirrors. The angle between the two mirrors is adjusted so that the centroid of the light distribution that is reflected off of the V-mirror is 90 degrees from the axis of the lamp. No cover glass is included in this system so that all geometries considered have the same average surface losses; however, thermal considerations would probably require the inclusion of a cover glass in a practical implementation.

FIGURE 1D. PARABOLA WITH LENSES - The lamp is placed at the focal point of a 10 mm focal length parabola. The angle from the first focus to the rim of the reflector is 35 degrees. The collimated light is then split with two off-center aspheric lenses. The center of each lens is adjusted so that the centroid of the focused light distribution is parallel to the axis of the lamp.

FIGURE 1E. BACK TO BACK ELLIPSES - Two of the 'Standard Ellipses' are cut along the plane that passes through the first focus of the reflector and joined back to back.

FIGURE 1F. DUAL ELLIPSES WITH TAPERED RODS - Two ellipses are truncated along the plane that passes through the first focus of the reflector. The opposite sections of the reflectors that are used in Figure 1E are selected. The reflector shape is also changed to make the reflectors more spherical. The hole near the apex of each of the two reflectors is large enough to ensure that the light which propagates from the first focus to the rim of the reflectors is not blocked during its propagation from the rim to the second focus. The angles with this system are reduced through the use of a tapered rod. The diameter of the output end of the rod is 1.3 times larger than the input diameter. The tapered rod that is pictured in Figure 1F has an input radius of 2.3mm and a length of 30mm.

FIGURE 1G. DUAL ELLIPSES WITH LENSES - The system shown in Figure 1F is used, except that the holes in the reflector are replaced with negative lenses. The negative lenses reduce the angular distribution of the light, performing the same function as the tapered rods.

ILLUMINATION MODELING

Given the geometries described in the previous section, a Monte Carlo ray trace approach was used to evaluate the spatial and angular distribution of the light for each case. The starting positions of equally weighted individual rays are randomly chosen within the 2.5mm sphere and then the direction of eacht ray is randomly chosen using an isotropic distribution. The ray is traced through the system, accurately modeling all refractive and reflective effects. The bundles of rays that are focused at each of the two output ports are analyzed to determine the location of their least-means-squared focus.

The illuminance distribution at the plane in space where the least-means-squared focus occurs is then

computed by averaging the rays in a grid of square collection bins. Smoothed raster plots of this binned data are created. The flux within circular apertures of monotonically increasing size are computed using the raw ray data, not the binned data. This flux versus aperture data is called an illuminance encircled energy distribution curve.

The angular distributions of the rays that pass through a 12 mm port are then used to compute the intensity distributions of the collected light. Because the single port distribution is not split into two, a 18 mm port was used to compute the intensity distribution for the 'Single Ellipse' system.

The intensity distributions are computed by averaging the rays in a polar grid of equal width elevation angles. Each annulus of elevation angle is then divided into bins of equal azimuth width. The bins look like arcs within a pie wedge. The flux within circular apertures of monotonically increasing elevation angles is summed using the raw ray data. This flux versus angle data is called an intensity encircled energy distribution curve.

ILLUMINATION MODELING RESULTS

The illuminance (lumen/mm^2) distributions near the input face of the lightpipes for each of the dual port systems are shown in Figure 2. For the 'Dual Ellipses with Tapered Rods' case the lightpipe input face is the large end of the tapered rod (Fig. 2E). Because all systems possess symmetry, only one of the two output ports is shown. The greyscale levels are chosen such that the peak illuminance in each plot is dark and the area of zero illuminance is white. The size of the fiber required to collect all the light for each of the systems can be qualitatively evaluated by looking at the size of the illuminance distributions. The 'Dual Ellipses with Tapered Rods' and the 'Dual Ellipses with Lenses' show the smallest spatial distributions.

The focusing can be quantified by examining the illuminance encircled flux distribution curves shown in Figure 3. The ordinate represents the diameter of fibers placed at the least-means-squared focus. The abscissa represents the flux collected at the fiber input face divided by the flux that passes out of the lamp. Perfect collection with no transmission losses would be a value of 0.5, since the light is divided between two ports. Because of reflectivity and transmission losses, the maximum is 0.5*0.95*0.95 = 0.45. The port size for the 'Dual Ellipses with Tapered Rods' is the size of the output end of the taper, which is 1.3 times the input size; however, the size of the input end is used to determine how much flux is collected. The illuminance encircled flux distribution results show that the 'Dual Ellipses with Tapered Rods' and 'Dual Ellipses with Lenses' systems possess the best collection efficiency of all the systems modeled. The primary reason is that these two geometries possess reflectors which nearly surround the lamp.

The illuminance distributions for the 'Standard Ellipse' and both ports of the 'Dual Ellipses with Lenses' are shown in Figure 4. Figure 5 shows the illuminance encircled energy plots for the 'Standard Ellipse' and both of ports from the 'Dual Ellipses with Lenses' case--where the outputs from both ports have been combined ideally into an 'effective' single round fiber. The diameter corresponds to the area of the combination of the two ports that are used. For example, if the relative flux for one of the 'Dual Ellipses with Lenses' ports is 0.35 at a diameter of 6.33 mm, then the combination is 0.70 relative flux at $6.33\sqrt{2}$ =8.95 mm.

The intensity distributions (candela vs. angle) for the light collected within a 12 mm aperture for each of the dual port systems are shown in Figure 6. The intensity distribution for the 'Single Ellipse' case is shown in Figure 7. Essentially, the raster plots represent the light distribution as it would appear to an observer looking at the internal surface of a large sphere. The light source is located at the center of the sphere and the observer is far from the sphere. Because the angles are relatively small for all these cases, the dual port intensity distributions are similar to what would be observed if the light that would be captured by a 12 mm diameter fiber were allowed to illuminate a wall on the far side of a room. The 'Split Ellipse', 'V-Mirror', and 'Parabola with Lenses' all show a somewhat half-circular shape because of the splitting geometries. The other cases all show a dip in the center of the distributions. The dip occurs because of the hole near the apex of the reflectors and/or because of bulb blockage. In the 'Dual Ellipses with Tapered Rods' and 'Dual Ellipses with Lenses' cases, some blockage because of the legs of the lamps occurs. This leg blockage depends upon the relative size of the reflectors and the lamp.

As light propagates within a large core optical fiber, the energy within a small annular range of angles will be mixed within that annulus. The intensity distribution becomes rotationally symmetric. Some mixing between small and large angles occurs, but that is determined by effects such as bends in the light pipe and scattering. For straight light pipes with negligible scattering effects, annular structure will be lost at the output end of a light pipe, but the average radial structure will be retained. As shown in Figure 8, by the time light from the focus of the 'Dual Ellipse with Lenses' has traveled through 60 mm of fiber, the intensity distribution is nearly identical to that of the 'Dual Ellipse with Tapered Rods' case.

The intensity encircled energy plots as a function of angle are shown in Figure 9. Since the collection efficiency information is contained in Figure 3, these curves were all normalized by their value at large angles. The curves show that all designs provide greater than 90% of the light within a 33 degree half angle (66 degree cone). These angular distributions are low enough to have good transmission through 0.55 numerical aperture (NA) fibers. The critical angle is arcsin(NA), which is 33.4° for a 0.55 NA fiber.

A very simple model of a short length of high transmission fiber is that it propagates light that is at angles lower than the fiber critical angle (Transmission = 1) and leaks light at angles higher than the critical angle (Transmission = 0). This is shown schematically in Figure 10. This model is an over simplification. Bends in the fiber, propagation of skew rays, and the tightness of the cladding will all cause the effective transmission curve as a function of angle to be less than unity at angles below the critical angle and greater than 0 at some angles above the critical angle. Even so, the binary transmission model provides reasonable performance estimates.

A direct result of the binary fiber transmission model is that the fraction of flux transmitted from one end of the fiber to the other end is the same as the fraction of flux within the critical angle of the fiber. Thus, Figure 9 can be used to determine what fraction of the flux coupled into the fiber propagates from the input end to the output end.

COMPARISON USING ANGLE TO AREA CORRECTION TO PROVIDE EQUIVALENT ANGULAR DISTRIBUTIONS

A nonimaging angle to area converter, shown schematically in Figure 11, can be used to increase/decrease the size of the focused light by decreasing/increasing the angular distribution. This is the technique used in the 'Dual Ellipse with Tapered Rods' case and represents a simple means to match the illuminator NA with the fiber NA. This type of angle-to-area conversion is described by **[7]**.

$$D_1 \mathrm{Sin}(\theta_1) = D_2 \mathrm{Sin}(\theta_2), \qquad \text{EQ. 1}$$

where D_1 and θ_1 are the aperture size and angle at the input to the nonimaging converter; D_2 and θ_2 are the aperture size and angle at the output of the nonimaging converter. An ideal lens-based magnification system is also governed by Equation 1, where D_1/D_2 is the magnification ratio. For example, assume that a system transmits 90% of the light within 20 degrees using a 15 mm fiber. If 90% of the light could be transmitted within 30 degrees, then the equivalent fiber size is 15*sin(20)/sin(30) = 10 mm.

One way to compare the potential performance of fiber optic illuminators that have different angular distributions is to select a percentage of the collected flux and use a nonimaging optical element to adjust the angle within which that flux is contained. Assuming the binary fiber transmission model, the fiber transmission is then independent of the details of the shape of the angular distribution. Of course, if the angular distribution is magnified then the spatial distribution must be demagnified using Equation 1. This scaled spatial distribution plot can then be used to compare the relative amount of flux collected by the fiber. Since the transmission through the fiber is the same for all cases, the scaled spatial distribution plots provide a relative comparison of how much light is delivered to the other end of the fiber.

This scaling approach is used to adjust all the curves shown in Figures 3 and 9 so that 90% of the light is collected within a 30 degree half angle. The scaled illuminance encircled energy results and scaled intensity encircled energy results are shown in Figures 12 and 13, respectively. The graphs show that the 'Dual Ellipses with Tapered Rods' and 'Dual Ellipses with Lenses' cases provide the highest amount of flux in the smallest fibers.

RECOMMENDATIONS FOR FURTHER WORK

If delivery of only a small fraction of the flux were needed for a particular application, then the scaling might be performed using a smaller percentage of the angular distribution. Since we are interested in high efficiency systems, those cases are considered outside the scope of this paper; however, the scaling concept introduced in this section could be used to compare such systems.

It would be valuable to extend the results presented here to include a more detailed fiber transmission model. The angle-to-area scaling approach could then be used to optimize the transmitted light as a function of all relevant fiber and illuminator variables.

CONCLUSIONS

Since the headlamps of automobiles are required to be on either side of the automobile, the natural configuration is one whereby the focusing geometry provides one port on the right and one port on the left. The fiber can be bent to provide this same type of geometry, but the space for the bend tends to require a larger total package size. The 'Ellipse with V-Mirror', 'Dual Ellipses with Tapered Rods', and the 'Dual Ellipses with Lenses' configurations all provide this desired geometry. The 'Dual Ellipses with Lenses' configuration maximizes the light collection into small fibers and also minimizes the required package space.

The analysis of dual focus systems is a complex problem. Computer modeling has been shown to be an effective tool to aid the design of this type of complex optical system. Especially important is the combination of pictures which show important ideas and graphs which quantify the tradeoffs.

ACKNOWLEDGEMENTS

The authors wish to thank John Davenport for his comments which improved the content of this paper. We also thank Mary Sue Kaliszewski for her comments which improved the readability of this paper and of previous SAE papers.

REFERENCES

[1] W.J. Cassarly, J.M. Davenport and R.L. Hansler, 'Uniform Light Delivery Systems,' SAE, Paper No. 960490.

[2] J.M. Davenport and R.L. Hansler, 'Discharge for Automotive Lighting,' SAE Congress, Mar 1988.

[3] J.M. Davenport, R.L. Hansler, and R.H. Springer, 'Starting the Instant Light Metal Halide Lamp', IESNA Annual Conference Technical Papers, pp 457-472, 1989.

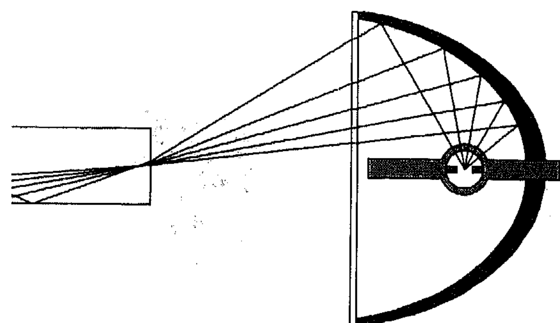
[4] For example: K. Jost, 'High Intensity Discharge Headlamps,' Automotive Engineering, pp. 38-42, Nov. 1995.

[5] H Tiesler-Wittig and J. T. Bindels, 'Discharge Light Source - The basis for fibre optics automotive lighting,' SAE, Feb. 1995.

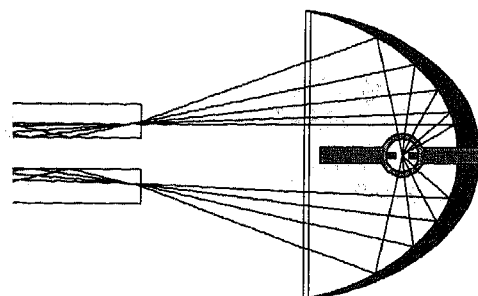
[6] W.J. Cassarly, J.M. Davenport, and R.L. Hansler, 'Distributed Lighting Systems: Uniform Light Delivery,' SAE Paper 950904, Feb. 1995.

[7] For example: W.T. Welford and R. Winston, 'High Collection Nonimaging Optics,' Academic Press, Inc. San Diego, 1989..

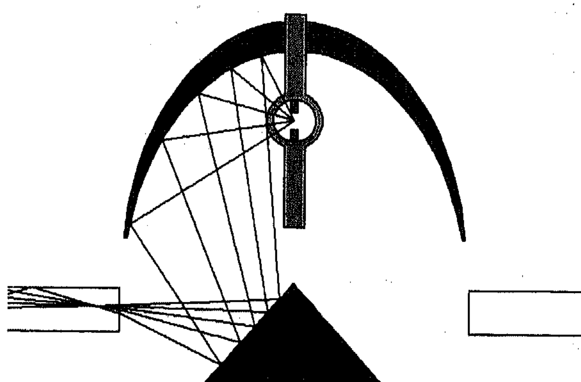
LightTools is a registered trademark of Optical Research Associates.



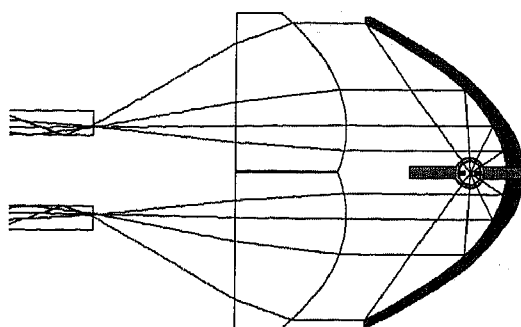
(A) Single Ellipse



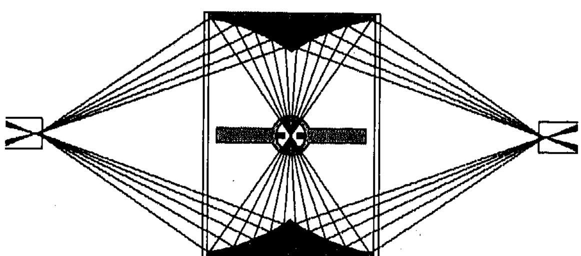
(B) Split Ellipse



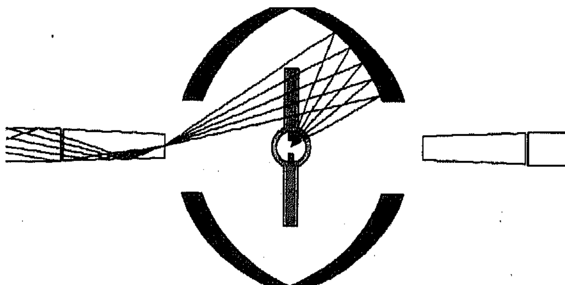
(C) V-Mirror



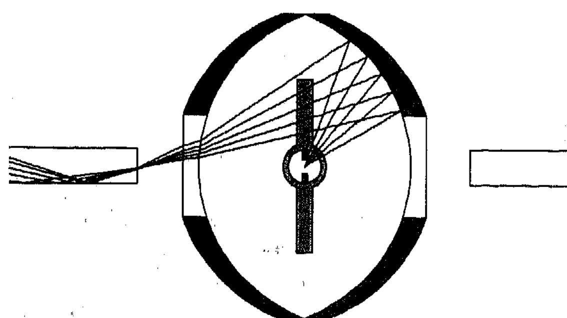
(D) Parabola with Lens



(E) Back to Back Ellipses



(F) Dual Ellipse with Tapers



(G) Dual Ellipse with Lenses

Figure 1. Collection Geometries

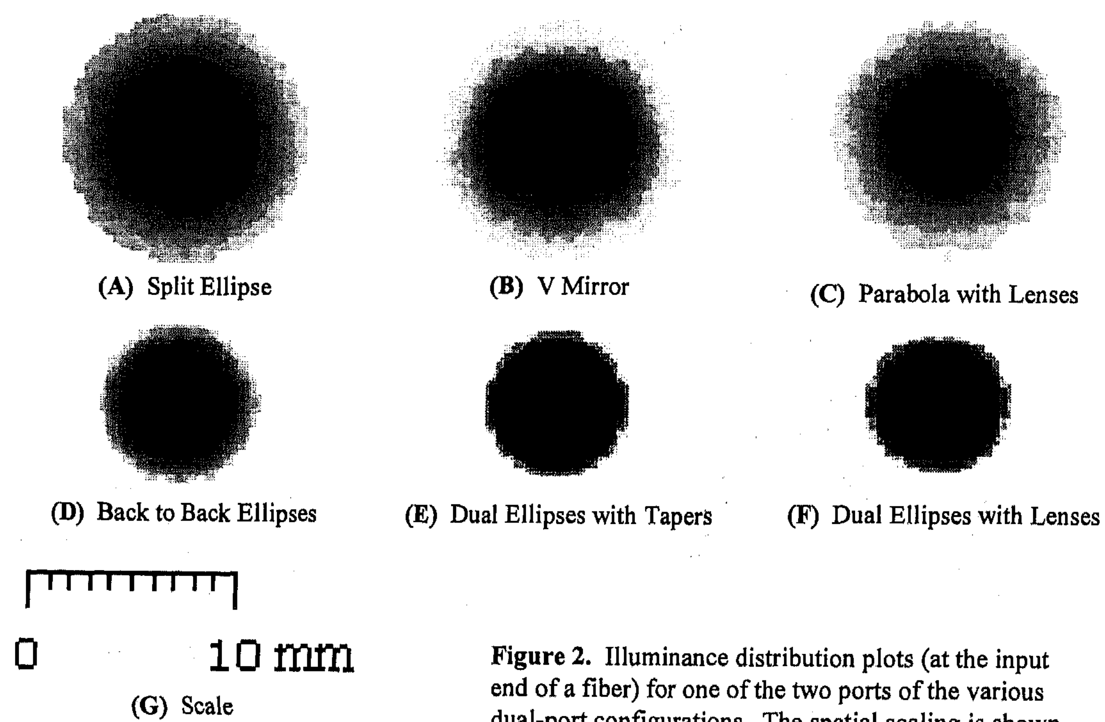


Figure 2. Illuminance distribution plots (at the input end of a fiber) for one of the two ports of the various dual-port configurations. The spatial scaling is shown in (G).

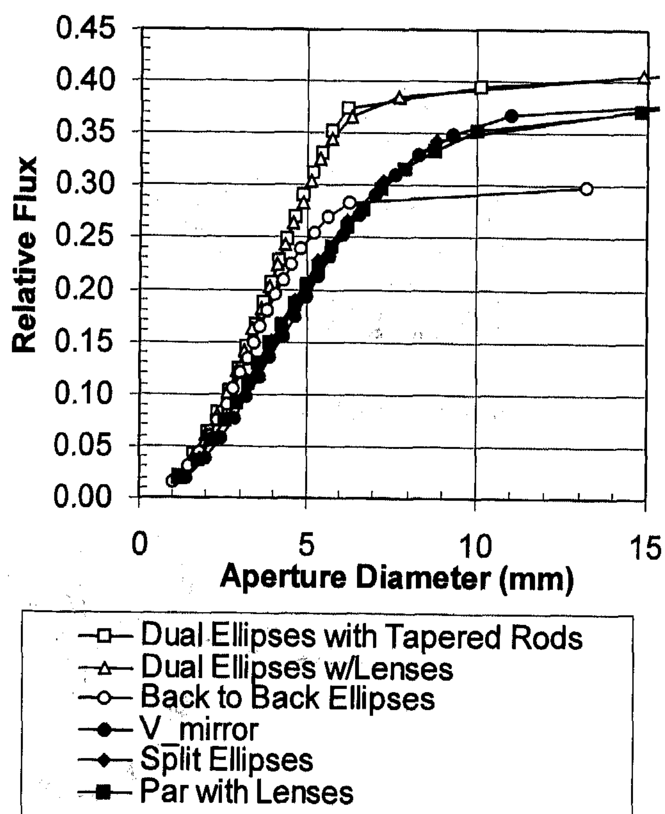


Figure 3. Illuminance encircled energy distribution curves (lumens collected by an aperture) for one of the two ports in each of the dual port configurations.

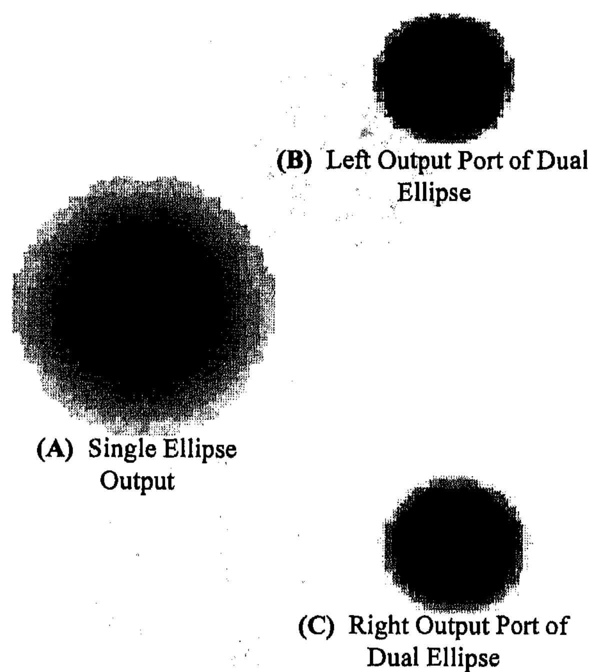


Figure 4. Illuminance distribution plots for the single ellipse and dual ellipse with lenses cases. (Same as above.)

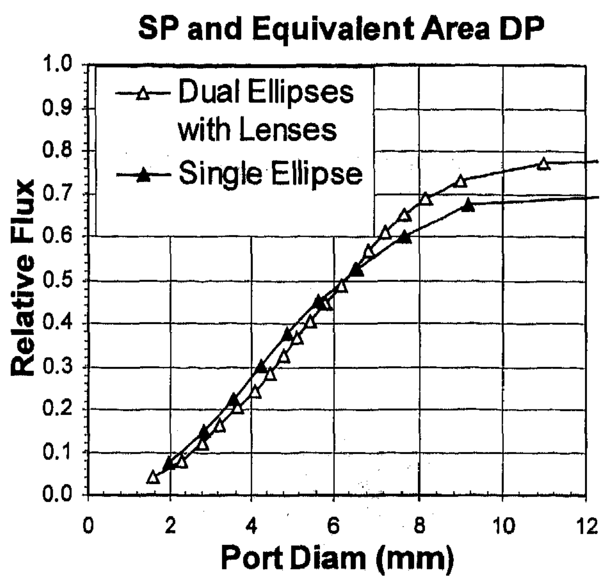


Figure 5. Single ellipse illuminance encircled energy distribution curve compared to the dual ellipses with lenses case with equivalent areas. The increase in collection efficiency can be seen.

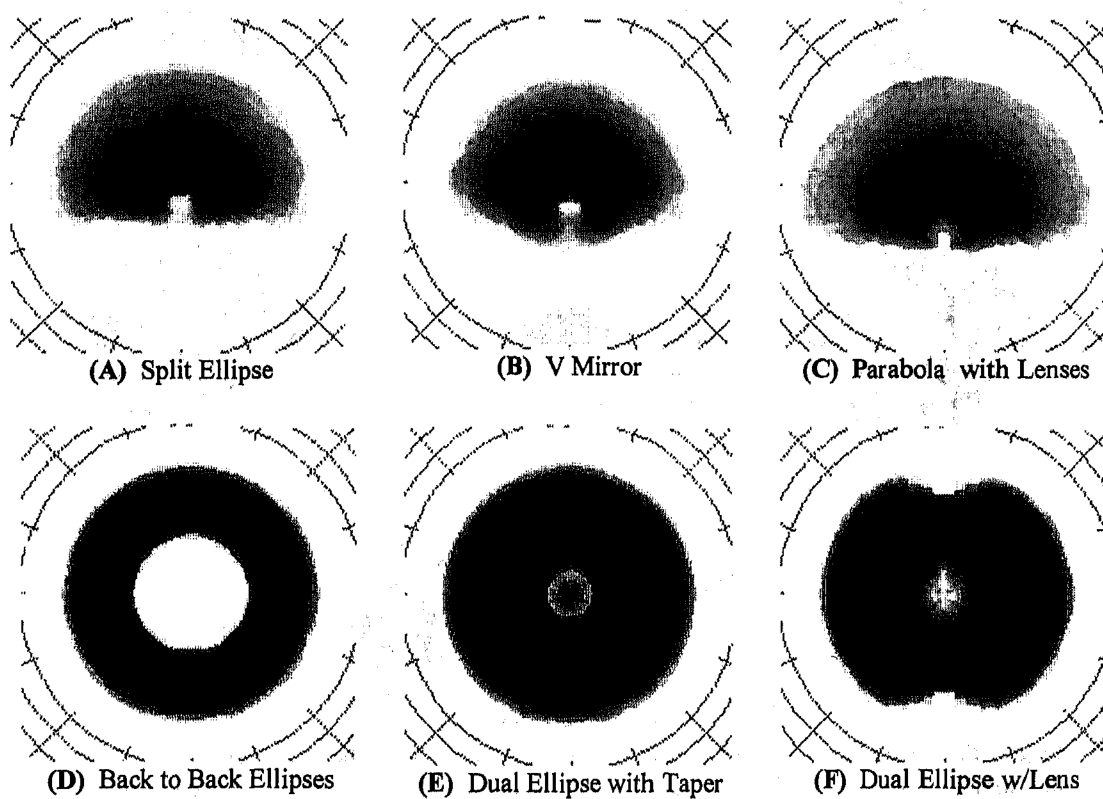
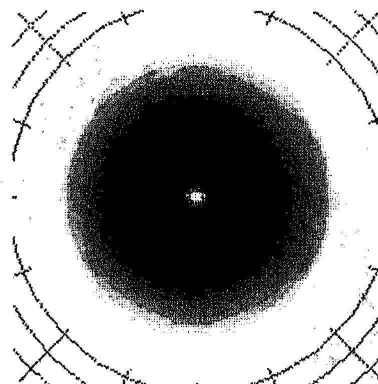
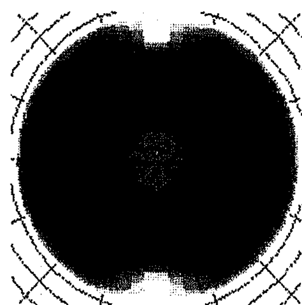


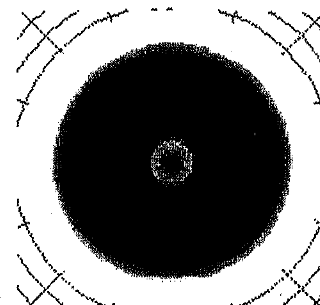
Figure 6. Intensity distributions for one of the two ports for each of the dual port systems considered. Note that first pictured circular grid-line around each distribution corresponds to +/-50 degrees, for an included angle of 100 degrees.

Figure 7. 'Single Ellipse' intensity (angular distribution.





Intensity (angular distribution) plots for the dual ellipse with tapers case at the entrance and exit of the taper.



(A) Entrance of Taper

(B) Exit of Taper

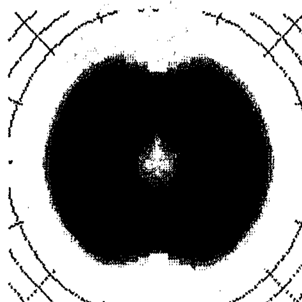


Illuminance(spatial distribution) plots for the dual ellipses with tapers case at the entrance and exit of the taper.

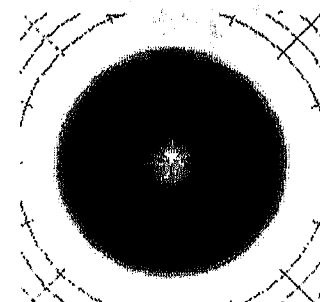


(C) Entrance of Taper

(D) Exit of Taper



Intensity plots for the dual ellipse with lenses case at the entrance and a short distance down a fiber.



(E) Entrance of Fiber

(F) A short distance down the Fiber

Input to Tapered Rod



Output of Tapered Rod

(G) Tapered Rod (Cylindrical) Orientation

Figure 8. Tapered rod and mixing in a fiber. (A)-(D) show what the illuminance and intensity distributions look like for the dual ellipse with tapers case before and after the taper. The angles have gone down and mixed annularly, and the spatial spot size has increased after the taper. (E) and (F) show what happens to the angular distribution of the dual ellipse with lenes case at the end of a fiber, and a short distance into the fiber. Note that the same annular mixing occurs as in (A) and (B), and that (F) is essentially identical to (B). (G) shows the orientation of the Taper.

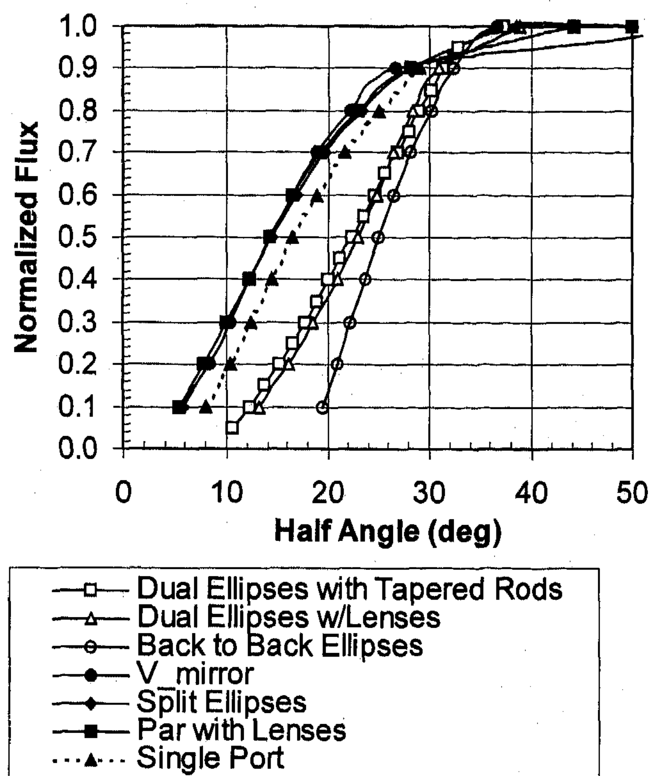


Figure 9: Intensity Encircled energy distribution curves (lumens within an angle) for one of the two ports in each of the dual port configurations.

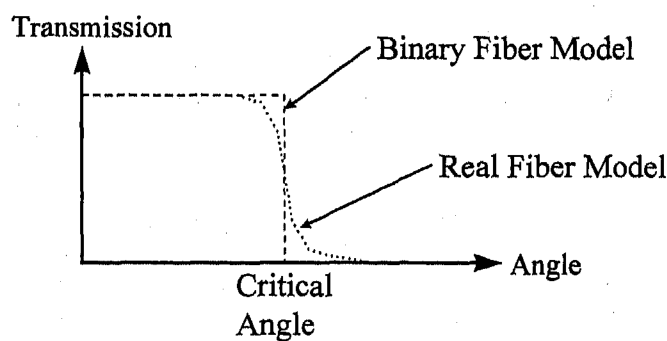


Figure 10: Comparison of binary and real fiber transmission models.

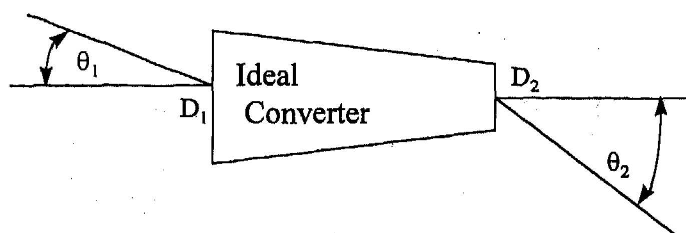


Figure 11: Ideal angle to area converter. The input and output are described by Equation 1.

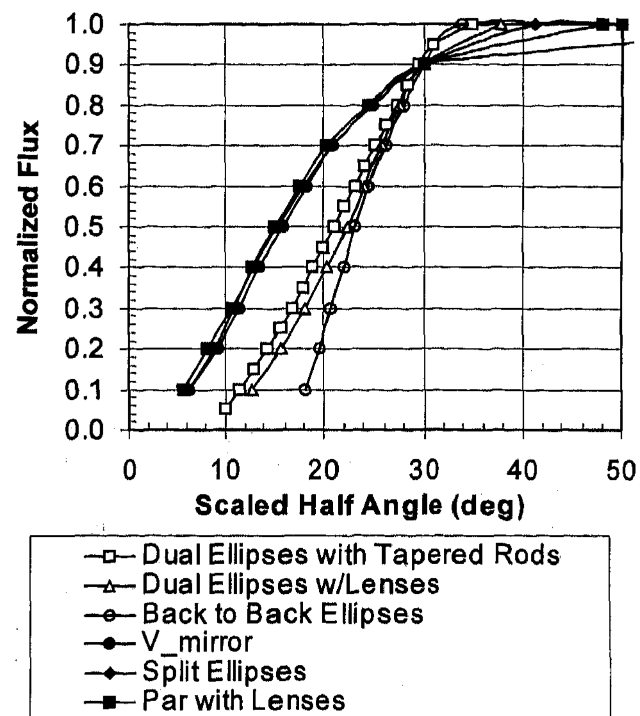


Figure 12: Intensity Encircled energy distribution curves (lumens within an angle) for one of the two ports in each of the dual port configurations, where the angles have been scaled such that the 90% point is the same for all cases.

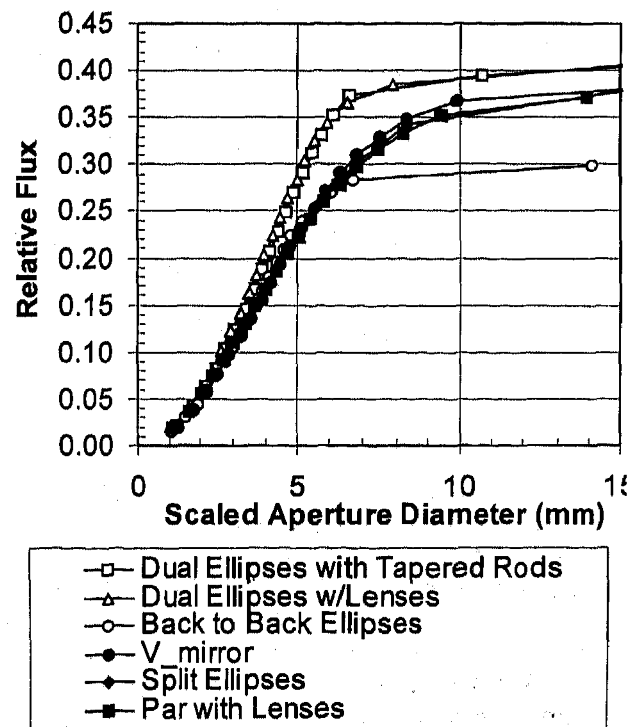


Figure 13: Illuminance Encircled energy distribution curves (lumens collected by an aperture) where the aperture diameter is scaled to reflect the angle adjustments shown in Figure12.

970261

Intelligent Audio Amplifier

Jan-Paul F. Huyser
Philips Semiconductors Nijmegen

Darby Hadley
Ford Audio Systems

ABSTRACT

An intelligent amplifier has been developed which delivers output power to either a two or four speaker audio system. The amplifier determines the number of speakers present during the start-up sequence and configures itself either as two BTL amplifiers or four single ended amplifiers. The system does not require any wiring modifications to operate in either mode and will reconfigure itself on each start-up cycle allowing system modifications, such as adding speakers, without additional wiring or hardware modifications.

Difficulties arise when determining the presence of speakers in an automotive environment. Extraneous noise sources such as door slams, hatch slams and engine cranking can cause erroneous measurements thus inappropriate configurations. The presented circuit was determined to be immune to these noise sources and is based upon:

1. Integration of measurements over a fixed period of time

2. Applying a large voltage to overcome noise sources without creating audible disturbances

3. Fixed switch on behavior

INTRODUCTION

A situation typically unique to OEM manufacturers is the requirement to offer a variety of audio systems on a particular vehicle. The problem that arises is complexity not only in the audio hardware but also in the vehicle wiring. A typical audio system line-up may range from a basic 2-channel am/fm radio to a 6 speaker CD system. In an effort to reduce radio complexity in the low to mid-range system and still offer an easy upgrade path, it is desirable to offer one wiring solution for a four speaker system, but have the radio work as either a 2 or 4-channel unit.

In a typical radio, the head unit would contain 4 amplifier channels. If the vehicle were only manufactured with two speakers, 2 of the amplifier channels would be wasted. Other solutions exist where the front and rear amplifiers are out of phase and in a two speaker system, the speaker could be connected between the front and rear channels to create a bridge amplifier delivering four times the power to the front individual speaker. The problem has been that this would either require a unique vehicle wiring harness or a modification to the head unit. Also, if the customer would ever want to add a second set of speakers, again the wiring would either have to change or the head unit would have to be modified.

The question then became, "wouldn't it be nice to offer one set of vehicle wiring and allow the radio to configure itself based on the number of speakers present in the vehicle".

This paper describes an audio system that detects the number of speakers present during a start-up cycle and either configures the radio to deliver 2 channels of 21 watts/channel or 4 channels at 6 watts/channel. The system will reconfigure itself on each power cycle to allow for modifications to the system, such as adding additional speakers at a later time.

The detection method is paramount to the successful operation of the system. Upon early investigations it was determined that there exists a large number of noise sources in the automotive environment that could have an impact on the load detection method thus the system had to be proven reliable under a variety of interference situations.

DESIGN CONCEPT

An auto-detect 2/4 channel amplifier can switch automatically from 4xSingle Ended (SE) to 2xBridge Tied Load (BTL) operation without changing wiring. During the start-up cycle of the amplifier (switching from 'off' to 'on') the presence of the rear speakers are measured. If one of the rear speakers is present, the amplifier is switched automatically to a 4xSE configuration with an output power of 6W per channel. A stereo BTL configuration is selected when **no** rear speakers are detected, allowing

output power of 21W per channel to be delivered to the front speakers (Supply voltage of 14.4V @ THD=10% and R_L=4 Ohm).

Problems associated with load detection in an automotive environment:

1. The speaker suffers from microphone effects. During door (hatch) slams the speakers generate large voltage swings which can disturb the load detection. To avoid incorrect load detection, a large voltage swing will be generated during the load detection.
2. Long wiring to rear speakers suffer from induced noises from other vehicle electrical sources while the engine is running.
3. At engine cranking the sudden supply voltage drop can disturb the load detection.

Extra design criteria are:

1. No plops allowed during load detection.
2. No additional external components

AMPLIFIER PRINCIPLE

Consider an amplifier with 'a common buffer', a 'front' amplifier and a 'rear' amplifier. The presence of a load between the common buffer and the rear amplifier is detected during the 'start-up sequence ' which determines the mode of operation.

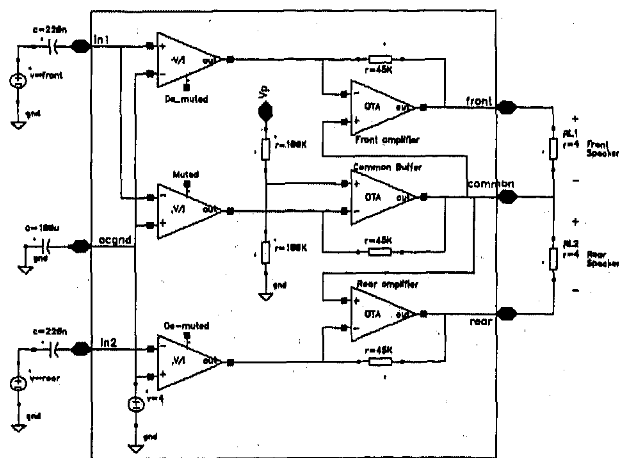


Figure 1 Amplifier in 2xSE configuration with common buffer

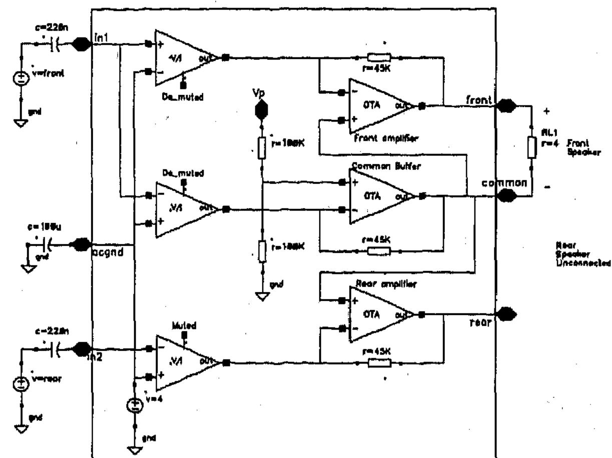


Figure 2 Amplifier in BTL configuration with common buffer output

The amplifier can be controlled in three states:

- Standby	amplifiers switched off (low standby current.
- Mute	common buffer dc biased
- On	input stages and front/rear amplifiers un-muted

In Figure 1, the SE application is given. The amplifier is switched to SE operation (after detection of the rear speaker) by muting the input signal of the common buffer. To avoid dc currents through the load the common buffer forces the output voltages of the 'front' amplifier and 'rear' amplifier at half supply voltage (Vp/2). The system is configured in such a way that the 'rear' amplifier compensates the voltage swing of the 'front' amplifier. When the 'front' and 'rear' input signals are equal and in phase, the output signals of both amplifiers have an 180^0 phase shift (which is compensated at the connector block of the radio). No current will flow into the common buffer and because of this configuration dissipation is minimized.

In the absence of a 'rear' load (Figure 2) the BTL mode is selected after the start-up sequence. During this mode of operation the 'rear' amplifier is disabled (high ohmic) and the common buffer is unmuted. Now the common buffer will complete the bridge of the BTL amplifier (the front amplifier now operates in BTL mode).

BASIC LOAD DETECTION

The load detection principle is explained in three steps:

1. Basic load detection principle by using Figure 3.
2. The load detection principle implemented in an BTL amplifier circuit (see Figure 4)
3. Reduction of external timing/integration components.

At first the 'common' buffer is switched 'on' and due to the fact that Va is rising gradually from 0V to Vp/2, the output (Vb -Figure 3) of this buffer will also rise gradually. The rear amplifier is disabled (high ohmic) during the start-up sequence in the conditions 'standby' and 'mute' and during load detection.
The output of the rear amplifier is forced low by the load detection circuit using the pnp transistor (Q1). When the output of the common buffer has reached the final output voltage level of Vp/2, the output of the rear amplifier is gradually released by the load detection circuit (Vd-Figure 3 increases from 0V to Vp) Due to the load (or an internal pre-load) the output of the 'rear' amplifier now gradually increases from 0V to the output level of the common buffer (Vp/2). From this point onwards, the base-emitter junction of the pnp transistor Q1 is forced reverse. When Vd reaches the voltage Vp, the load detection is finished. Thus a voltage swing of Vp/2 is generated which is applied over the 'rear' load (see Figure 3, voltage Vb-Vc in timing diagram).
During the load detection transistor Q1 will only conduct current when a load is present. This current is measured and integrated over time by the control-block. When the integrated value of this current exceeds a fixed level, a an output signal is generated which indicates a load is present. This output signal is stored in a memory cell which is used to put the amplifiers in SE or BTL mode. Now the amplifiers are enabled.

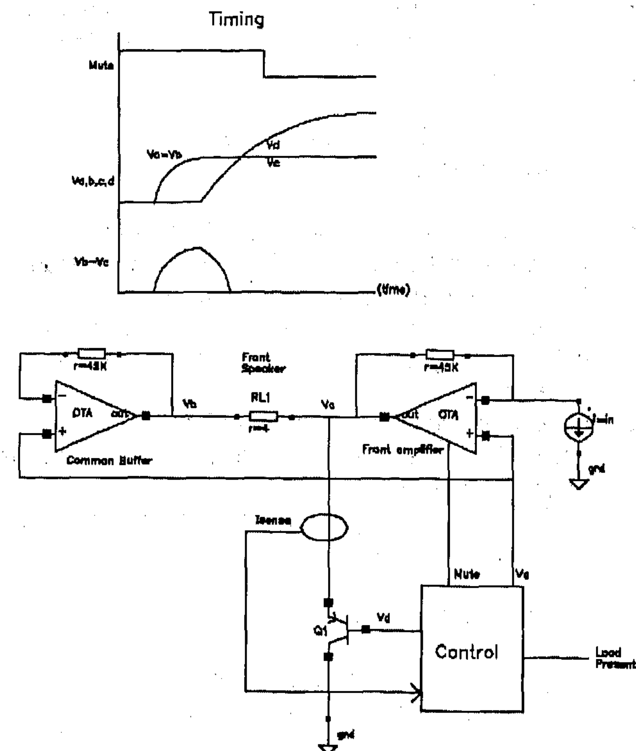


Figure 3 Simple load detection circuit

LOAD DETECTION IMPLEMENTATION

Implementation of the control block is given in the application circuit of Figure 4. The intelligent amplifier can be controlled by the mute/standby interface. When a voltage of :

$$V_{E2}+V_{E3} < V_{msb_on} < V_{E2}+V_{E3}+V_{E4}$$
$$V_{msb_on} = 4V \text{ (= mute level)}$$

is applied to the 'msb_on' pin of Figure 4, the voltage on capacitor C4 gradually starts to rise from 0V to V_{msb_on}. At the moment the voltage V_{C4} exceeds the value V_{E2}, comparator 1 will switch on the supply voltage of resistors R1 and R2. Now, because of the presence of C1, the voltage Va will rise slowly from 0V to Vp/2. The timing C4xR4 should be larger than C1xR1/2. The buffer which is connected as a voltage follower stage, will generate an output voltage equal to Va. The output of the 'rear' amplifier is disabled (high ohmic) by comparator 3 (Comp3) but will be forced low by means of Q3, Q4 and Vd (=0V).
Comparator 2 will switch when V_{C4} exceeds $V_{E2}+V_{E3}$. Now the voltage Vd (=V_{C3}) will rise from 0V to Vp, causing the voltage Vc to rise also. When a load is connected, a current will flow through both Q3 and Q4. Capacitor C2 is being charged when the current of Q4 exceeds the threshold level of I_{ref}. Capacitor C2 will integrate the load current during the load detection. When the voltage Vd exceeds the half supply voltage (Vp/2), the state of C2 (charged/uncharged) will be

stored into a memory cell. The contents of this memory cell indicates if a load is present. Furthermore, because a large voltage swing of Vp/2 is applied to the speaker, a large load current will flow, minimizing the potential of false measurements. Due to the large voltage and the period of integration, the interference generated by the microphone effects of the speakers will have little impact on the detection mechanism. To avoid 'pops', the signal applied to the speaker:

1. must be smooth with no sudden changes.
2. voltage change dV/dt must be small ->t_{detect}>0.4sec.

By applying 9V to the V_{msb_on} pin the voltage on capacitor C4 (V_{C4}) will rise again from it's previous value (4V) to the voltage applied on pin V_{msb_on} (9V). When the voltage V_{C4} exceeds V_{E2}+V_{E3}+V_{E4}, the output of comparator 3 (Comp3) will change state. Now the 'rear' amplifier is disabled.

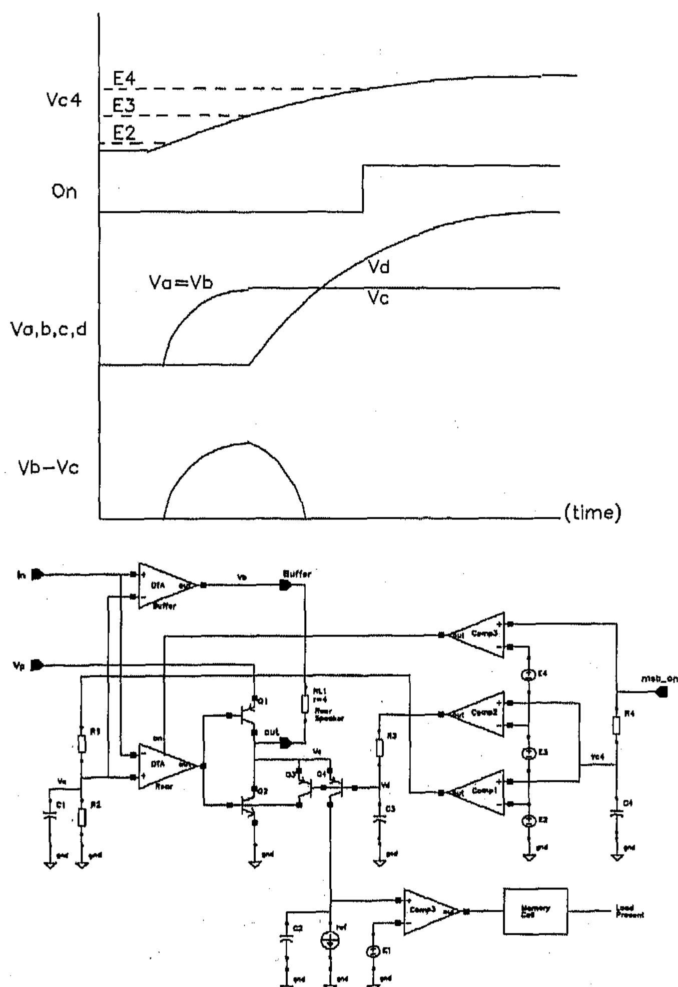


Figure 4 Load detection Implementation

NTERFERENCE FROM THE VEHICLE

Measurements were taken in a typical audio ıystem with a variety of speakers and locations to determine the contribution of various sources. Figure 5 shows that voltage induced by the microphone effect from a typical speaker during sudden cabin pressure changes caused by door slams. As can be seen, the large voltage swing and amount of induced energy onto the speaker terminals can not be ignored.

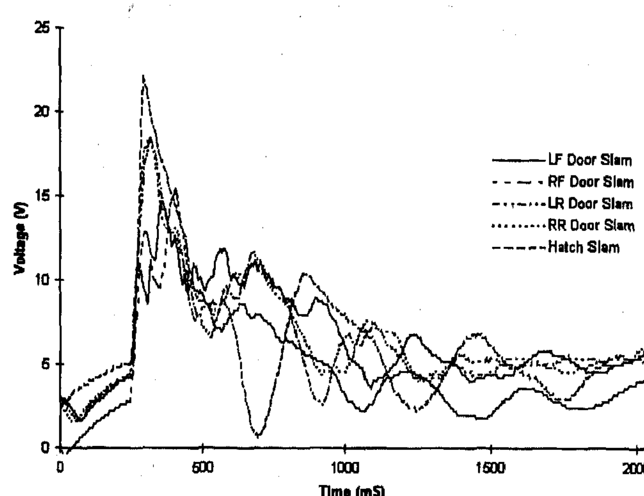


FIGURE 5 - Speaker Voltage from Door Slam

Figure 6 shows interaction of the detection signal with the door slam interference. The integration time for the detection circuit had to be long enough to prevent interference from the microphone effect of a door slam.

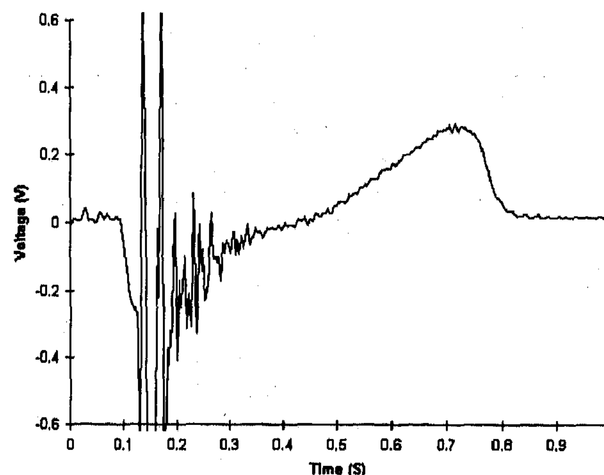


FIGURE 6 Differential Speaker Voltage of Door Slam Interference During Load Detection

REDUCTION IN TIMING COMPONENTS

As can be seen from Figure 4, many timing and integration components are needed. By combining circuits, the extra timing and integration components can be eliminated. In the intelligent amplifier, a signal-ground capacitor and a capacitor on the mute-standby pin are used as the timing capacitor and the integration capacitor. Because of this reuse, no extra components are needed.

The application circuit of the complete intelligent amplifier is given in Figure 7. and in Figure 8, the switch-'off' (=standby) and switch 'on' curves are given.

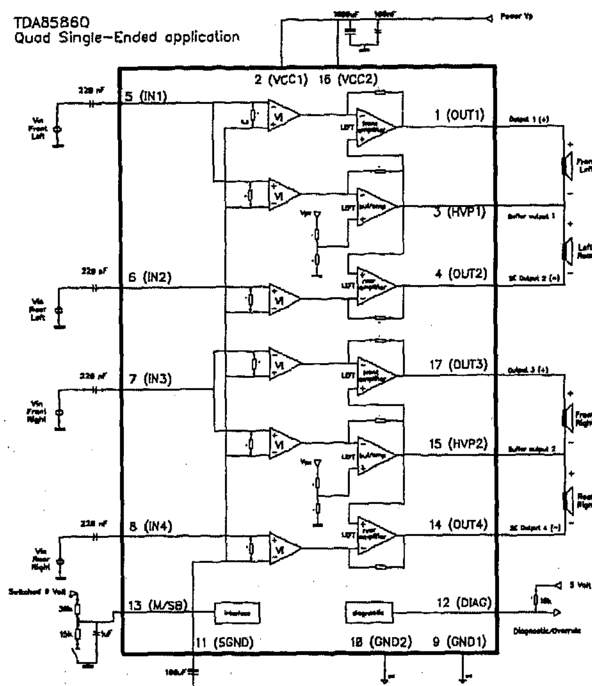


Figure 7 Application Intelligent Amplifier

For integration, the capacitor on the standby-mute-on pin (pin 13) is used. Timing is accomplished by charging the 'signal ground' capacitor on pin 11 (see Figure 7). When the amplifier is switched from 'standby' to 'mute', the 'signal ground' capacitor is being charged from 0V to 4V. During this charge action: the control signals and the voltages Va + Vd (see Figure 7) are generated. Now only one capacitor controls the complete switch on time of the amplifier. Integration is accomplished by the capacitor on the standby-mute-on pin of the IC. Normally this capacitor is needed to avoid switching pops during the transition from 'mute' to 'on'. When a load is present, an additional charge current will flow out of this control pin during the load detection, which will lift the voltage on this pin (see Figure 8 V_{msb_on} curve from time frame: 2sec<t<4sec). The extra charge in this capacitor is now being detected by the load detection control block, which will force the IC into 4 channel or 2 channel mode. In Figure 8, the curves are given when rear loads are connected (amplifier will switch to 4 channel mode).

Smooth voltage changes are generated during the load detection as can be seen from Figure 8. When 4 Ohm speakers are used, a maximum voltage of 4.5V is generated during the load detection sequence. Because of the slope, no pops are heard during the load detection. The diagnostics pin signals the result of the load detection process to allow the radio software to configure itself as needed, such as disabling the fader control. Also, there is the opportunity for the microcontroller to override the automatic amplifier configuration if there is a need.

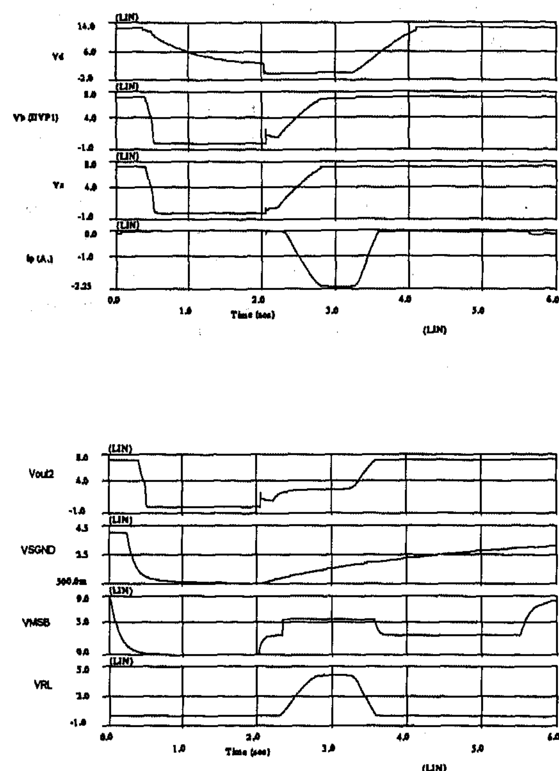


Figure 8 Wave-forms of intelligent amplifier

CONCLUSION

The intelligent amplifier offers a unique design solution to car radio manufacturers allowing a reduction in head unit complexity as well as offering flexible options to system designers.

The principle for load detection presented allows for the successful operation of the intelligent amplifier but also has other uses for future designs such as enhanced diagnostics of the system.

The detection method reuses existing components attached to the amplifier thus alleviating the need for any additional external components to the amplifier.

The principles presented in this paper have been tested in a real automotive environment and have proven to be robust. Now, a car radio can be designed which will determine its optimum configuration depending on the audio system. No changes in wiring will be needed allowing for standardized vehicle wiring.

970262

The Evolution: From Car Audio to Digital Mobile Multimedia

Richard Zerod
Ford Motor Co.

ABSTRACT

New advances in wireless and digital communication technologies are causing the traditional car audio system to undergo a radical transformation. The historic function of providing high-quality music and sound reproduction to the consumer will continue; however, additional data and information services will soon be available to consumers using ancillary high-speed data channels available in the new Digital Audio Broadcasting (DAB) systems.

The European developed EUREKA/147 DAB system represents leading-edge technology and has already been introduced in several global markets. This proven system has demonstrated the capability for providing excellent quality music, along with significant data-casting capabilities. Potential benefits to the consumer include the ability to receive traffic management information, digital news and weather services, slow-speed video images, and even subscription services such as personal paging and messages. The traditional "audio" system will soon become an integral part of the wireless-link to the many information and data-services which are currently available.

INTRODUCTION

A HISTORICAL PERSPECTIVE - A historical perspective can often provide insight into trends for new technological developments. The evolution of sound and data-broadcasting has been based on a history of continuous innovations and developments over several decades. Some of the more significant events are shown in **Figure 1** [1].

Prior to 1980, the primary emphasis was placed exclusively on improving the sound quality of analog broadcast systems. The re-allocation of the FM broadcast band (i.e., to minimize interference during tropospheric propagation) and implementation of FM stereophonic broadcasting, represent example improvements.

During the 1980s, the initial consumer application of *data* broadcasting occurred. Initially deployed in Europe, the Radio Data System (RDS) was the first wide-spread application of data/information broadcasting that was targeted at the consumer market. With a total channel capacity of 1.2 kbps, the data throughput is relatively slow; however, the system has proven useful for conveying a traffic announcement signaling information flag via a 57 Khz subcarrier on existing FM analog broadcasts.

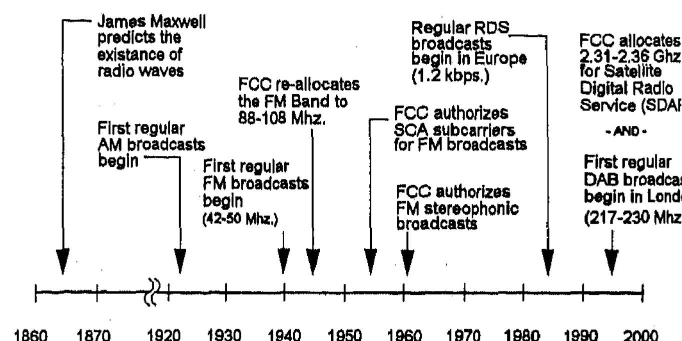


Figure 1 - A Historical Perspective

In the 1990s, initial deployment of *digital* broadcasting systems occurred. Once again, Europe is the global leader in the technology with the implementation of the EUREKA/147 DAB system. Since this system is designed specifically for the transmission of digital information, which includes digitized audio and additionally ancillary data, potential cross-talk and data-channel capacity issues – typically associated with transmitting data subcarriers on an analog FM broadcast system – are eliminated.

ENABLING TECHNOLOGIES - The synergy among several new technologies has provided the foundation for the development of the EUREKA/147 DAB system. Some of these include:

<u>Digital Coding and Bit-Rate Reduction Schemes</u> The bandwidth required for digital transmission is directly related to the serial data-rate of the source information. A clear advantage exists when audio (or video) compression techniques are utilized for eliminating perceptually unnecessary information from the source material.

The Motion Pictures Experts Group (MPEG), Layer II protocol is one of several popular audio bit-rate reduction schemes and is utilized in the EUREKA/147 DAB system. This compression technique can produce high-quality stereo audio at a data-rate of 256 kbps[1]; the resulting audio provides "no perceptible difference between an original sequence and its coded version."[2] The flexible EUREKA/147 standard also has provisions for the use of higher compression levels for applications where lower data-rates can be tolerated (e.g., talk material, etc.) and higher bandwidth reductions are desired.

The Joint Picture Experts Group (JPEG) protocol, which is widely used in Internet applications, is gaining popularity for the transmission of still-frame video images on the EUREKA/147 data channels. This lossy compression technique is well-suited for "real-life" images where some distortion of edges or boundaries can be tolerated. The variable compression technique utilized in the JPEG protocol provides the ability for the user to compress an image until a given quality-level is achieved; very useful images can be obtained using compression ratios of >10:1.

Forward Error-Correction Coding Systems - Signal fading or noise in a mobile environment can produce random or burst bit-errors which ultimately degrade the received signal quality. The addition of redundant data, used for Forward Error Correction (FEC) coding, is an essential element for successful wireless data transmission. (This is particularly important for one-way, high-speed data transmission to mobile receivers where the re-transmission of corrupted or missing data is not practical.)

A Bit-Error-Rate (BER) of 10^{-4}, after error-correction, is generally required for reproducing high-quality audio material; the requirement for video-image transmission is more stringent. To achieve a high-degree of consumer acceptance, these BER's must be attained over substantially all driving conditions and environments within the DAB service area.

The EUREKA/147 DAB system utilizes a concatenated error-correction scheme consisting of Reed-Solomon and Convolution encoding to protect against both random and burst error conditions. The amount of error-protection, which is correlated to signal reception robustness and service area coverage, can be chosen by the individual broadcaster according to their performance requirements. For audio signals, Unequal Error-Protection (UEP) is employed where the protection coding is applied non-uniformly through the audio frame to provide better protection to the more critical portions of the audio data. The UEP aids in producing a "graceful degradation" effect where the quality of the decoded audio signal slowly deteriorates as the receiver exits the DAB service coverage area.

Five different error-protection levels have been established to facilitate different operational requirements. The average code rate, R_{ave}, varies from 0.34 for very high protection, to 0.75 for applications where low protection levels can be tolerated (e.g. cable distribution).[3] For mobile broadcast applications, an R_{ave} of 0.5 is typically used.

Complex, Digital Modulation Systems - Traditional analog broadcasters have superimposed the audio source material onto an RF carrier using Amplitude Modulation (AM) or Frequency Modulation (FM) techniques. However, data-transmission systems typically utilize a combination of amplitude, phase, or frequency modulation techniques to transpose the serial data stream onto an RF carrier.

For a given modulation technique, the required RF spectral bandwidth is directly related to the serial data rate. The bandwidth efficiency, which is quantified in bits-per-second-per hertz (bits/sec/Hz), indicates the required transmission bandwidth as a function of the data rate. Through the utilization of complex, multi-level modulation schemes, bandwidth efficiencies of several bits/sec/Hz. can be readily achieved. Unfortunately, the improved bandwidth efficiency generally requires a higher signal-to-noise ratio (SNR) to maintain a good BER. Additionally, the demodulation circuitry becomes more complex, and generally, more expensive.

The EUREKA/147 system uses a form of Quadrature Phase-Shift-Keying (QPSK), where an RF carrier is shifted in increments of 90°. Four possible carrier phase states are possible, where each phase represents the logic-state of 2 successive bits (often referred to as a "symbol"). A common method for generating a QPSK signal is shown in **Figure 2**.

---

[1] Assuming each channel of a 16-bit CD source is originally sampled at 48.0 Khz, the resulting 256 kbps stereo audio is compressed by a factor of:

(2 x 16 bits x 48.0 Khz) / 256 kbps = 6.0 : 1

[2] Based on extensive testing by the ISO and ITU-R during the period between 1990-1993.

[3] The average code rate, R_{ave}, is defined as:

R_{ave} = [Audio data / (Audio data + FEC data)]

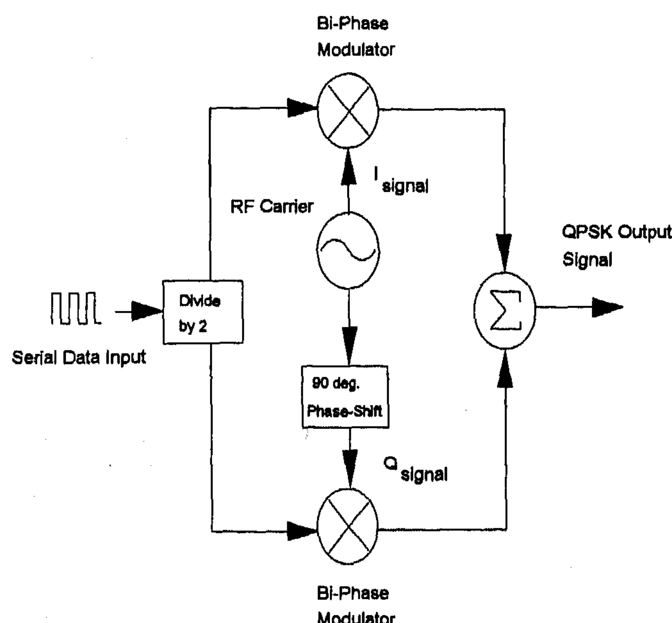


Figure 2 - QPSK Signal Generation System

The theoretical bandwidth efficiency of QPSK modulation is 2 bits/sec/Hz. However, low-pass filtering is utilized to roll-off the low-energy sideband components and minimize the occupied RF bandwidth. After filtering, an aggregate bandwidth efficiency of 1.5 bits/sec/Hz. has been achieved.

Sub-micron Levels of IC Integration - Practical implementation of the required digital demodulator and error-correction algorithms requires the use of integrated circuits employing high levels of integration. Complex gate structures, requiring several hundred-thousand gates, are required for sophisticated demodulation and decoding schemes. This requires the use of sub-micron integration levels to achieve the required functionality using only a relatively small number of IC's in the end product.

Through the use of technologies allowing IC geometries of <1 micron, high-density gate structures have been successfully developed for the EUREKA/147 DAB system. Due to the significant engineering development time required, the decoding systems are still relatively expensive and represent a major portion of the receiver cost. However, as with other new consumer technologies, increased competition and volume-levels are expected to drive costs down with time.

Direct Broadcast Satellite Systems - Direct broadcast satellite delivery can potentially provide a cost-effective, alternative solution for program delivery to large geographic areas where the development of a terrestrial broadcasting infrastructure is not economically feasible. Advances in power generation and storage systems, high-efficiency RF power amplifiers, and high-gain antenna systems have resulted in satellite broadcasting systems capable of delivering an aggregate effective radiated power between +50 dBw to +60 dBw.

However, the path-loss for a satellite in geo-synchronous orbit can approach -200 dB under line-of-sight conditions. Accounting for the relatively small receiving antennas (i.e., with relatively low-gain) that would be required for a practical automotive application, link-margins of <10 dB are to be expected. Certain geographic areas, such as the "urban canyons" encountered between tall buildings, would result in signal blockage that results in loss-of-signal. However, work is continuing on improvements in coding-gain, modulation techniques, and other parameters to further improve the system performance under marginal reception conditions.

DISCUSSION

BRIEF OVERVIEW OF THE EUREKA/147 DAB SYSTEM - The EUREKA/147 DAB system is designed to provide high-quality, multi-service digital radio broadcasting for reception by vehicular, portable, and fixed receivers. It is designed to operate at any frequency up to 3000 MHz for terrestrial, satellite, or hybrid (satellite and terrestrial) systems. The system uses advanced audio compression techniques to remove perceptually unnecessary information. FEC is provided utilizing Reed-Solomon and Convolution encoding. The transmitted signal is spread in the frequency and time-domains to provide a high-quality signal, even under frequency selective multipath fading conditions. A total channel capacity of 2.4 Mbits/sec allows for multiplexing several different audio programs, as well as high-speed ancillary data, into a 1.55 MHz wide RF channel. The conceptual block diagram of the DAB signal generation system is shown in **Figure 3** [2].

Each digitized audio channel is applied to an audio encoder that performs the MPEG, Layer II bit-rate reduction algorithm. The amount of compression is adjustable for each stereo channel to achieve a balance between the audio quality and channel capacity utilized. A channel encoder then adds the FEC data to the compressed audio serial data; R_{ave} is adjustable to achieve the desired signal robustness.

Time interleaving is performed to improve the ruggedness of the error-correction algorithms in a time-variant transmission channel. [4] The resulting signal is multiplexed into the Main Service Channel with other similarly processed signals.

The multiplexer output is subdivided into a large number of bit-streams, each having low individual bit-

[4] In the case of mobile reception where deep fading is encountered, burst-errors are likely to occur. During the de-interleaving portion of the decoding process, the burst-errors are spread out in time. The resulting bit-errors are more easily corrected by the FEC algorithms.

rates, which are used to phase modulate individual orthogonal RF carriers. By inserting a temporal guard band between successive symbols, which is larger than the delay spread expected from multipath interference, inter-symbol interference will not occur.

The multiplexer control information, which contains the configuration information of the individual service channels, is transmitted on the Fast Information Channel (FIC). Since this information must be rapidly

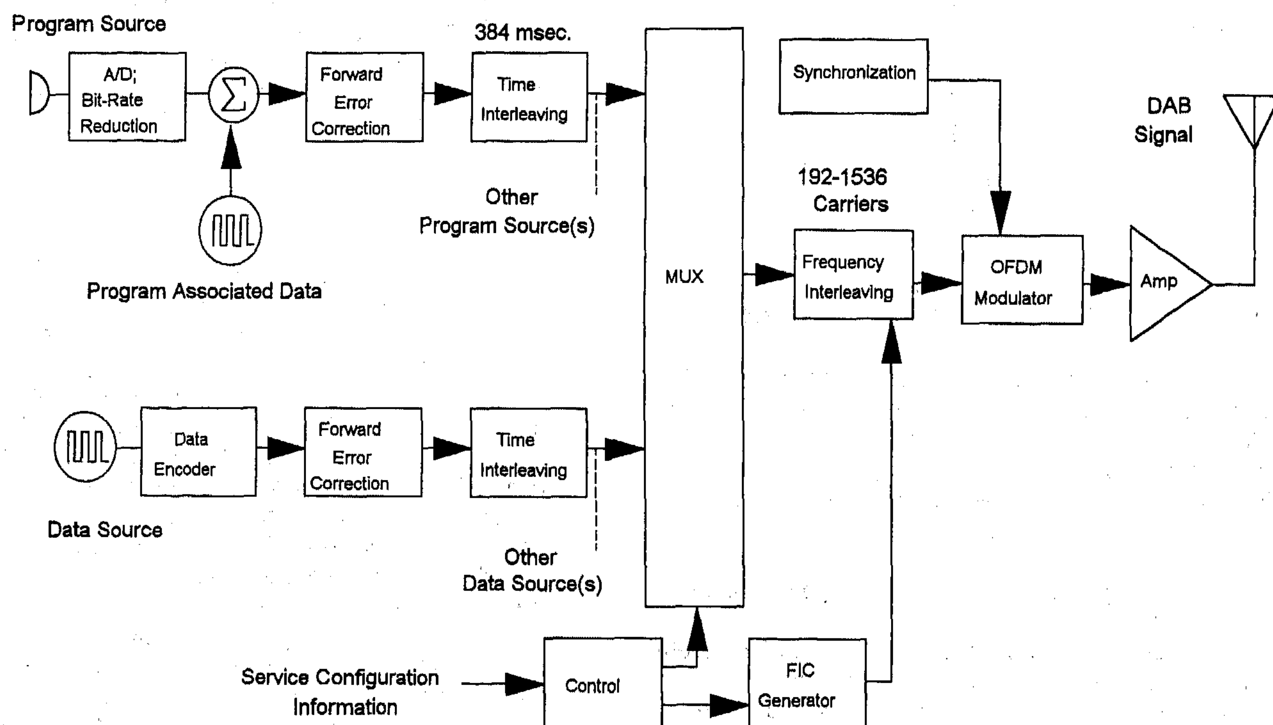


Figure 3 - EUREKA/147 DAB Signal Generation System

available at the receiver to facilitate the decoding process, it does not go through the time-interleaving process to avoid the associated time-delay. However, this data is highly-protected with a low R_{ave} and frequently repeated to assure the integrity.

The multi-carrier Coded Orthogonal Frequency Division Modulation (COFDM) system results in an output spectrum that is approximately rectangular and occupies a bandwidth of 1.55 MHz. A residual amplitude modulation artifact exists due the individual carrier amplitudes reaching their respective minima values at different times. The significance of low-noise receiving system designs, to minimize the BER during periods of low-instantaneous signal level, becomes apparent upon examining the wide-dynamic range of the COFDM signal. For reference, examples of both frequency and time-domain measurements of a EUREKA/147 DAB signal are shown in **Figure 4** and **Figure 5**, respectively.

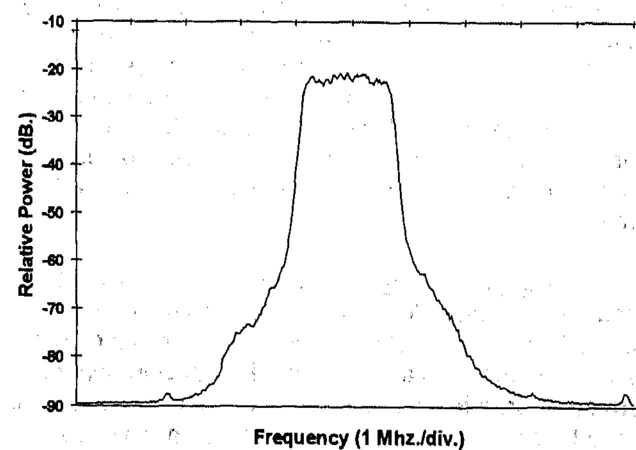


Figure 4 - EUREKA/147 Frequency Domain Signal

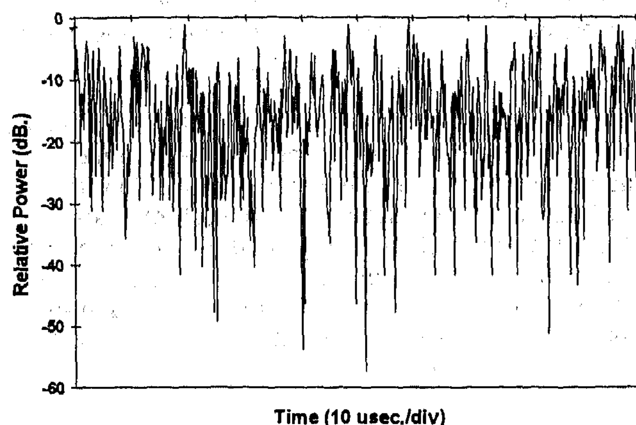


Figure 5 - EUREKA/147 Time Domain Signal

The EUREKA/147 system has a special feature which allows a broadcasting network to be extended by the use of multiple transmitters, each operating on the same frequency. The Single Frequency Network (SFN) feature permits the DAB service area coverage to be specifically tailored to suit the desired reception area contour. Through the use of an appropriate guard-interval, which is the difference between the transmitted symbol period and the symbol analysis time within the receiver, inter-symbol interference - resulting from the reception of multiple, time-delayed signals - is effectively mitigated.[5] Provided the maximum delay-spread of the different SFN signals do not exceed the guard-interval, the receiver can constructively integrate the signals into a coherent data stream [3]. This attribute also allows for the use of low-power gap-filler transmitters to supplement the service coverage in specific, localized areas of poor reception.

GLOBAL DEPLOYMENT STATUS - The EUREKA/147 DAB system has been adopted as the European Telecommunication Union Standard (ETS) 300 401, Radio Broadcast Systems; Digital Audio Broadcasting (DAB) to mobile, portable, and fixed receivers. Global implementation is leading in Europe and Canada where full-time DAB service is presently available, or shortly planned for, several major cities. Many other countries are conducting field-tests or have allocated the required frequencies necessary for the future support of EUREKA/147. The global deployment status is shown in **Figure 6**[6]; market expansion is expected to continue for the next several years.

The U.S. has not adopted EUREKA/147 as the domestic DAB standard. Efforts are presently underway to test and evaluate several different DAB systems for the U.S. market. Although the high-technical performance level achieved by the EUREKA/147 system is widely acknowledged, issues presently exist in the U.S. regarding the additional spectrum required for deployment of this system. The majority of effort is presently focused on developing an in-band DAB system that utilizes the existing FM broadcast band, and can co-exist with the present analog broadcast stations.

DATA-TRANSMISSION CAPACITY - Considerable interest exists within the broadcast industry regarding the potential benefits of *data-casting* and transmission of ancillary information to supplement the main audio program material. The flexible architecture of the EUREKA/147 DAB system contains several possible methods for multiplexing additional serial-data onto the composite transmission ensemble.

The Extended Program Associated Data (X-PAD) channel can be used for the transmission of information which is related to the audio material for the particular stereo channel. The ancillary data, which is time-multiplexed into the audio data frame, has a maximum capacity of 64 kbps. In addition to carrying information such as song title, artist, etc., initial development testing has shown the X-PAD channel could be useful for the transmission of still-frame video images (e.g., traffic and weather maps) and other types of data that could potentially be useful to the vehicle occupants.

Additional general-purpose data sub-channels are also available within the main service channel with a capacity of n* 8 kbits/sec. The information carried in this mode is not multiplexed with the audio frame data, and consequently, is not necessarily related to any specific audio program.

The high data-channel capacities available with the EUREKA/147 system provides a foundation for future, wide-area data-broadcasting applications. Initial applications are likely to include Intelligent Transportation System (ITS) functions, where the real-time transmission of large amounts of traffic-management data could be required.

CONCLUSIONS/FUTURE TRENDS

The function and role of the traditional audio system is expected to evolve into one of the principal transport mechanisms for the reception of data and information into the vehicle. Certain trends in the not-too-distant future can be reasonably extrapolated based on recent technological developments; some of these include:

- Europe and Canada will continue to lead in the global implementation of Digital Audio Broadcasting services. Multimedia applications will take advantage of

[5] In this sense, the EUREKA/147 system can be considered somewhat analogous to a "cellular-type" broadcast system. The service coverage can be extended by the use of a network of additional transmitters that provide additional reception coverage.

[6] For purposes of illustration, the EUREKA/147 deployment status is shown on a country-by-country basis. However, DAB service is not necessarily available in all the geographic areas within each country listed as "On Air".

the flexible, high-speed data channel capacity of the EUREKA/147 system; initial applications will be used as a supplement to the existing audio program material.

- "Data-casting" for vehicular applications will become increasing more significant. High-capacity data channels will be required for reliable, wide-area broadcast coverage of real-time information. Future applications will focus on information which is useful to the vehicle occupants; this is likely to include traffic management information, weather conditions, and customized subscription information services.

- Satellite DAB delivery, particularly in areas presently under-served by the terrestrial broadcast system, will likely gain in popularity. Earth-station receiving requirements will be minimized due to the relatively high satellite transmitting power.

- Significant hardware and software changes within the audio system will be required to take advantage of all the new features available with the EUREKA/147 DAB system. Initial system costs are expected to be relatively-high and the first applications will be on high-end vehicles. New opportunities exist in the design and manufacture of new antenna systems, tuners, digital decoders, displays, and the human-machine interface.

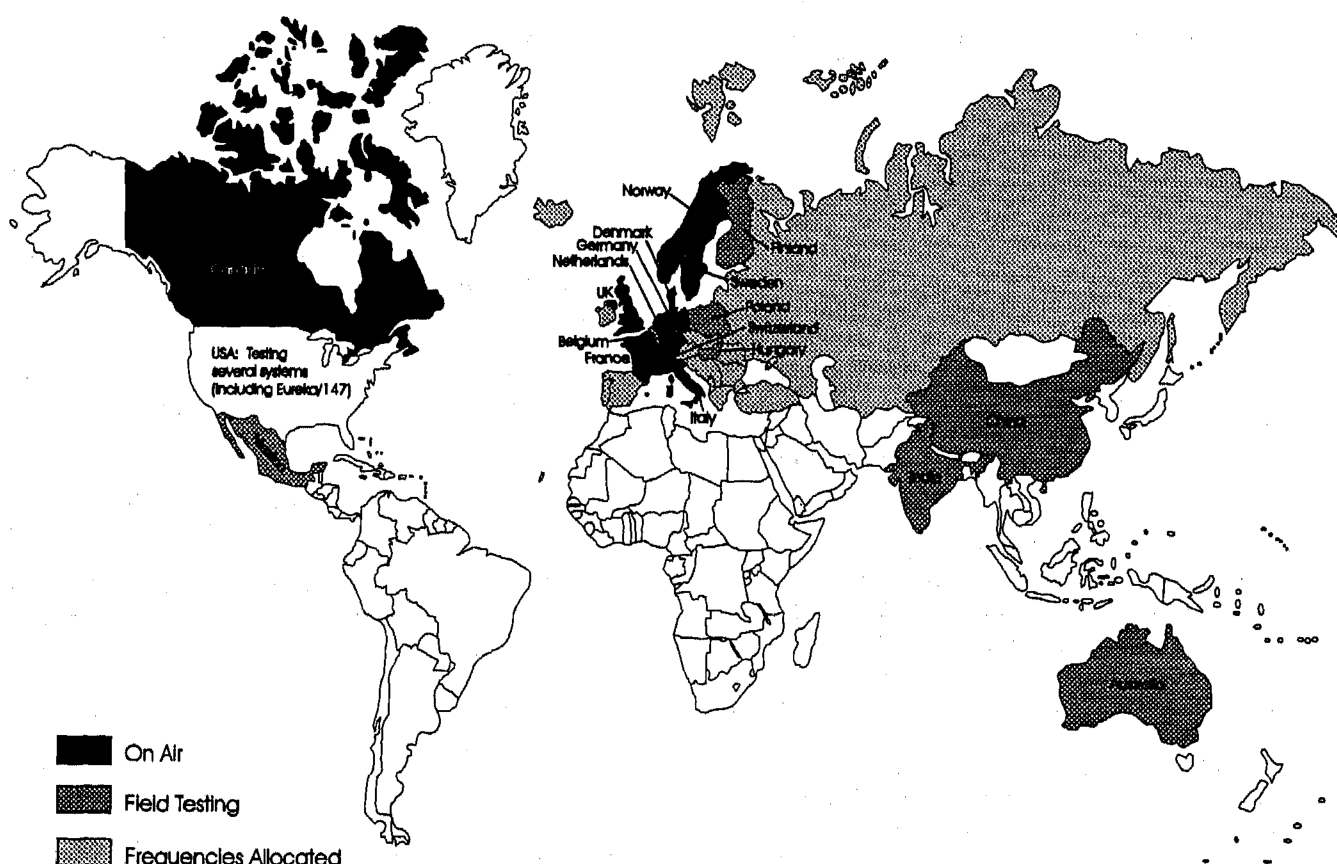


Figure 6 - EUREKA/147 DAB Global Deployment Status

ACKNOWLEDGMENTS

Acknowledgment is given to Mr. Carlos Altgelt of Ford/ACD for his assistance in compiling some of the information presented in this paper.

REFERENCES

1. "Broadcasting and Cable Yearbook 1993"; R.R. Bowker Publications; New Providence, New York.

2. "Digital Audio Broadcasting System, Guidelines for Implementation and Operation"; July, 1995; EUREKA Project 147.

3. "The Digital Audio Broadcast Single Frequency Network Project in Southwest Germany"; T. A. Prosch, IEEE Transactions on Broadcasting, Vol. 40., No. 4, December, 1994.

ABOUT THE AUTHOR

Rick Zerod is a Senior RF Technical Specialist with Ford Motor Company, Automotive Components Division. Some of his previous projects have included the design of several radio chassis and leadership of the development team in the recent introduction of Ford's concealed AM/FM antenna systems.

Presently, Rick is involved in the development of Ford's DAB system design strategy. He is also involved in several RF field measurement/simulation projects, and has participated in the NRSC DAB Field-Test Task Group. Rick has a BSEE degree from Lawrence Technological University and holds 4 U.S. patents. He is an Amateur Radio operator (W8CZ) with a principal interest in the study of antennas, UHF/microwave propagation and digital-modulation techniques.

970268

DynoCal - A Chassis Dynamometer Calibrator

Severino D'Angelo, Charles Brownell, and Curtis Brownell
Real Time Instruments

Garrett Torgerson
State of California Bureau of Automotive Repair

William G. Mears
Dynamotive Engineering

ABSTRACT

The DynoCal system for auditing chassis roll dynamometers has been developed for the California Bureau of Automotive Repair. It evaluates performance of new dynamometer designs and operational dynamometers in the field. The system includes a 50 kW electric motor with torque, thrust, and speed instrumentation. It functions as a vehicle to drive the chassis rolls under computer-controlled test modes. A trailer transports the system and houses the battery power supply, chargers, flux vector power converter, and computer. The controls are user-friendly, with graphic displays and high-speed data acquisition. Among other functions, this unique test instrument measures dynamometer road load and inertia simulation accuracy.

INTRODUCTION

DynoCal was developed to improve the State of California's new Inspection and Maintenance (I/M) program. This program requires many passenger cars and light trucks to be periodically tested on chassis dynamometers. The dynamometers for the program will come from several manufacturers. Each dynamometer design will undergo a procedure to verify that it meets the standards of the program. In addition, dynamometers being used in the field will be audited.

For verification that the dynamometers in the program comply with required operational accuracy and performance specifications, the California Bureau of Automotive Repair (BAR) contracted with Real Time Instruments to develop a dynamometer testing device.

The resulting DynoCal system provides a number of features that make it a very capable and flexible dynamometer test instrument:

- Travels in a van-towed trailer.
- Exploits AC flux-vector drive technology.
- Operates on rechargeable batteries.
- Provides automated test functions.
- Uses Integrated data acquisition.
- Tests dynamometer road load and inertia simulation accuracy.
- Monitors dynamometer simulation error in real time.
- Gives color graphical displays.
- Generates comprehensive reports.

BACKGROUND

Whenever a chassis roll dynamometer is used for vehicle testing, its performance accuracy can influence test results. For valid and repeatable test results the dynamometer must be accurately calibrated and stable, so it provides the loading conditions specified for the test being conducted.

Chassis dynamometer calibration and performance specifications normally include measurement of the roll surface speed and tractive

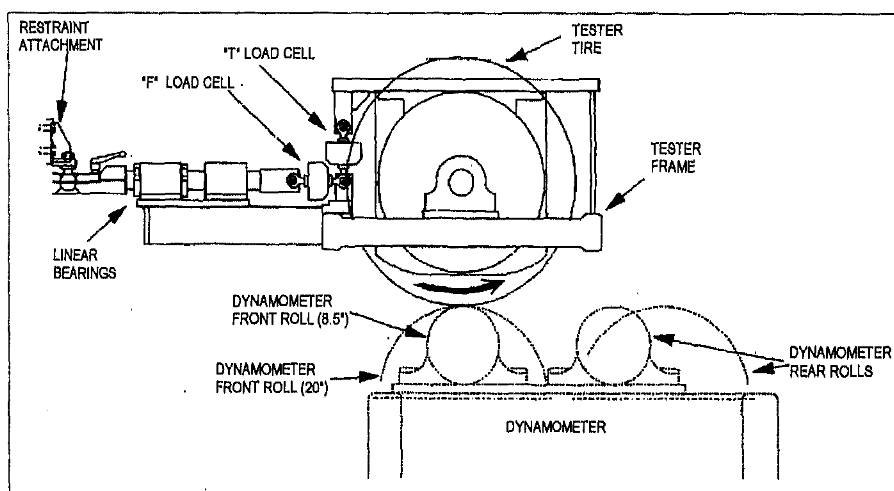
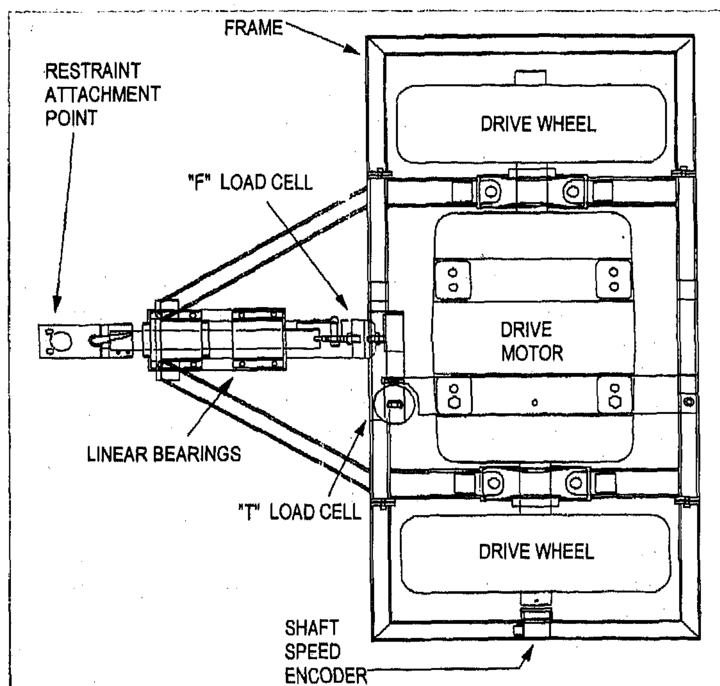


Figure 1. Side View in "F"-Mode



Fiaure 2. Toρ View Of Tester

force. Inertia simulation accuracy and response time of the load control system are also included in some dynamometer performance specifications.

Each dynamometer manufacturer interprets specifications and provides instrumentation and controls, with calibration instructions intended to demonstrate compliance with the specifications. This can result in a wide variation in methods and results, with possibly wide variation in outcome of the vehicle tests ultimately done on the dynamometers. For example, some dynamometers do not have the self-diagnostic capability needed to evaluate inertia simulation accuracy, leaving the user with the task of verifying accuracy.

Dynamometers which appear to be calibrated satisfactorily may not perform accurately when actually being driven by a vehicle. Attempts have been made to evaluate dynamometer performance using vehicles with torque wheels. Difficulties in calibration, control, and data interpretation have limited use of this approach.

To simplify the task of evaluating the many dynamometer designs being proposed and used for California consumer vehicle emission programs, the California Bureau of Automotive Repair (BAR) developed the concept of a system to test dynamometer calibration and performance by accurately measuring roll speed and tractive force independently of the dynamometer instrumentation. In this way, the tester would provide a transfer standard which could be used to compare dynamometer calibration and performance directly against specifications. This was opportune for Real Time Instruments, which has advocated standardized chassis dynamometer performance tests, [1[ ]2] and had, independently of the BAR concept, considered approaches to building a dynamometer calibration machine.

What is required to evaluate dynamometer performance independently of the instrumentation provided on the dynamometer? To answer this question, a brief description of how a dynamometer works will lead to the general approach applicable to all chassis dynamometers.

In the most general terms, a dynamometer is a device that measures mechanical power or that applies a controlled load to a mechanical power source. In the case of a chassis dynamometer, the vehicle's driving tires are the power source. The dynamometer is loading the vehicle correctly if the tangential force between the dynamometer roll surface and the vehicle's driving tires is exactly the correct value for the vehicle speed, acceleration rate, and dynamometer settings representing the vehicle inertia and road load friction. However, this tractive force cannot be measured directly when a vehicle is being driven. And when there is no vehicle on the roll, there is no tractive force at the roll surface. A "vehicle" capable of measuring tractive force and roll speed continuously and accurately under repeatable driving cycles could measure how well the dynamometer functions. By doing this the same way on all dynamometers, the tester could provide direct comparisons of dynamometer performance. DynoCal does this.

DESIGN OBJECTIVES

The design requirements from BAR can be summarized as follows:

- Capability to determine dynamometer:
  Speed and torque measuring accuracy.
  Load and inertia simulation accuracy during transient mode testing.
- Recording acquired speed, torque, and related data.
- Comprised of no more than two modules

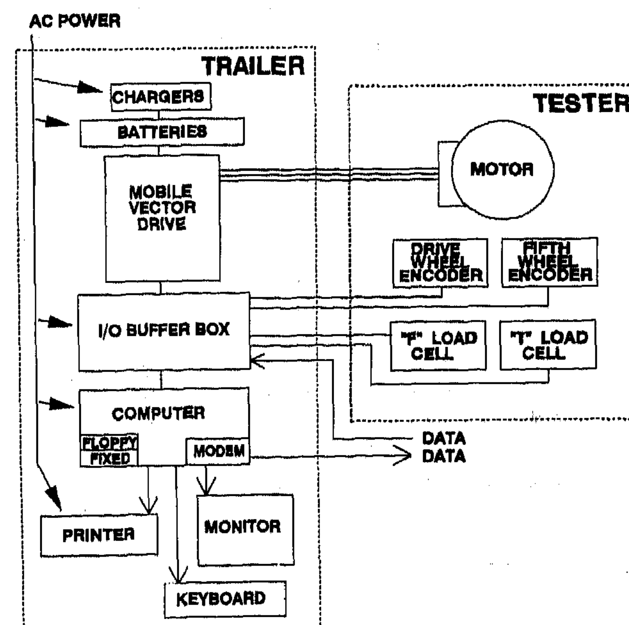


Figure 3. Schematic

- Transportable by minivan or equivalent vehicle.
- Speed and load determined independently from dynamometer sensors.
- Self-contained, with 25 hp available independent of external power, and 50 hp when connected to building power.
- Able to accommodate bi-directional dynamometers, in-ground to 24" above ground.
- Able to run ASM, BAR31, IM240 and LA92-S cycles on dynamometers simulating from 1500 to 6000 lb inertia up to 70 mph
- Instrumentation accuracy at limits of best dynamometer practice, traceable to NIST:
  - Load ±0.1% of full scale, 0.1 lb resolution
  - Speed, ±0.5% to 5 mph, then ±0.025 mph to 70 mph, 0.025 mph resolution
- Response time of torque response to step change command: 90% within 50 ms, 98% within 150 ms.
- Data acquisition:
  - Eight channels, analog (to 25 volts) or digital.
  - Selectable, high-speed sampling frequency.
- Computer, and software:
  - IBM-compatible, Pentium™ CPU
  - Minimum 16 Mb RAM
  - 1 Gb hard drive.
- Control of driving cycles, test sequences, data acquisition, data reduction, report generation, display and printing, data storage and transfer.

These were performance specifications, allowing the supplier maximum flexibility in designing the system. To develop a system to meet the BAR specifications, Real Time Instruments formed a team with expertise in dynamometer control, user application, equipment design, and fabrication, which had successfully completed earlier dynamometer projects.

DESCRIPTION

OVERVIEW. – Figure 1 shows a side view, and Figure 2 a top view of DynoCal, the device that is placed on the roll or rolls of a chassis dynamometer. The dynamometer is set up to simulate a given road load and inertia or other load such as ASM horsepower. DynoCal then "drives" the dynamometer rolls much like a vehicle. While DynoCal is driving the rolls, however, it measures the force being transmitted to the rolls, and the speed at which this force is being transmitted. Using this procedure, the DynoCal computer is able to determine how accurately the dynamometer is simulating the preset loads and inertia.

In addition to the basic driving test just described, DynoCal is capable of a variety of other procedures for determining the dynamometer calibration accuracy and performance.

The overall DynoCal system is illustrated in the form of a block diagram in Figure 3. The block labeled "TESTER" includes the components that are positioned on the dynamometer, such as the AC motor, wheels, force transducers, speed sensors, bearings, and frame. The block labeled "TRAILER" includes the batteries, charging system, computer, signal conditioners, data acquisition system, safety controls, operator station, power switches, and mobile vector drive. The entire system is housed inside the trailer for transport from site to site.

MAJOR COMPONENTS – Refer to Figures 1 and 2. The figures illustrate the locations of the motor and other major components. The tester is essentially a single-axle electric vehicle. The axle is the motor shaft, with a tire at each end carrying the weight of the tester. Bearings on the shaft support the tester frame.

The tester is transported inside a trailer, as shown in Figure 4. The trailer also houses all the other items required to operate the tester. When the tester is operated on a dynamometer, it can be attached either to a restraint mounted in the trailer, or to a floor-mounted fixture. Both restraint positions provide horizontal and vertical adjustment for DynoCal. These adjustments enable DynoCal to be positioned

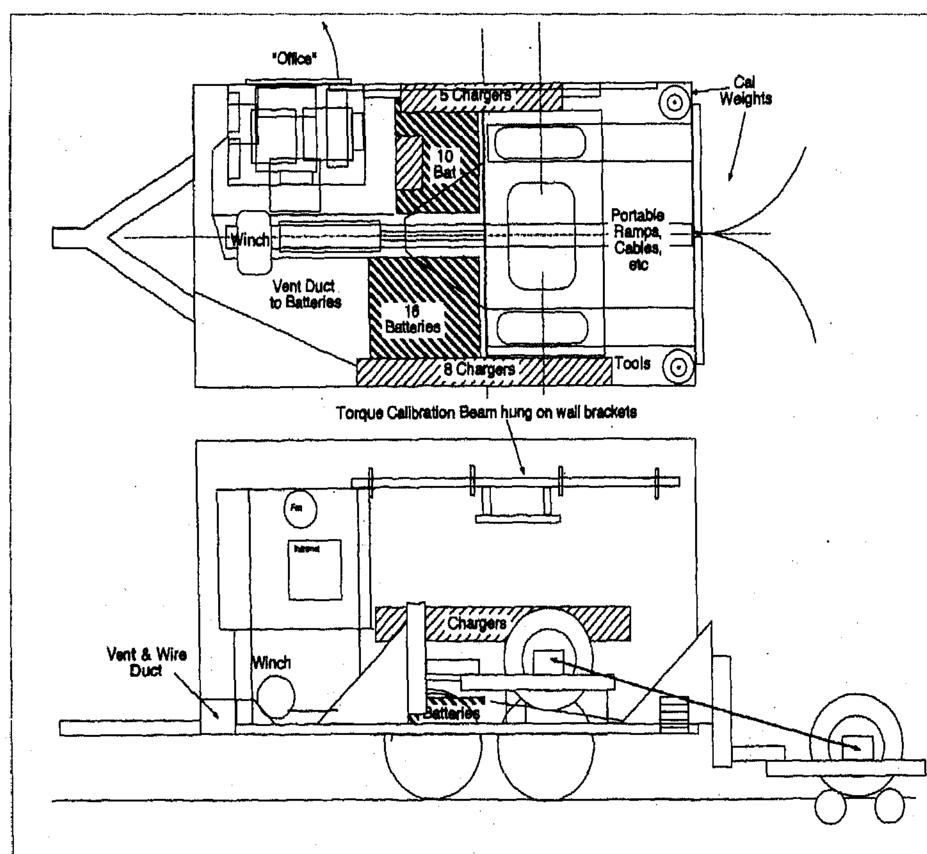


Figure 4. Layout

cradled between twin rolls, or on top of any given rolls. The two positions are needed to perform the tests described below.

LOAD MEASUREMENT – The tester employs two load cells (see Figure 5). One, termed the "Torque" load cell or "T" load cell, measures the force generated as a direct reaction to the torque produced by the AC motor. The other, called the "Force" or "F" load cell, measures the horizontal force generated in reaction to the wheels applying tractive force tangential to the dynamometer roll.

The tester can be used in either "T-mode" or "F-mode". In "T-mode", the measurement of tractive force is derived from the "T" load cell. In this mode, the tester driving wheels can be operated either in a cradled position between two dynamometer rolls as shown in Figure 5, or on top of the roll as shown in Figure 1. For the "F-mode", the tester driving wheels must be positioned exactly on top of the dynamometer roll. For twin-roll dynamometers this is normally the front roll because this is the roll that absorbs most of the thrust force from a test vehicle.

When the DynoCal wheels are cradled between front and rear rolls as a vehicle's driving wheels are, the force measured by the "F" load cell is no longer correct vehicle thrust. It also includes various force components produced by the cradling of the tires. In the cradled position, a hard acceleration tends to cause the wheel to climb the front roll, and a deceleration tends to cause it to climb the rear roll. Both of these actions result in errors in measuring true tractive force by the "F" load cell. The "T" load cell measures motor torque the same for both modes, but tire losses will be different when cradled and when running on top of one roll.

In "F-mode", if the tester wheels are not exactly on top of the roll, the curvature of the roll generates a horizontal force component from the tester's own weight. This is an artificial horizontal force not resulting directly from the thrust of the tester on the rolls, and leads to measurement error by the "F" load cell. During "F-mode" setup the wheel position is set accurately with alignment instruments, and maintained in this position by the rigid floor-mounted restraint. "F" load cell measurements are compensated for force shifts resulting from the slight deflection of the restraint produced by tractive force. Deflection vs. tractive force was calibrated by measuring movement of the DynoCal frame as the tester was operated on a dynamometer at a range of force levels. The correction is calculated for each operational force data point by multiplying the force measured by the deflection calculated for this force, and dividing by the sum of the tire and roll radius. To verify this correction, the tester was moved across the top of the roll and the change in apparent tractive force compared to measured off-center movement.

SPEED MEASUREMENT – The tester also has two speed-sensing encoders. The encoders are optical, digital devices. They are immune to drift and do not require calibration. One is mounted to the motor shaft and measures the rotational speed of the tester motor. The other is mounted to a "fifth wheel" which has a thin rubber "tire" running on the surface of the dynamometer roll. The first encoder (the "shaft speed encoder") is required by the flux vector drive to control the motor. The second encoder (the "fifth wheel encoder") is calibrated to measure the true speed of the dynamometer roll surface, which is required for accurate dynamometer performance measurement.

ELECTRICAL DESCRIPTION – DynoCal drives on a dynamometer as an electrical vehicle would. Much of the time, the tester expends energy and the dynamometer absorbs it. The energy expended by DynoCal comes from 26 12-Volt lead-acid gel-cell batteries wired in series to provide a nominal voltage of 312 VDC. Between the batteries and the AC motor that drives DynoCal is an AC flux vector power converter. This power converter is called a "Mobile Vector Drive"

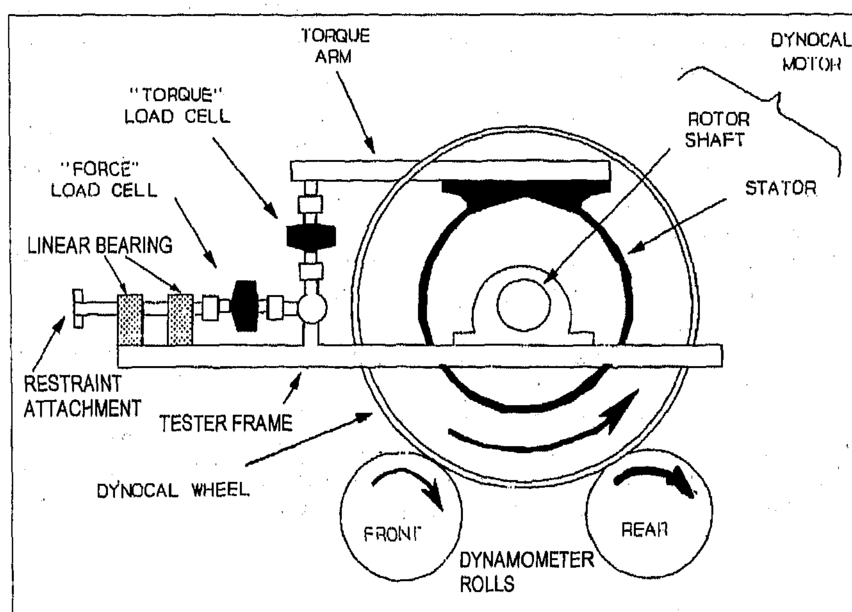


Figure 5. Load Measurement

(MVD), originally designed for electric vehicle use. It generates the appropriate 3-phase voltage and frequency to produce the motor torque required to control DynoCal operation.

The power converter in many electric vehicles can charge the batteries by absorbing power from the dynamometer during deceleration. DynoCal uses the MVD in the same way. Most of the time, however, DynoCal draws energy from the batteries, and a charging system is required to replenish battery energy.

Sealed gel-cell lead-acid batteries were selected for two primary reasons: safety, and their ability to accept a rapid charge. Some of the notable features of the batteries are:

OCT 10-96 LOAD CELL CALIBRATIONS 09:03:24

SEP 19-96 11:22:02

RECORD #7

		T-LOAD CELL	0.00 lbs	F-LOAD CELL	0.00 lbs
		OFFSET	11.4363	OFFSET	-19.932
		+GAIN	14.6329	+GAIN	13.7997
		-GAIN	14.5336	-GAIN	13.8375
Nothing	0.00	-0.02	-0.36	-0.00	-0.12
10 wts	689.14	689.13	-689.18	689.15	-689.13
9 wts	620.57	620.81	-620.47	620.41	-620.59
8 wts	552.00	551.68	-551.78	552.10	-552.01
7 wts	483.43	483.98	-483.29	483.64	-483.61
6 wts	414.86	415.49	-414.63	414.84	-414.67
5 wts	346.29	346.88	-346.30	346.42	-346.08
4 wts	277.71	278.09	-277.80	277.82	-277.91
3 wts	209.14	209.33	-209.24	209.19	-209.07
2 wts	140.57	140.61	-140.70	140.54	-140.31
1 wts	72.00	71.84	-72.28	71.91	-71.80
Hanger	3.43	3.15	-3.69	3.36	-3.26

Esc Exit Home First End Last Page Up Previous Page Down Next

Figure 6. Load Cell Calibration

- High volumetric and gravimetric power densities at high rates of discharge (provides more power in less space).
- Thin plates (provides high-rate discharge capability).
- Low internal resistance (better terminal voltage characteristics under high-rate discharges).
- Excellent high-rate recharge capability (>90% recharge in one hour when enough charging power is available).
- Negligible gassing.
- Maintenance-free.
- Non-spillable (can be mounted in any orientation; can be transported by air or ground without restriction).
- Excellent storage capabilities (long shelf life).
- Wide operating temperature range.

The batteries can be recharged either by the battery chargers provided or by motoring the dynamometer with DynoCal on the rolls and using the DynoCal AC motor as a generator. Each pair of batteries is charged by a separate 24 VDC charger designed for gel cell batteries.

CONTROLS – The MVD is controlled in torque mode via a ± 10 Volt analog signal generated by the computer and fed through the I/O Buffer Box. This box houses a multifunction printed circuit board and has several push-buttons and LED's on its front panel. It also receives outputs from the load cells and encoders and converts them to a format that is usable by the computer. In addition, the I/O Buffer Box provides *START*, *STOP*, and *QUICK STOP* buttons.

The computer includes a keyboard and 14" SVGA monitor to provide interfaces for the operator. It also has a 1 Gb hard drive for storing operational and data files, a modem for sending out data and reports by telephone line, and a printer for generating reports locally. The computer is IBM-compatible with an Intel Pentium™ microprocessor, 16 Mb of RAM, and a 3 1/5" floppy drive.

It is the software in the control computer that permits DynoCal to perform its many automated functions. These include calibrations, parameter setting, data logging, dynamometer checks, drive trace storage, report generation, and other functions.

FUNCTIONAL DESCRIPTION

DynoCal acts as a substitute for a vehicle on a chassis dynamometer. Instead of a human driver watching a driver's aid and controlling vehicle speed (and in turn, roll speed) with the throttle, DynoCal controls its own wheel speed (and in turn, roll speed) directly via its AC motor. DynoCal follows a drive trace that is loaded onto the hard drive in the form of an ASCII file. It can be created and modified by the user, providing flexibility in the testing program. If desired, the drive trace can be constructed to cause DynoCal to mimic a human driver (by imitating the anticipation abilities and the response time of an actual driver). But unlike a vehicle, DynoCal accurately measures and records the thrust force it applies to the dynamometer and the speed at which the force is applied. Based on this information, DynoCal is able to determine how accurately the dynamometer is simulating road load and inertia.

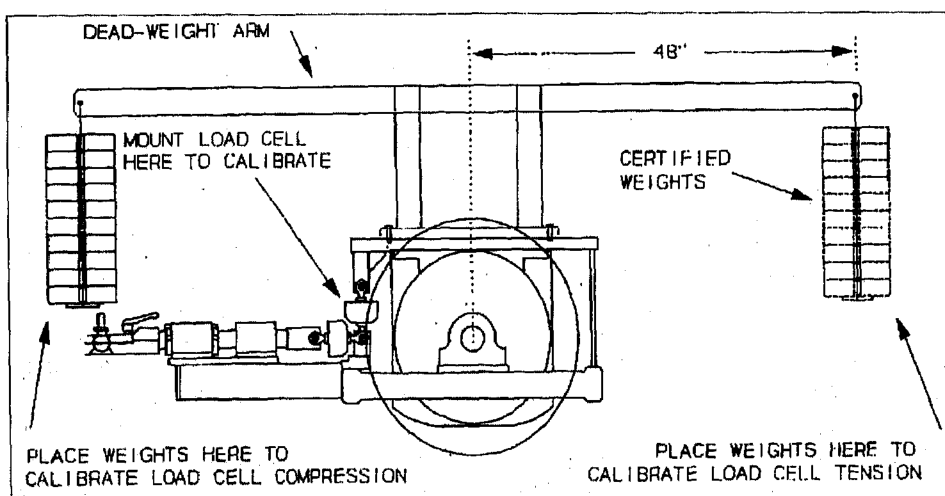


Figure 7. Load Calibration Method

In "T-mode", the tractive force being applied by DynoCal to the dynamometer is calculated by subtracting DynoCal parasitic losses, inertia acceleration losses, and tire losses from the force measured by the "T" load cell. In "F-mode", the tractive force is obtained from the "F" load cell with compensation for roll surface curvature effects.

The following is a summary of the principal DynoCal functions accessible through the computer. The first six functions form the foundation for DynoCal data measurement accuracy.

- TIMEBASE ACCURACY
- LOAD CELL CALIBRATION
- SPEED ACCURACY AND CALIBRATION
- INERTIA MEASUREMENT
- MEASURE TIRE LOSSES
- RESPONSE TIME RECORD
- DRIVING CYCLE

Dynamometer calibration and performance are also measured using all but the TIMEBASE ACCURACY and LOAD CELL CALIBRATION functions.

TIMEBASE ACCURACY – The basic 10 ms control cycle time is measured for 0.05% accuracy either by stopwatch over an extended time interval or by an NIST-traceable frequency counter

LOAD CELL CALIBRATION – The "T" and the "F" load cells are both calibrated by dead-weighting. A screen is dedicated to facilitating the calibrations. Both load cells and their signal conditioners are calibrated in the "T" load cell location to 0.1% accuracy using a balanced 8 ft (4 ft on each side) checking arm and NIST traceable weights covering the full range of operation. Figure 6 is a typical torque calibration screen and Figure 7 illustrates the torque calibration method.

SPEED ACCURACY AND CALIBRATION – DynoCal also has two speed measuring devices, one for measuring dynamometer roll speed (the "fifth wheel") and one for measuring DynoCal drive wheel speed. A screen, Figure 8, is assigned for making these calibrations. and provides for manual input of the dynamometer indicated speed for accuracy determination. The fifth wheel is calibrated by using DynoCal to motor a dynamometer at constant speed with no dynamometer load. Before starting the calibration, the dynamometer diameter is measured with a "Pi" tape accurate to 0.001", and traceable to NIST. Then, as the dynamometer roll is motored at constant speed, the roll rpm is measured with a stroboscope with 0.01% timing accuracy. The fifth wheel calibration is set from the measured rpm and roll diameter. Then the driving wheel speed is calibrated to be equal to the fifth wheel speed by adjusting the effective drive wheel radius.

DEC 05-96 12:49:24

SPEED ACCURACY AND CALIBRATION

	Diameter	Speed
Dyno Roll:	8.580 in	0.0 rpm**
5th-wheel:	11.529 in	0.00 mph
Drive wheel:	24.34 in	0.00 mph

Highlight and press ENTER to set speed	Wait to settle and enter displayed dyno speed	Dyno speed error	5th wheel speed	Drive wheel speed
10 mph	10.09	0.01	10.08	10.15
20 mph	20.00	-0.11*	20.11	20.25
30 mph	30.10	-0.10	30.20	30.39
40 mph	40.00	-0.19*	40.19	40.46
50 mph	0.00			
60 mph	0.00			
70 mph	0.00			

*exceeds specification of 0.10 mph
**if roll rpm does not match strobe value, DynoCal diameters can be adjusted by entering the value displayed on the strobe here: 988.0

Alt-X Exit F1 Stop rolls

Figure 8. Speed Calibration

INERTIA MEASUREMENT – DynoCal is able to determine its own mechanical inertia, the mechanical inertia of the dynamometer, and how much inertia the dynamometer is actually simulating. Figure 9 gives a typical inertia measurement screen. To calibrate the DynoCal base mechanical inertia, DynoCal is lifted off the rolls and then accelerated and decelerated over a range of measured constant torque with simultaneous speed-time and acceleration measurements. The base inertia is determined for each accel/decel pair. The average base inertia and standard deviation are computed. A standard deviation less than 4 lb is required for the DynoCal calibration. Dynamometer inertia is measured in the same way, but with wheels on the rolls and speed input from the fifth wheel.

MEASURE TIRE LOSSES – Whenever DynoCal is used on a dynamometer in "T-mode", its tire losses

REALTIME INERTIA MEASUREMENT

DEC 09-96 Printing 15:39:42

RECORD #13

	SET	MEASURED	
INERTIA	2000	2095	lbs
ERROR	40	95	lbs
STD DEV		2	

	ACCEL					DECEL					
SET FORCE lbs	TIME sec	WHEEL FORCE lbs	ROLL FORCE lbs	WHEEL ACCEL mph/s	ROLL ACCEL mph/s	TIME sec	WHEEL FORCE lbs	ROLL FORCE lbs	WHEEL DECEL mph/s	ROLL DECEL mph/s	INERTIA lbs
200	18.44	247.7	199.6	1.90	1.89	15.42	169.1	200.9	2.27	2.30	2097.2
250	14.72	292.9	243.0	2.38	2.35	12.64	219.5	249.4	2.77	2.81	2092.8
300	12.22	340.4	289.0	2.86	2.83	10.76	269.3	297.2	3.25	3.31	2093.8
300	12.22	340.0	288.6	2.86	2.83	10.76	269.5	297.3	3.25	3.31	2094.3
250	14.67	293.2	243.3	2.39	2.36	12.69	219.6	249.4	2.76	2.80	2095.0
200	18.37	246.4	198.2	1.90	1.89	15.54	169.0	200.8	2.25	2.29	2096.2

LOW SPEED 15 mph
HIGH SPEED 50 mph
FACILITY BAR
REMARKS
MEASURE DECEL YES
LANE

Esc Exit Page Up Previous Page Down Next ALT-P Print ALT-F File F2 Remarks

Figure 9. Inertia Calibration

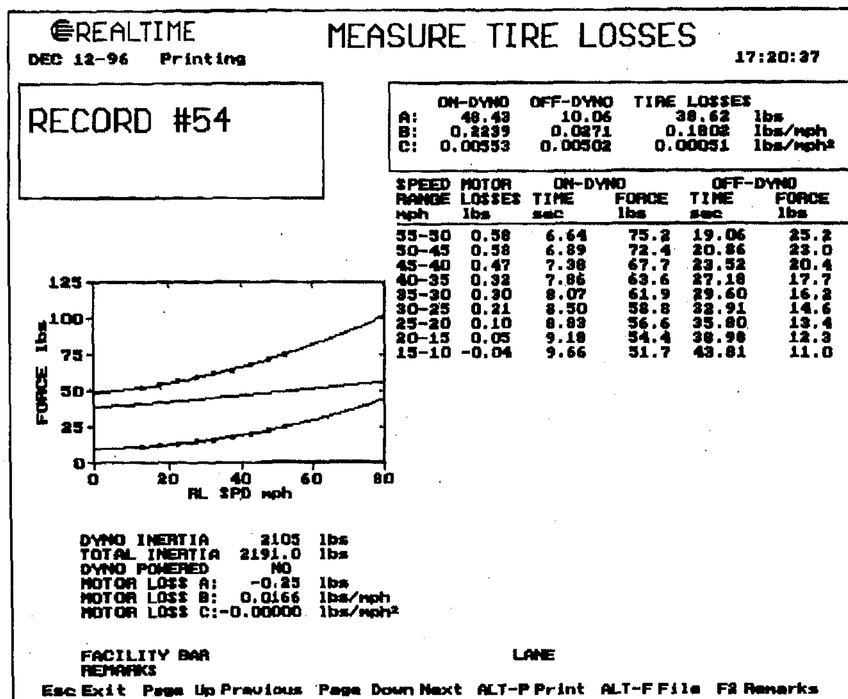


Figure 10. Tire Losses

are used in the calculations. To ensure that the values used in the calculations are correct, a function is provided to measure tire losses. Figure 10 is a typical tire loss screen. After warming up the tires and the dynamometer, a coastdown is run with the dynamometer de-energized and Dynocal on the rolls, also de-energized during the coastdown. Then the dynamometer is run through the same de-energized coastdown, but with DynoCal lifted off the rolls. The fifth wheel is used for dynamometer coastdown speed measurement in both cases. The DynoCal "T" load cell is also recorded during the wheel-on coastdown to derive the parasitic losses of the motor bearings. As shown on the MEASURE TIRE LOSSES screen, tire losses are calculated by subtracting the wheels-off coastdown and motor bearing parasitic loss coefficients from the wheels-on coastdown coefficients. The OFF-DYNO coefficients describe the dynamometer parasitic friction.

RESPONSE TIME RECORD – DynoCal uses its fifth wheel to measure the rate of acceleration or deceleration following a rapid change in DynoCal load, which is also measured. Figures 11 shows a typical screen of response time. The control drives DynoCal at a constant load and instantaneously commands a constant acceleration or deceleration load. On the RESPONSE TIME screen, acceleration, load, and time are displayed for a 4-second interval beginning just before the load change. The display can then be inspected to determine the amount of delay in dynamometer response.

DYNO LOADING ACCURACY – DynoCal measures the tractive force it is applying to the rolls while the dynamometer is simulating road load at steady speed. Inertia sumulation is not measured in this test. Figure 12 shows a typical road load accuracy test screen.

SIMULATION ERROR – During a driving cycle, Figure 13, DynoCal compares the actual thrust that it applies to the rolls to the thrust that should have been applied based on the road load and inertia settings entered into the dynamometer control. From this, simulation error is calculated according to the methods described in Reference 2.

DATA LOGGING CONFIGURATION – The computer logs up to eight channels of data on its hard drive while carrying out its standard test functions. These data are used by the computer to generate its reports. The data are also available for export to a spreadsheet or statistics package for analysis.

PRINTING REPORTS – The results of the various tests are collected in the form of reports which may be printed out.

TRANSMITTING FILES VIA MODEM – The computer includes a modem card that permits data files and report files to be sent over the phone lines.

TEST RESULTS

Testing on new and in-service dynamometers has validated the need and value of DynoCal.

Much of the initial testing has been part of evaluating the DynoCal software and control functions, and learning how to interpret test results. The ability to measure tire losses has been used to investigate tire warmup and develop the relationship between loss characteristics and tread temperature during warmup. This was done to provide a method of correcting for tire losses when DynoCal is used to evaluate a dynamometer which has been cold-soaked, and must be tested as soon as the insulated enclosure is removed. The MEASURE TIRE LOSSES function, Figure 10, was used for this study, and the resulting loss coefficients analyzed in a spreadsheet. The tread temperature for each loss determination was measured with an infrared pyrometer. Figure 14 shows the tire losses over the operating speed range for each measured temperature, and Figure 15 shows the losses expressed for the 15 and 25 mph speeds used for the two ASM cycles. The smooth plots, Poly. (25) and (15) in Figure 15, are from the equation, for comparison with the plots from the data curves on Figure 14. These results offer promise of giving excellent tire loss correction for accurate testing in the "T"-mode during dynamometer warm up.

Some other results from this early testing are summarized. DYNO LOADING ACCURACY has already found a number of dynamometer control errors. Figure 12 shows how the dynamometer overloads 34 lb when in the test mode. This error was not evident during calibration, and was traced to incorrect dynamometer parasitic friction compensation. The SPEED ACCURACY AND CALIBRATION screen, Figure 8, indicates a type of operational speed error observed. INERTIA MEASUREMENT, Figure 9, has been found to be a

quick and repeatable method for accurately determining true mechanical inertia of the assembled dynamometer. RESPONSE TIME, Figure 11, reveals the actual response of the dynamometer load control to a change in vehicle acceleration rate. This figure shows how the dynamometer load takes about 0.1 second to begin adjusting to the roll acceleration, and then takes about 0.9 seconds to reach required load.

Inertia simulation has been found to require very accurate tire loss compensation when running in the "T"-mode. In the "F"-mode, the load zero has a major influence on correlation with the "T"-mode. Obtaining a valid zero has been difficult, and may be influenced by the tendency of DynoCal to shift sideways slightly on rolls which are threaded for increased traction.

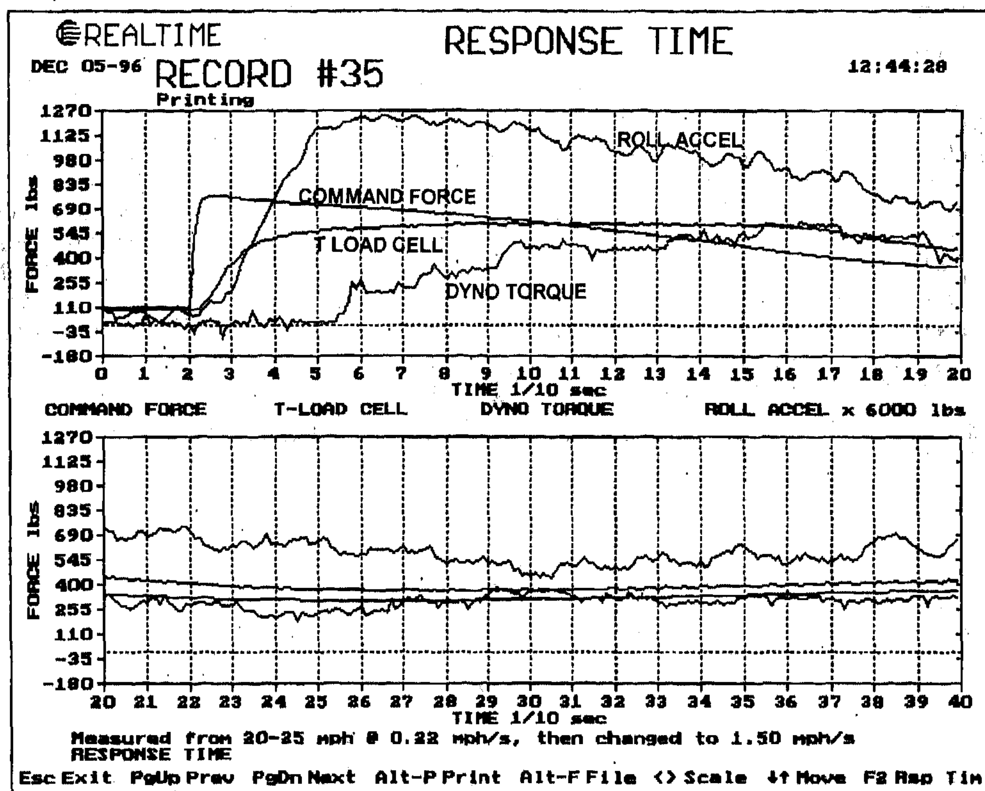


Figure 11. Dynamic Response

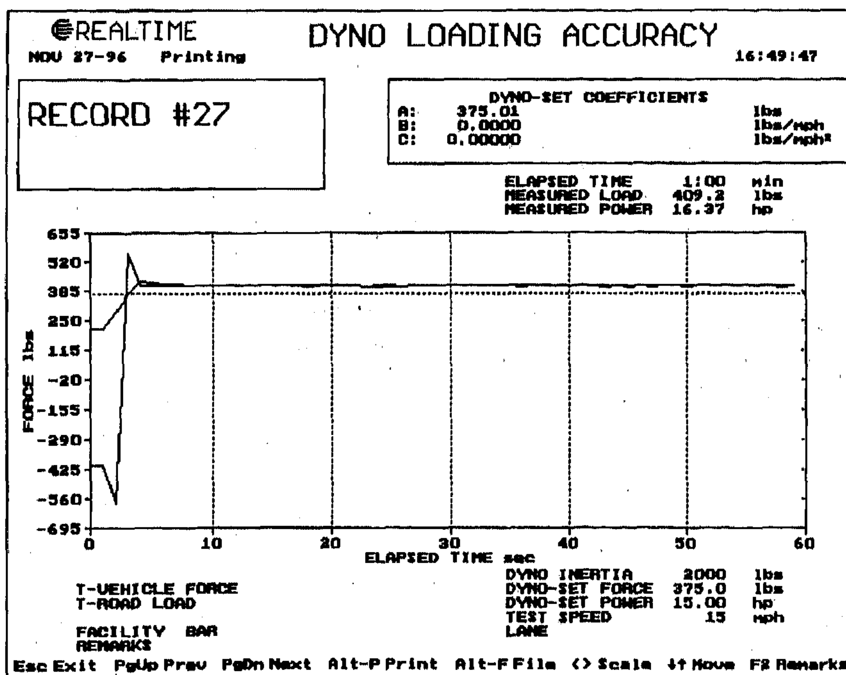


Figure 12. Road Load Accuracy

DynoCal has already shown benefits in detecting the following problems and performance characteristics in dynamometers evaluated by BAR:

- Discovery of incorrect load calibration.
- Discovery of speed-load measurement interference.
- Load command errors.
- Significant differences in inertia simulation response time.
- Non-linear speed measurement.
- Slow settling time for constant load control.
- Display of actual operation of augmented braking.

In addition, testing experience has resulted in modification to the automated test functions to accommodate inertia simulation measurement on dynamometers while augmented braking occurs on deceleration, and where eddy-current absorbers simulate inertia only during acceleration.

CONCLUSIONS

DynoCal is an innovative dynamometer testing instrument. It has speed and load measurment accuracy not previously available. The powerful, fast-responding, computer-controlled motor provides measurement of dynamometer response time, inertia simulation, road load simulation, and other critical dynamometer functions under the same operating mode as used during vehicle testing. As a result, the following conclusions are drawn:

1. A practical dynamometer tester has been developed.
2. Design objectives have been achieved.
3. Test results verify the need for and usefulness of the tester.

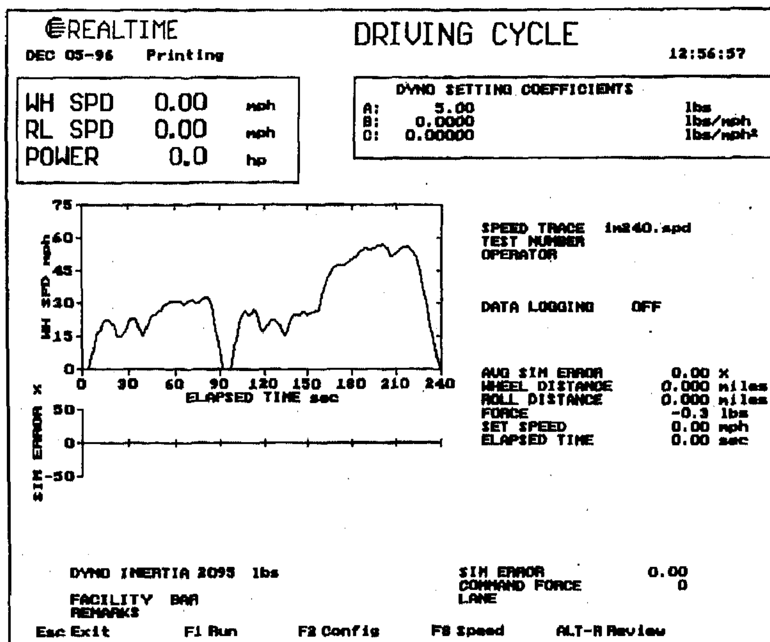


Figure 13. Simulation Error

ACKNOWLEDGMENTS

Successful completion of the DynoCal project was made possible by the dedication and skill of many contributors. Especially notable are: David Amlin, Debby Berardino, Scott Counts, Barbara D'Angelo, Duane Dann, Chuck Kizlauskas, Bill Murphy, and Gary Trudeau.

REFERENCES

1. S. D'Angelo, Charles Brownell, Curtis Brownell, and W. G. Mears Evaluating the Performance of Chassis Dynamometers with Electric Inertia Simulation. SAE Technical Paper No. 960716, 1996.
2. W.G. Mears, S. D'Angelo, and C.D. Pausell, Performance Tests of a Large-Roll Chassis Dynamometer with AC Flux-Vector PEU and Friction-Compensated Bearings. SAE Technical Paper No. 930392, 1993.

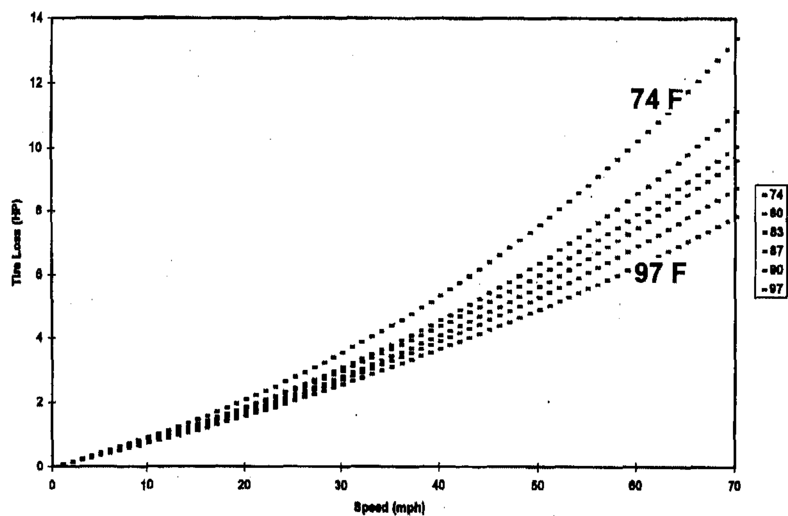


Figure 14. Tire Loss vs. Tread Temperature

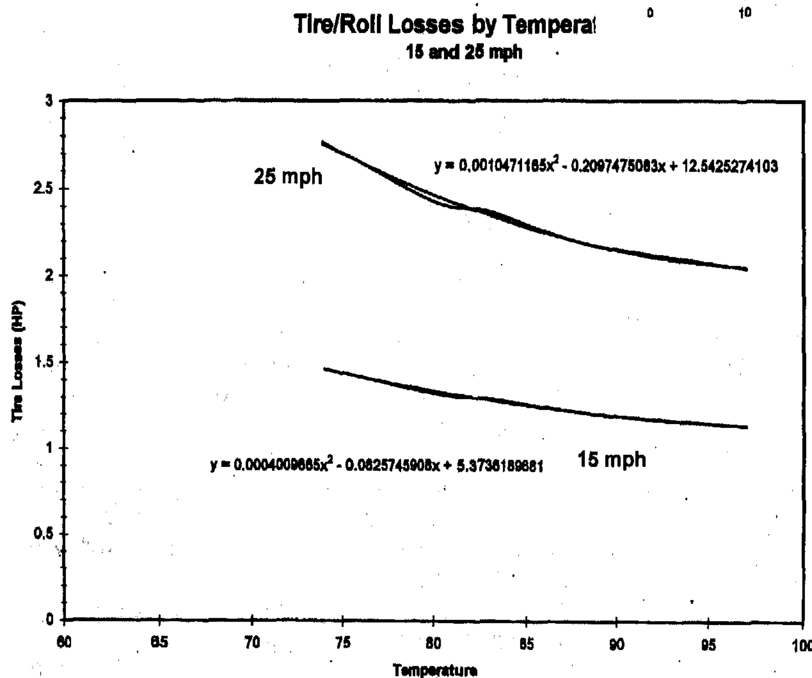


Figure 15. Loss vs. Temp at 15 & 25 MPH

APPENDIX A

DynoCal SPECIFICATIONS

GENERAL RATINGS

Continuous power (30 min)	67 hp at 47-70 mph
	35 hp at 25 mph
	21 hp at 15 mph
Continuous force at wheel surface	534 lb-force
Peak force at wheel surface	664 lb-force
Maximum Speed	70 mph

MECHANICAL EQUIPMENT

Drive wheels	Automobile tires
Wheel preparation	Tread removed
DynoCal width (outside of wheels)	53 in
DynoCal width (inside of wheels)	38 in
DynoCal width (outside of frame)	65 in
Bearing type	Spherical roller
Bearing life	50,000 hr
Mechanical inertia (nominal)	75 lb

ACCURACY/RESOLUTION

Time base accuracy	± 0.05%
Force measurement accuracy	± 0.1% of full scale
Speed measurement accuracy	± 0.05 mph
Inertia repeatability, std. dev. of 10 runs	≤4 lb
Response ± 2% of commanded torque	≤150 ms

TORQUE SENSORS

"T" and "F" load cells	Same
Transducer type	Strain-gauge load cell
Range	± 1,000 lb
Measuring system accuracy	± 0.1% of full scale
Torque arm length	14.000 in
Calibration arm length	48.000 in

SPEED MEASUREMENT

FIFTH WHEEL	
Nominal diameter	11.459 in
Encoder transducer type	Optical encoder, quadrature
Resolution	0.004 mph
Encoder Lines	12,500 direct
Pulses	50,000 per rev (with quadrature)
DRIVE WHEELS	
Nominal diameter	27 in
Encoder transducer type	Optical encoder
Resolution	0.08mph
Encoder Lines	1,024 per rev
Pulses	1,024 per rev

COMPUTER SPECIFICATIONS

CPU	Intel Pentium™ 133 MHz
RAM	16 Mb
Hard Drive	1 Gb
A/D Converter Resolution	16 bits
Serial Communication	Two RS232C ports

UTILITIES AND POWER

DynoCal motor power	Batteries, 312 VDC nominal
Chargers	115/230 V, single-phase AC
Computer and I/O Buffer Box	115 V, 5 A, clean power
Shop Air	Not required
DynoCal motor cooling	Not required
Battery charger cooling	Self-contained fans

BATTERIES

Type	Sealed lead-acid gel cell
Capacity (10 hr rate)	38 Amp-hr
Dimensions	6.5" long, 8" wide, 7 " high
Weight (each)	35 lb
Number used	26
Cycle life at 25°C:	
15 min rate	300 for 80% DOD
3-5 hr rate	500 for 80% DOD
Operating temperature range	-40°C to +60°C

BATTERY CHARGERS

Nominal output	24 VDC, 8 A
Input power	230/115 V, 1.5/3.6 A, 50-60 Hz
Weight (each)	16 lb
Number used	13
Approx. total charging time from 50% discharge	5-6 hr

DYNOCAL MOTOR NAMEPLATE DATA

NEMA frame size	405TZ
Horsepower	67
Volts	240 V
Speed	585 rpm
Frequency	30 Hz
Temperature rating	40°C above ambient (30 min)

DIMENSIONS AND WEIGHTS

Trailer dimensions (nominal)	7 ft by 12 ft
Trailer outside dimensions	188" L, 94" W, 103" H
DynoCal tester weight	1,900 lb
Trailer weight without tester	3,800 lb
Trailer weight with tester	5,700 lb

ENGINEERING CENTER

Real Time Instruments, Inc.
22622 Lambert St., #306
Lake Forest, CA 92630
Tel: 714-452-0660

MANUFACTURING CENTER

KLC Enterprises, Inc.
4765 Holland Rd.
Saginaw, MI 48601
Tel: 517-753-0496

970277

The Effect of Driver Age on Traffic Accidents

George Z. Libertiny
Design and Manufacturing Defect Consultants

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

There have been numerous publications on the frequency and severity of traffic accidents as effected by age. This is a recent review of the current findings.

While very old drivers have increased accident rates, the impact of their accident rate on the society is small, because they drive very short distances and thus the total number of accidents in which they are involved is low. To improve the traffic accident situation the main effort has to aim the young drivers. Due to their alcohol consumption prior to driving, their reluctance to wear seat belts and the high mileage they drive, they are one of the main causes of the high accident rates. Driver education programs aimed to improve driving skill have minimal effect on accident rates. However, driver education and law enforcement programs aimed at reducing the consumption of alcohol prior to driving and encourage the usage of seat belts could contribute to a decrease in accident rates.

INTRODUCTION

Motor vehicles are integral parts of our lives. They provide mobility for much of our society. At the same time they create the potential for injuries and death. Young people are particularly affected by motor vehicle related accidents. Half of the 19 year olds who die, die because of traffic accidents. That is, there is approximately a 50 percent chance that the death of a 19 year old will be caused by a traffic accident. This probability for traffic related death does decrease with age, as shown on Figure 1[Ref. 1]. In fact all the major factors affecting the frequency and severity of traffic accidents change with age. Some of them have positive effects (e.g., increase in seat belt usage, less alcohol consumption by drivers) and other have negative effects (i.e., decreased reaction time, increased consequence of impact). The types of accidents also change with age. Consequently the effect of safety regulations and traffic related laws have different results for different age groups. For example, older people are almost never involved in rollover accidents. Consequently attempts

(and funds) aimed to make vehicles less likely to rollover will not have any effect on the safety of older drivers. On the other hand older drivers have accidents more frequently because they do not notice oncoming vehicles when changing lanes. Thus, for older drivers, it is more important to have better rear view mirrors than to have an increase in rollover protection. It is necessary to know what is causing accidents in various age groups in order to make the most effective safety decisions.

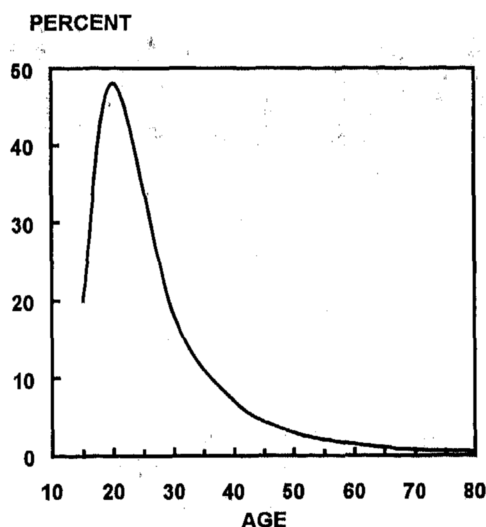


Figure 1. Probability that a given death is due to motor vehicle crash

The purpose of this paper is to review available accident data in order to establish how age effects the frequency and severity of traffic accidents.

DISCUSSION

Effect of Age and Gender on Fatality Risk - The consequence of traffic accidents increases with age. Physical changes with age cause the body to be more susceptible to serious injury in an impact. The risk of bone fractures is particularly increased. The "fragility" effect is particularly noticeable above the age of 60. In general, there is broad evidence of an increasing trend in the risk of injury and fatality with age [Ref. 2]. In a study by Oreskovich et al. [Ref. 3], it was found that 92 percent of people over 70 years of age who suffer severe trauma injuries never return to their previous level of independence. Evans [Ref. 4] found that the fatality risk significantly increases with age.

The results of Evans' study are supported by a recent NHTSA publication [Ref. 5] including the rate of injuries and fatalities at various ages. A comparison between injuries and fatalities is given in Figure 2 [Ref. 5].

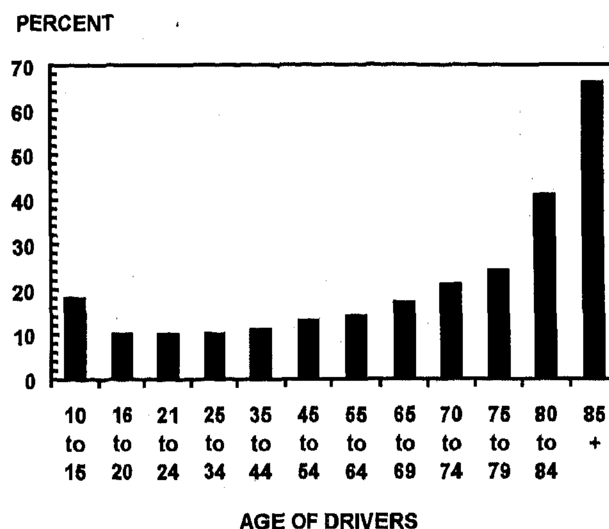


Figure 2. Fatalities as a percent of traffic injuries

Evans [Ref. 1] found that fatality risk due to a given impact is about 25 percent greater for females than for similar aged males from about age 15 to 45 years. The very young and very old males are most at risk. For both sexes, fatality risk is least at age 20; at age 70 it is about three times what it was at age 20.

On the positive side, older people tend to compensate for their increased risk of injury; they do so by more frequently using large vehicles which significantly decrease injury severity, by not consuming alcohol prior to driving, by using seat belts, and by driving less miles.

Effect of Age, Gender and Race on Vehicle Crashes - A North Carolina study [Ref. 6] indicated that the crash rate is lower for females when compared to males. The crash rate per licensed drivers decreases with

age. While young white and nonwhite drivers have approximately the same crash rates, a difference develops with age. The crash rate for nonwhite males is half at age 65 than of the 16-24 age rate. The decrease of crash rates with age is significantly larger for white males than for black males. Females have fewer crashes than males. Non-white females have more crashes than white females. These results are shown in Figure 3.

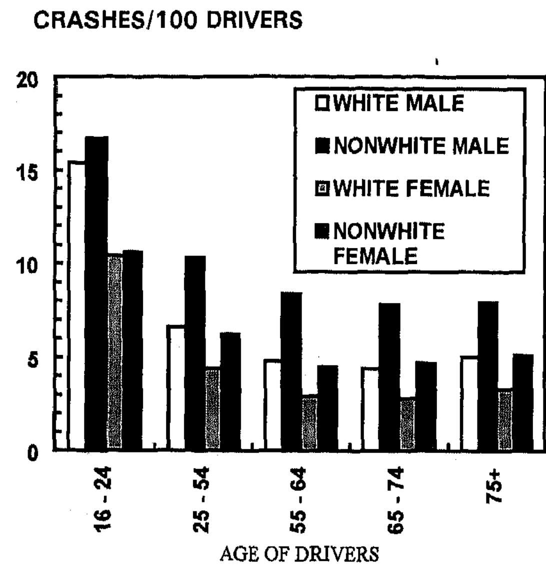


Figure 3. North Carolina Vehicle Crashes per 100 licensed drivers; in 1988.

The problem with the above chart is that it does not allow for the effects of mileage. When the same data is expressed as per estimated million miles [Ref. 6] the differences between the genders and races decreased a great deal. The effect of age also changed, as shown on Figure 4.

Recent data on fatalities in the U.S. [Ref. 7] is shown in Figures 5 and 6. As shown, the present fatality trend in the U.S. is similar to the crash trend in North Carolina at the time of that study. It is interesting to note, that while the fatality rate for the very old drivers increases with age. To put it into perspective, the fatality rate of the 80 to 85 years age group (Figure 6) is similar to the fatality rate of the 16 to 19 years age group. Furthermore, the loss of life years caused by the accidents by elderly drivers is negligibly small when compared to the loss of life years of young drivers. Consequently, the significance of the higher fatality rate of the elderly drivers for the society is an order smaller than the significance of the fatality rate of the young drivers.

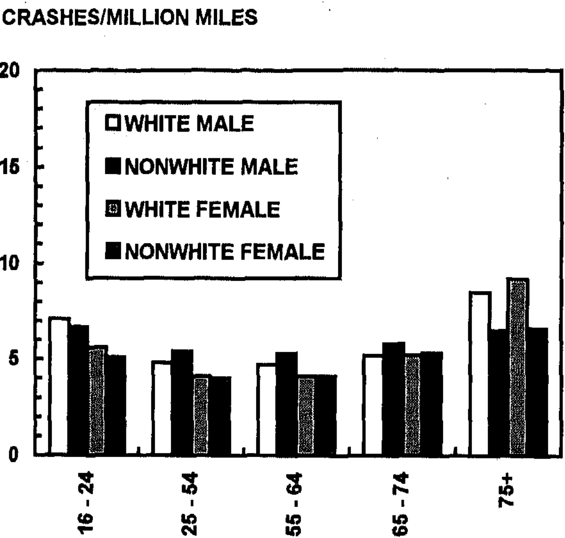


Figure 4. North Carolina Vehicle Crashes per estimated million vehicle miles; in 1988.

Fatality rates of female drivers are usually lower than male drivers at all ages. The difference is not as prominent, however, when the fatality rate is expressed as fatalities per miles compared to when it is expressed as fatalities per population (or drivers), as shown in Figures 5 and 6. From these figures it is apparent that one of the reasons for the lower female fatality rates is due to the fact that females drive less total miles than males do.

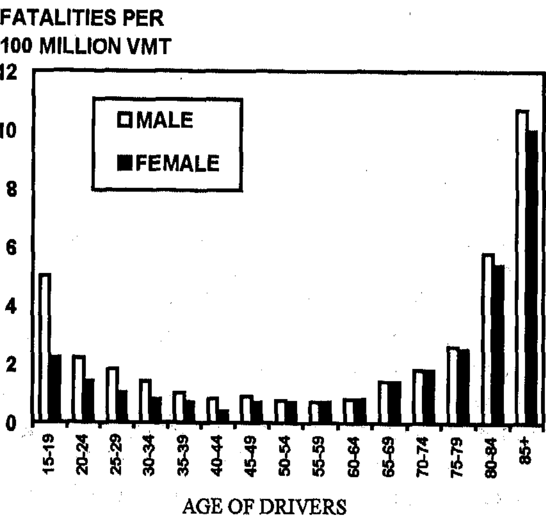


Figure 5. Fatality per 100 million VMT in the U.S.

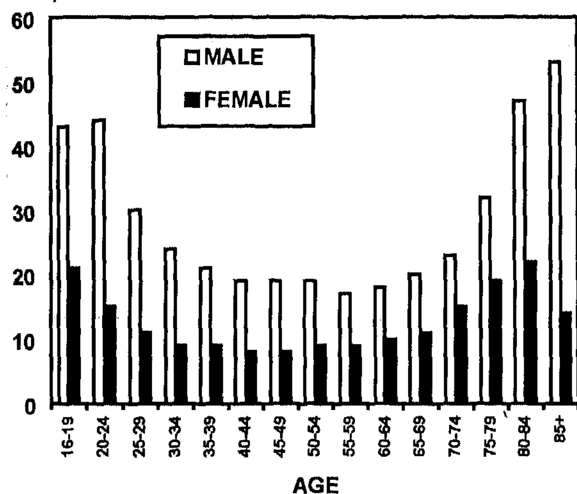


Figure 6. Fatality per 100,000 people in the U.S.

The effect of night driving and age on crash rates - The North Carolina study [Ref. 6] indicates that the older the driver, the less the difference between nighttime and daytime crash rates, Figure 7. This could be due to the fact that older drivers do not drive under the influence of alcohol and are likely be more careful when driving during at night in order to compensate for their physical impairments.

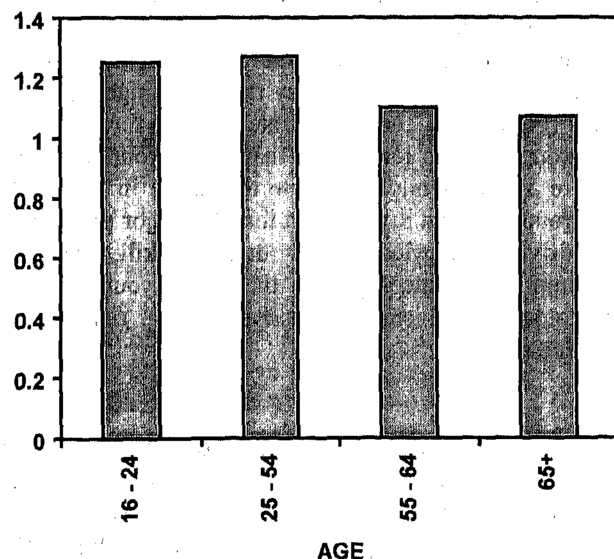


Figure 7. Ratio of Nighttime to daytime Crashes per Mile.

The effect of the very young age on crash frequency - The driving license age varies from state to state, as shown on Figures 8 and 9 [Ref. 8].

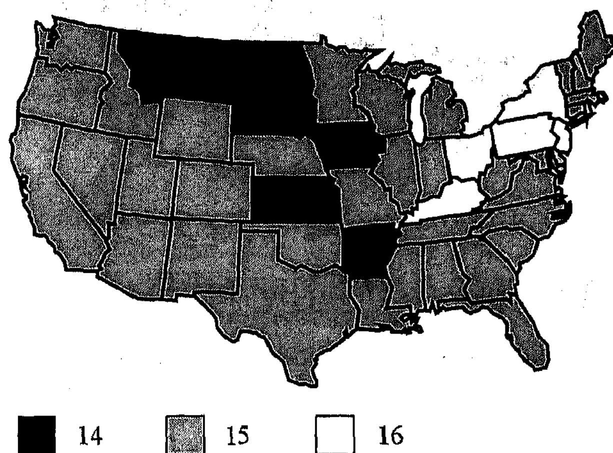


Figure 8. Minimum Age to Receive Learner's Permit

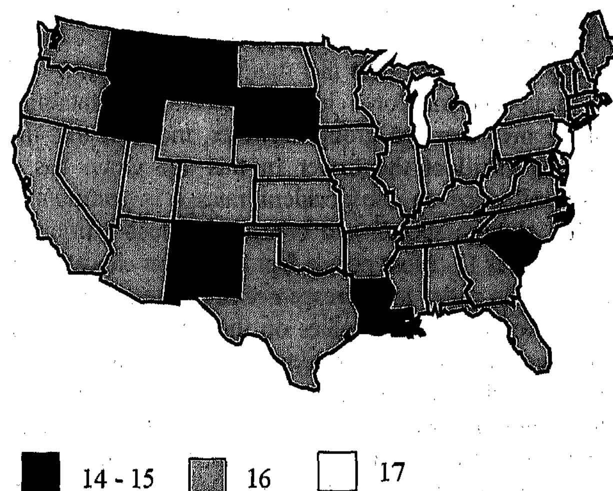


Figure 9. Minimum Age to Receive Driver license.

Intuitively, many people feel that the lower the age to obtain a driver license, the greater the problem with immature and inexperienced drivers. They reason that the better the driver education system and older the novice driver the lower the number of traffic accidents. If this is true, then one would expect high fatality rates in the 16 to 20 age groups in those states that issue driving licenses at younger ages. A comparison of Figures 8, 9 and 10, shows that such correlation is not apparent. While the accident rates are the highest in the

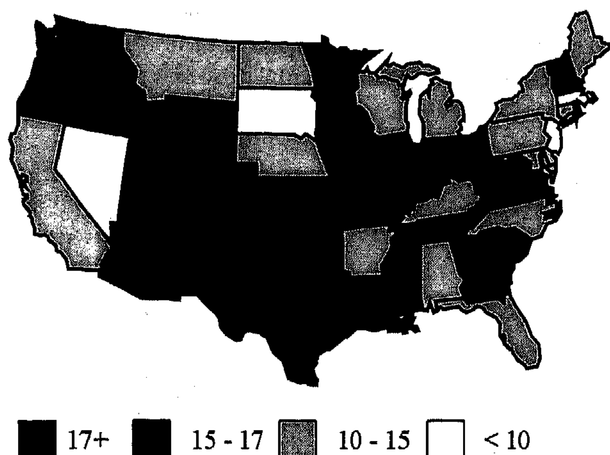


Figure 10. Fatalities in the 16-20 age group as a percentage of the total fatalities in a given state.

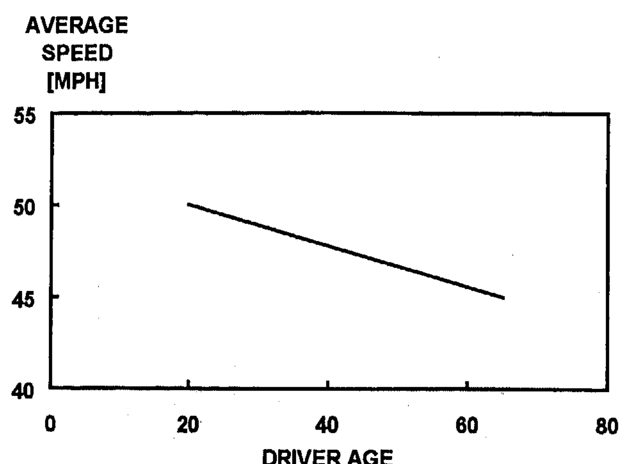


Figure 11. Observed travel speed (1984)

16 to 19 age group, these accidents are not caused by the lower age group to which the driving licenses are issued. This in turn, could mean that driver education program and driving experience are the possible causes for the high accident rates in this age group rather than is the driving license age. The fact that rollovers in the 16-19 age group account for a larger percentage of fatalities than in any other age group, however, indicates that the experience/skill of the driver in this age group has a major influence on the accident rates (in addition to driving at high speeds and consuming alcohol prior to driving) [Ref. 1]. Most people feel that driver education is important in order to decrease accident rates. Although there have been many studies on the influence of driver education programs on accident rates, none has shown that those who receive driver education have lower crash rates than those who do not. Studies show that whatever sort of driver education one received, the crash rate is typical to the age group to which that driver belongs. There is no reliable evidence for increased safety benefits from driver training [Ref. 1]. In the younger ages the driving experience and the age are the dominating factors. The age effect is strongly influenced by the mentality and local culture of the youth considered. The young age groups' affinity for "showing off" and "taking risk" may be the major factor that explains the high accident rates in this age group. The relationship between travel speed and age, shown on Figure 11 [Ref. 1] indicates that younger people more apt to take risks.

Once the driver passes the drinking age (21 yr.), then the improved driving skill is counteracted by the dramatic increase in the alcohol related accidents (as will be shown later in this paper). In any case, the driving skill itself does not necessarily decrease accident rates in most age groups (the youngest age group is a possible exception). The accident rate on public roads for very skillful racing car drivers is not any better than it is for the general public. It appears that the greater the skill, the greater the risk taking. This is also true for drivers of high quality and high performance vehicles. The old saying is that: one needs to have a high quality, high performance car in order to get out of the trouble he/she got into because of a high quality high performance car. Consequently, although the driving skills are improved by age 21, the fatality rate per population is similarly high for both the 15-19 and the 20 -29 age groups (Figure 6). Since the older (20-24) age group drives more miles than does younger (16-19) group (Figure 12), the fatality rate per mile for the 20-24 age group is much lower than that of the younger group (Figure 5). This might indicate that those in the 20-24 age group are good drivers, and that the major problem is not driving knowledge and driving skill, but the consumption of alcohol and the other youth related psychological problems (aggressive driving, risk taking, showing off, etc.).

Annual mileage driven and the age of the drivers - Annual mileage is one measure of exposure to the risk of traffic accidents. Older people usually drive less than younger people do (Figure 12). Thus, the

accident rates per drivers and the accident rate per million miles is usually very different. While the accident rate per miles is an indication of the driving

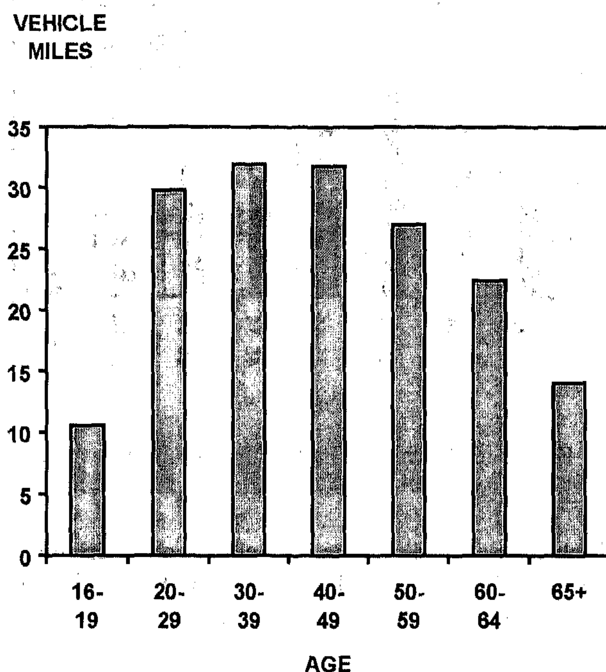


Figure 12 . Average Daily Vehicle Miles Traveled (1990)

ability, the accident rate per drivers is an indication of the effect of the driving behavior on the society. If a very old driver has great difficulty to drive without accident, and he knows this and has thus decided to drive very seldom and only for short distances in good driving conditions, then the likelihood of this driver having an accident could be less than that of a very able young driver who does, however, drive great distances each year. It appears, then, that it would be unfair to drivers to revoke their driving licenses because of age impaired driving, if other drivers, who drive large distances and thus cause and are exposed to a greater number of accidents, are not restricted. The cost to society is not different nor is the danger to the individual driver is different (young, able drivers who drive very great distances vs. old age impaired drivers who drive very limited distances). Any restrictions on older drivers appears to be unfair as long as the effect of their driving, on society as a whole, is not greater than the effect of the driving of the young drivers. When the need to restrict the driving of the older people is discussed, one should remember the effect of driving restrictions. Villeneuve [Ref. 9] attempted to draw attention to this problem and wrote: "I discovered the enormous misery that is generated by the loss or threatened loss of independent mobility. Many seniors simple refused to discuss a future without a car. Others told me that they could not conceive of such future, and some confessed they would rather die than quit driving ... [those] who were clearly unfit to drive ... denied their functional status and continued to drive in defiance of their families because independent mobility was central to their definition of self".

Fatality rates by income and age - Fatality rates in an area of residence depends upon the income of the area residence, as shown in Figure 13 [Ref. 10]. The higher fatality rates in lower income areas are not explained by differences in the amount of miles traveled, since the amount of travel increases with income (Figure 14 [Ref. 13]). Up to a given age, however, income and age are related; the older the person, the higher the income. Thus, the effect of income on traffic accidents may be partially age related. In addition to age, the socioeconomic status of the drivers is important. For example, it was found by Adams [Ref.12], that "An unskilled manual worker is four and a half times more likely to be killed in a motor vehicle accident than a self employed professional ..." We could not find direct data in the literature on the effect of income by age group on accident rates.

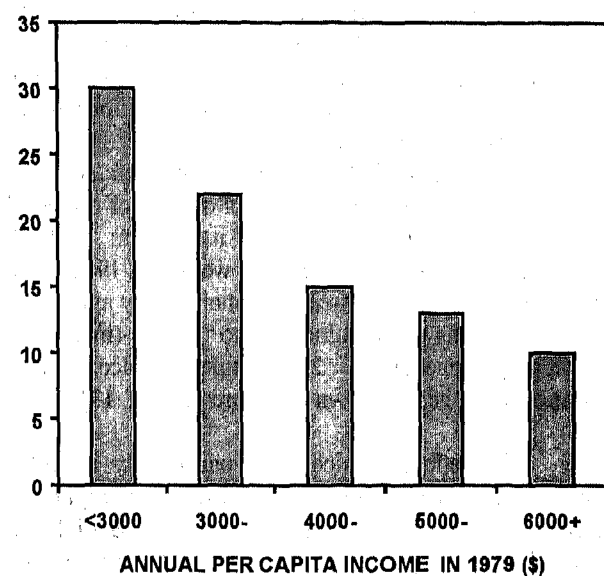


Figure 13. Fatality Rates per Capita Annual Income of Area of Residence (1977-79; Ref. 13)

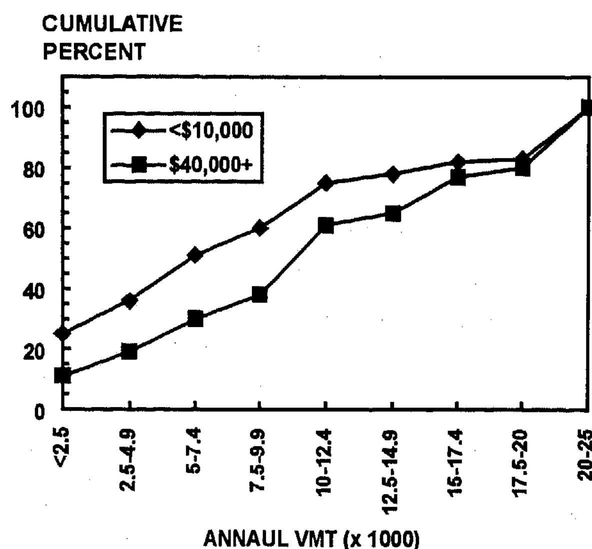


Figure 14. Cumulative Vehicle Miles Traveled for Selected Income Classes.

Type of accidents and the age of the drivers - Older drivers are less frequently involved in single car accidents, rollovers, alcohol and speed related accidents. On the other hand, older drivers are more often a party to accidents at intersections, making left turns, changing lanes and merging. [Ref. 2].

Alcohol consumption at various ages - The same amount of alcohol per body weight usually results in a higher BAC for older people. Furthermore, alcohol may also result in a greater impairment of performance in elderly persons. Women also achieve higher BAC for a given amount of alcohol per body weight [Ref. 2]. There was no material found on the effect of alcohol on elderly women, but, from the above, it appears that the significance of alcohol for elderly women is likely to be greater than for elderly men. Alcohol, in any age, makes the body organs more vulnerable to injuries (in addition to impair judgment) [Ref. 2]. Fortunately, alcohol usage declines with age, as shown in Figure 15 [Ref. 11].

The effect of alcohol is particularly important during nighttime driving. As shown in Figure 16 [Ref.17], close to 80 percent of the weekend nighttime driver fatality is alcohol related. One of the reasons why older people have much lower alcohol related accident rate is because older people less frequently go out to

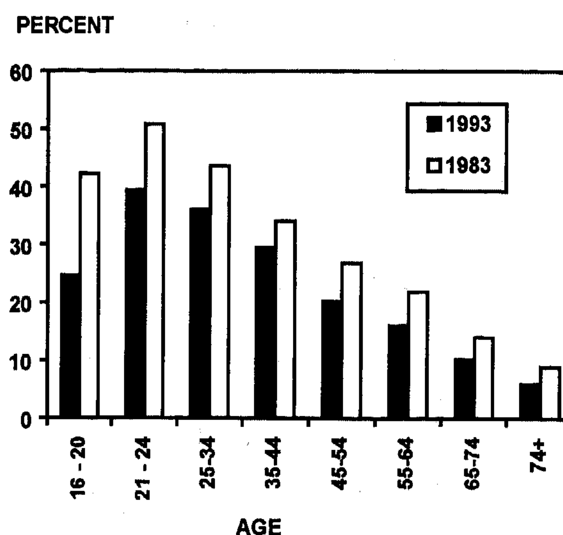


Figure 15. Drivers in Fatal Crashes; BAC = 0.01+

drink on weekend nights. Drinking and driving do not mix. The effect of alcohol drinking by young drivers is very significant on traffic accident rates.

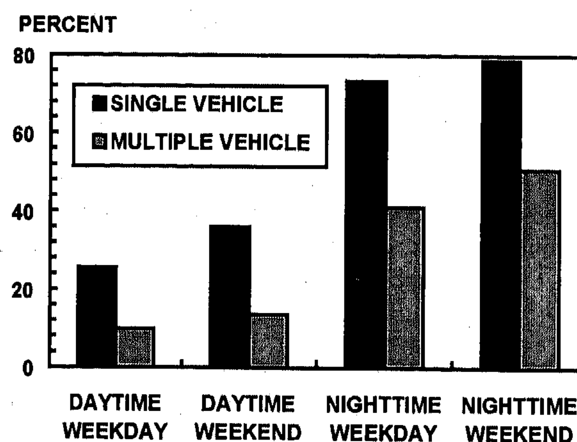


Figure 16. Percent of 21 and older drivers killed with alcohol involvement [Ref. 17]

Seat belt usage at various ages - The restraint usage by occupant survivors of fatal crashes and the occupant fatalities of passenger cars is given in Figure 17 [Ref. 14]. As it is shown, restraint usage is increasing with age. It is also apparent that those who are involved in fatal accidents are less likely to wear seat belts than are

those who survive these accidents. Seat belts clearly save lives at any age. Efforts should be made to ensure that young people do use their seat belts.

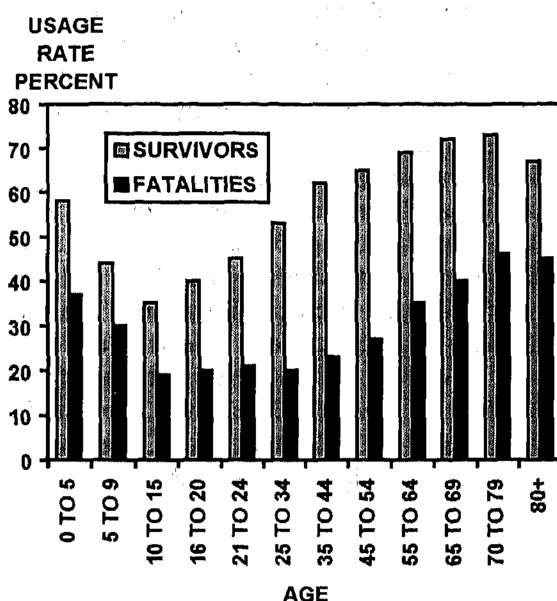


Figure 17. Restraint Usage; Occupant Survivors and Occupant Fatalities (1991; Ref. 14)

CONCLUSIONS

- A very large segment of fatal accidents are caused by young drivers who drink prior to driving and who, in addition, do not wear their seat belts. The probability that a given death is due to motor vehicle crash is fifty times higher for a person age twenty than for a person age seventy.

- Older people are more "fragile" than young people. Traffic injuries resulting in fatalities are seven times higher for an 85 year old than for a 20 year old.

- Fatality rates appear to be significantly lower for females. However, when it is defined as fatalities per miles, the difference between male and female is quite small. That is, females drive less and therefore they are less frequently involved in traffic fatalities. When both males and females drive the same distance, then the fatality rates are close (the female rate, however, is still slightly lower.)

- The importance of driver education is questionable. The skill of the drivers does not appear to effect the frequency with which a driver is involved in accidents, except for the youngest of drivers. Skilled racing car drivers do not have lower general traffic accident rates than does the general public.

- The age at which driver licenses are issued does not appear to influence accident rates.

- More traffic accidents occur at night than during the day. The likely primary cause is increased alcohol usage prior to night driving.

- The reason for lower accident rates and injuries for older drivers are due to the fact that older drivers drive at lower speeds, they are less frequently consume alcohol prior to driving, they drive fewer miles, and they are more likely to wear their seat belts.

- While very old drivers have increased accident rates, the impact of their accident rate on society is small; because they drive few miles and thus the total number of accidents in which they are involved is low. To improve the traffic accident situation, the main efforts have to be aimed at the young drivers. Their alcohol consumption, reluctance to wear seat belts and the greater number of miles they drive all contribute significantly to high accident rates. Driver education aimed at improving driving skill has a minimal effect on accident rates. However, driver education and law enforcement aimed at reducing the consumption of alcohol prior to driving and to encourage the usage of seat belts might contribute to a decrease in present accident rates.

REFERENCES

1. Evans. L. " Traffic Safety and the Driver." Van Nostrand Reinhold. 1991. SAE Paper No. 950871; entitled: Air Bag effectiveness - Trading Major Injuries for Minor .

2. UMTRI-94-43. December 1994. The Safety and Mobility of older drivers: What we Know and Promising Research Issues.

3. Oreskovich, M. R., et al. " Geriatric trauma: Injury pattern outcome." Journal of Trauma, Vol 24, p 565-572. DATE?

4. Journal of Trauma. Vol.28, No.3; March 1988. L. Evans. " Risk of Fatality from Physical Trauma versus Sex and Age.

5. NHTSA. 1994 Traffic Crashes, Injuries, and Fatalities - Preliminary Report. DOT HS 808 222, March 1995

6. Stutts, J.C and Martell, C. "Older Driver Population and Crash Involvement Trends, 1974--1988". Accident Analysis and Prevention. Vol.24, No.4, August 1992.

7. NHTSA. Traffic Safety Facts 1994. Received from NHTSA on Aug. 1995. Older Population.

8. IIHS Status Report, Vol. 30, No. 6, July, 1995. State by State Licensing Regulations Applying to Teenagers

9. Villeneuve, E. "Driving decision for seniors: A cost-effective community-based program for intervention and research with older drivers". Proceedings of Elderly Mobility Conference. Detroit, MI. Southeast Michigan Council of Governments. October, 1994.

10. Baker et. al. ,"The Injury Fact Book" Lexington Books. 1984.

11. NHTSA. Traffic Safety Facts 1993. Miscellaneous tables on age effect.

12. Adams, J.G.U. "Risk and Freedom - The Record of Road Safety Regulations." Bottesford Press, UK. 1985.

13. Nationwide Personal Transportation Survey, FHA, DOT, February 1995.

BIOGRAPHY

Dr. George Z. Libertiny (Ph.D. Mechanical Engineering) is the Principal Consultant of Design and Manufacturing Defect Consultants, Dearborn, Michigan. Prior to his present position he was a Principal Research Engineer Associate in the Automotive Safety and Engineering Standards Office of Ford Motor Company, Dearborn, Michigan. Prior to his position with Ford he was a professor at the Illinois Institute of Technology, Chicago. In addition to his employment with Design and Manufacturing Defect Consultants he is a Professor of Mechanical Engineering, Adjunct Faculty, University of Michigan-Dearborn. He has published more than 50 papers. He is a member of SAE and a Fellow of ASME. He is holder of the Ralph R. Teetor Award (SAE), the Forest R. McFarland Award (SAE) and the Outstanding Engineer Award (Michigan Society of Professional Engineers).

970278

A Study of an Evaluation Based on Physiological Responses of a Driver's Task and Stress Level While Maneuvering a Vehicle

Yuzuru Matsuura and Takatoshi Yanagida
Osaka Sangyo University

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

To evaluate a driver's task and stress level objectively while maneuvering a vehicle, we paid attention for human's physiological responses under the condition. The eccrine sweat on a palm and a forehead, heart rate and a face temperature of a subject driver were monitored and measured continuously and quantitatively using the specified direct sweat measuring equipment, the electrocardiograph and the infrared thermometer respectively.

As a result of having analyzed these data physiologically, we propose that the driver's stress level can be objectively determined by the quantity of eccrine sweat, changes of heart rate and changes of face temperature.

INTRODUCTION

The safety of a vehicle-road traffic greatly depends on a judgment and an action of a driver as well as road environments and vehicle performances. A driver's physical task to operate a vehicle is not so high, but the mental stress and the tension of a visual and hearing system and brains nervous system are extremely high.

Up to now, the evaluation of drivers' stress level of driving task and the mental stress generally depends on subjective evaluations of drivers' feeling by themselves. These values are always uneven because of their individualities and different way of their expressions.

In this study, to evaluate these levels objectively, we paid attention for human physiological responses while maneuvering a vehicle.the eccrine sweat on a palm and a forehead, heart rates and a face temperature of a subject driver were monitored and measured continuously and quantitatively using the specified direct sweat measuring equipment, electrocardiograph by a medical telemeter system and an infrared thermometer camera. These data moreover compared with similar data collected when subjects watched a TV and tried to solve arithmetic problems as conventional basic tasks.

EXPERIMENTAL APPARATUS AND METHOD

EXPERIMENTAL APPARATUS - To investigate human physiological responses, the subject's sweats on his forehead and palm were measured directly, momentarily and quantitatively with the direct sweat measuring equipment that uses carrier nitrogen gas as shown in Figure 1.

There are two types of the sweat. One is Thermal Eccrine Sweat (abbreviated as TES) and the other is Mental Eccrine Sweat (MES). Capsules pasted on two parts of a subject's forehead and left palm collected information on thermal reaction (TES) and mental reaction (MES) respectively. measured using the multi-telemeter system and their heart rates were calculated from the electrocardiograms (cf. in Figure 1).

Their face temperatures were also measured by the infrared thermometer camera (Thermotracer TH2100), Their operating behaviors were also recorded onto the 8 mm video tape recorder through a CCD camera mounted in front of them.

These data moreover compared with similar data collected when subjects watched a TV and tried to solve arithmetic problems .

EXPERIMENTAL METHOD - The experiment consists of fundamental experiments in the laboratory and experiments of a vehicle traveling in the test course.

(1) Fundamental experiments in the laboratory

From various daily occurrences that we generally experienced in our life, we selected typical tasks accompanied with visual stimuli or brains stimuli, and conducted several experiments to obtain the state of human physiological stress under these stimuli.

(2) Experiments of visual stimuli

Subjects watch a video monitor as a one of visual

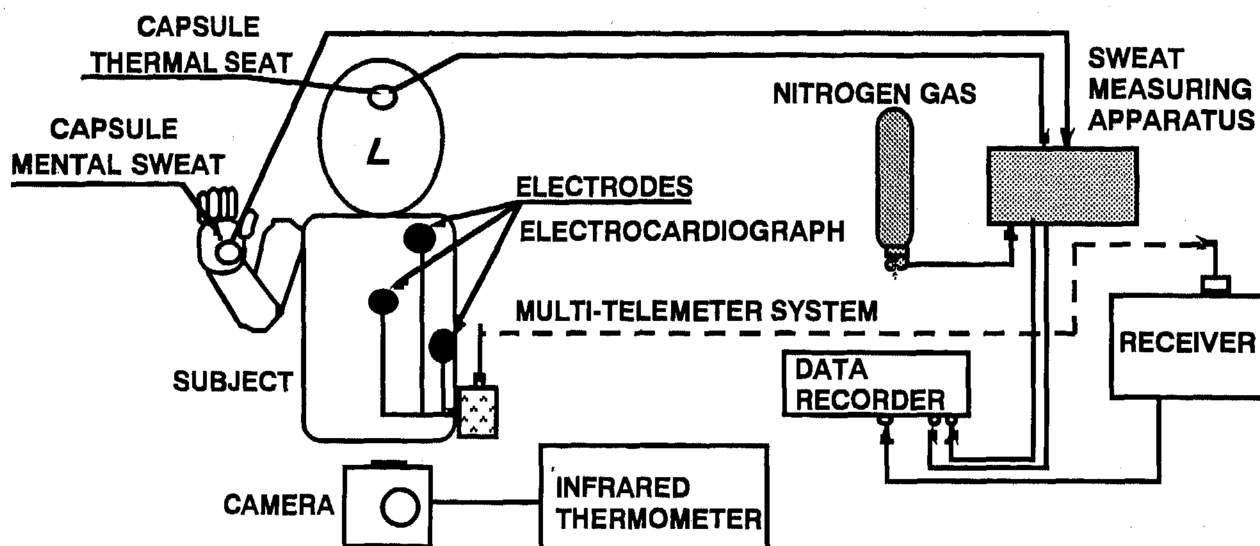


Figure 1 : The experimental apparatus

stimuli. Two subjects of pictures "mountains' nature scenery" and " motorcycle's race" were selected, so that the former gave subjects calm scenes and the latter gave them exciting scenes, in 6 minutes each.

(3) Experiment of brain stimuli

Fifty arithmetic calculation problems were given to subjects as brains stimuli. They calculated these problems for 10 minutes, and meanwhile their brains' stress were physiologically examined.

In each experiment, the room temperature kept approximately 24 centigrade. Their face temperatures were taken with the infrared thermometer camera every 4 seconds.

The subjects had the operating task of slalom-travels at 15 km/h and 30 km/h (each pylon's interval of 15m on the course) using a FF type passenger car on the test ground, and furthermore had constant radius turn travels of 15m radius at velocity of 30km/h.

It is difficult to be kept constant the inside temperature of the car cabin. It also needs to avert the subject's face from outside fresh wind or from air flow of the air-conditioner, to avoid an influence of cooling operation to the face. Therefore all windows of the vehicle were closed, and its inside conditions were kept freely without operating the air-conditioner. The face temperature was filmed every 2 seconds with the infrared camera.

The subjects consisted of 5 male adults of 21 to 23 years of age.

EXPERIMENTAL RESULTS AND CONSIDERATION

FUNDAMENTAL EXPERIMENT IN A LABORATORY - Figure 2 shows a result of sweat changes of a subject measured in the fundamental experiments of visual stimuli and brain stimuli in the laboratory. In thermal sweat of the forehead part, a change is hardly observed.

At early stage of the rest state, sweating on the forehead part was hardly found, but the quantity of the sweat on the palm part showed a high value. As this reason, many mental sweat glands are distributed on the palm part. This sweat phenomenon is considered to have shown a tension feeling of the subject just before a start of the experiment.

While the subject watched the mountains' scenery video, the quantity of his sweat didn't change greatly, but it began to rise just after the start of the experiment and afterwards its level kept high while watching the motorcycle's race video. For the continuation about exciting scenes of the race video and the delay of sweating phenomena, however the finish of the experiment of the video, the quantity of his sweat remains high level. The quantity of the average sweat in this period was 0.105 $mg/cm^2 \cdot min.$, and the total quantity of sweat during 8 minutes was 432.66 mg/cm^2. This shows the mental stress of the subject caused to be excited by the race video.

When the subject calculated the arithmetic problems,

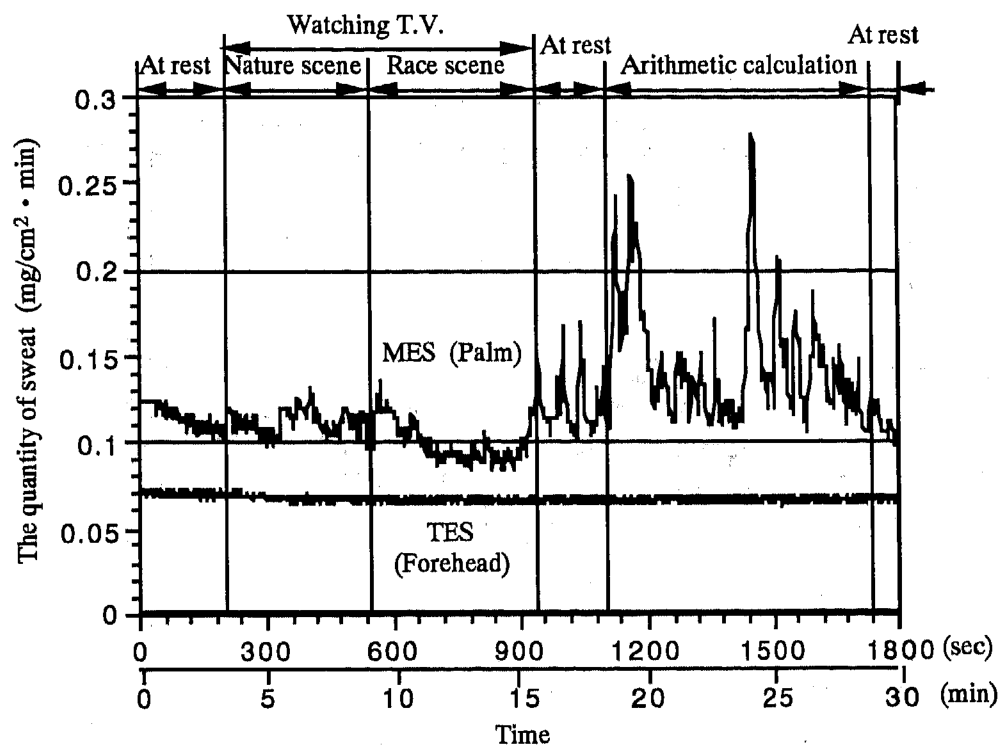


Figure 2: The result of sweating measured in the laboratory

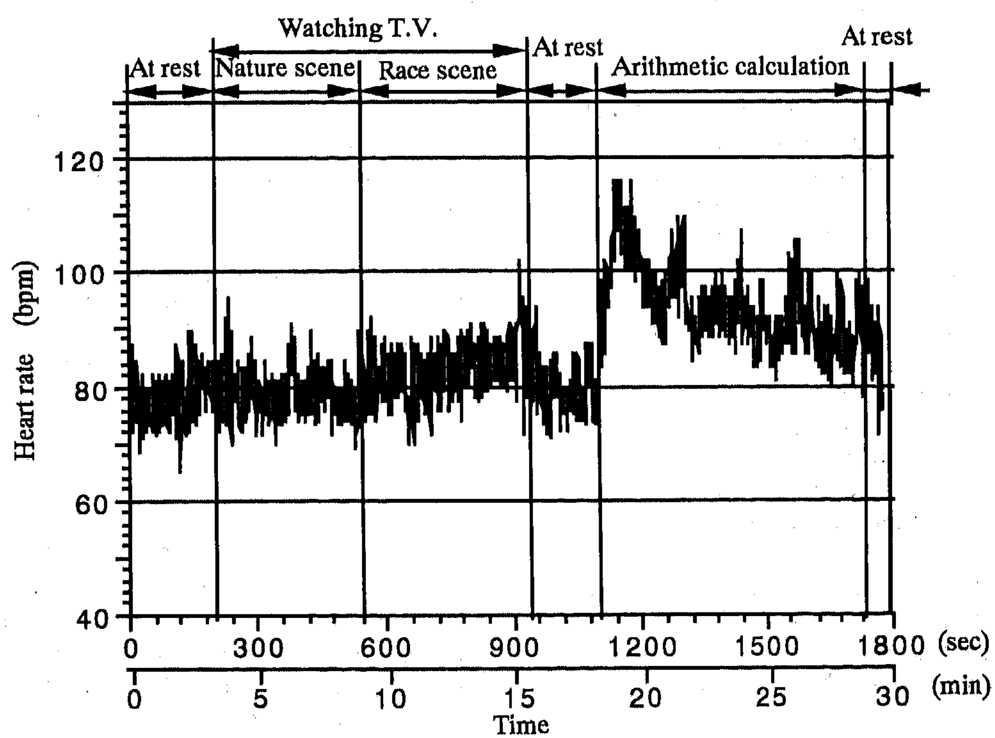


Figure 3: The result of heart rate measured in the laboratory

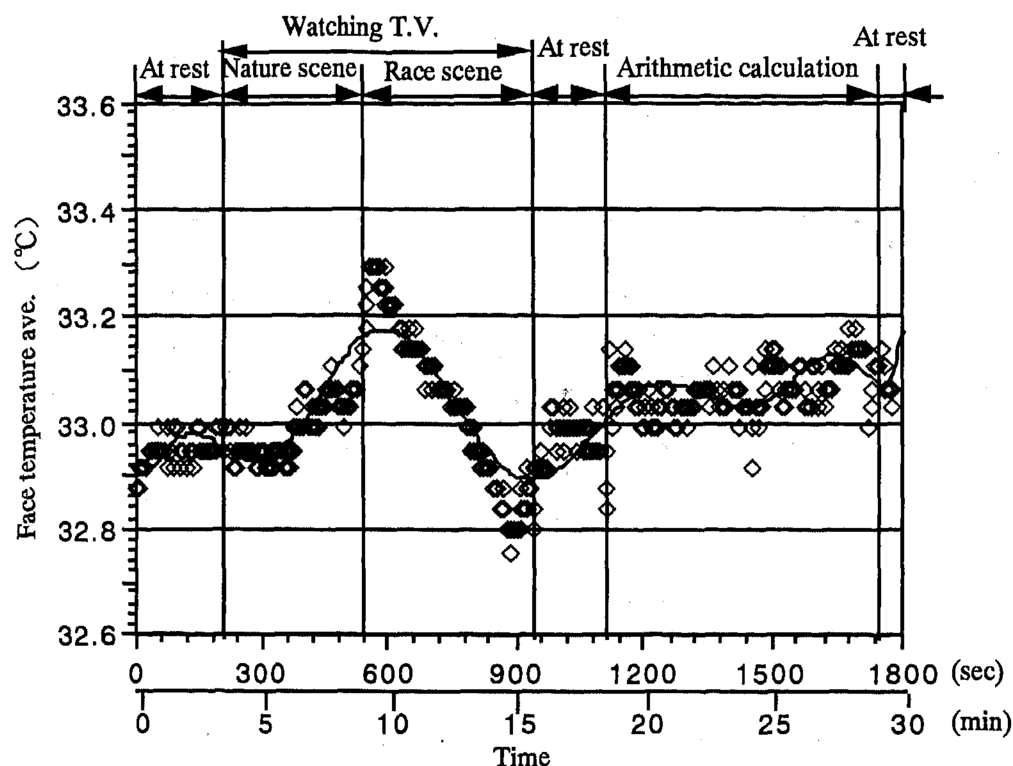


Figure 4 : The result of the face temperature measured in the laboratory

the quantity of his sweat greatly increased immediately after the calculation started, and fluctuated with a high level between 0.108 and 0.279mg/cm²•min. afterwards.

This trend was observed in all subjects. From this fact, when subjects input information, and processed and output ones, it is considered that they had many mental stresses.

The quantity of the average sweat in this period was 0.145mg/cm²•min., and the total quantity of sweat during 10 minutes was 1042.83mg/cm².

Now, let compare the quantity of the sweat in each experiment with that in each rest state, when subjects were watching the mountains' scenery on video monitor, the quantity of their sweat then is 0.73 to 1.02 times larger that in the rest state.

On the other hand, while they were watching the motorcycle's race, and were calculating the arithmetic problems, the quantity of their sweat is 0.90 to 1.23 times and 1.31 to 1.53 times larger that in each rest state respectively.

About the heart rate of subjects, the average heart rate in the rest state before the experiment showed 78bpm, and while their calculating the arithmetic problems, the heart rate is 1.16 to 1.24 times larger that in the rest as shown in Figure 3. In this way, we found that the stress state of the subject also could be shown by the change of heart rate.

Now, about a change of face temperature as shown in Figure 4, the average face temperature of the subject was fixed 32.0 degrees in the rest state and rose about 0.05 degrees while they were watching the mountains' scenery on video in comparison with the rest state, about 0.1 degrees while watching the motorcycle's race and about 0.1 degrees while calculating the arithmetic problems respectively.

We consider that the mental excitements by each stimulation have brought the rise of their face temperature.

EXPERIMENT OF A VEHICLE TRAVELING - As shown in Figure 5, a large change of the quantity of sweat on the driver's forehead part wasn't observed in the experiments of the vehicle traveling. The above fact shows that the operating task wasn't intense muscle exercise such as causing a thermal eccrine sweat.

On the other hand, a remarkable change of the quantity of sweat on the palm part was observed. The average quantity of sweat of the driver was 0.184 mg/cm²•min. in the rest state before the experiment, and the quantity of his sweat increased remarkably just when the slalom

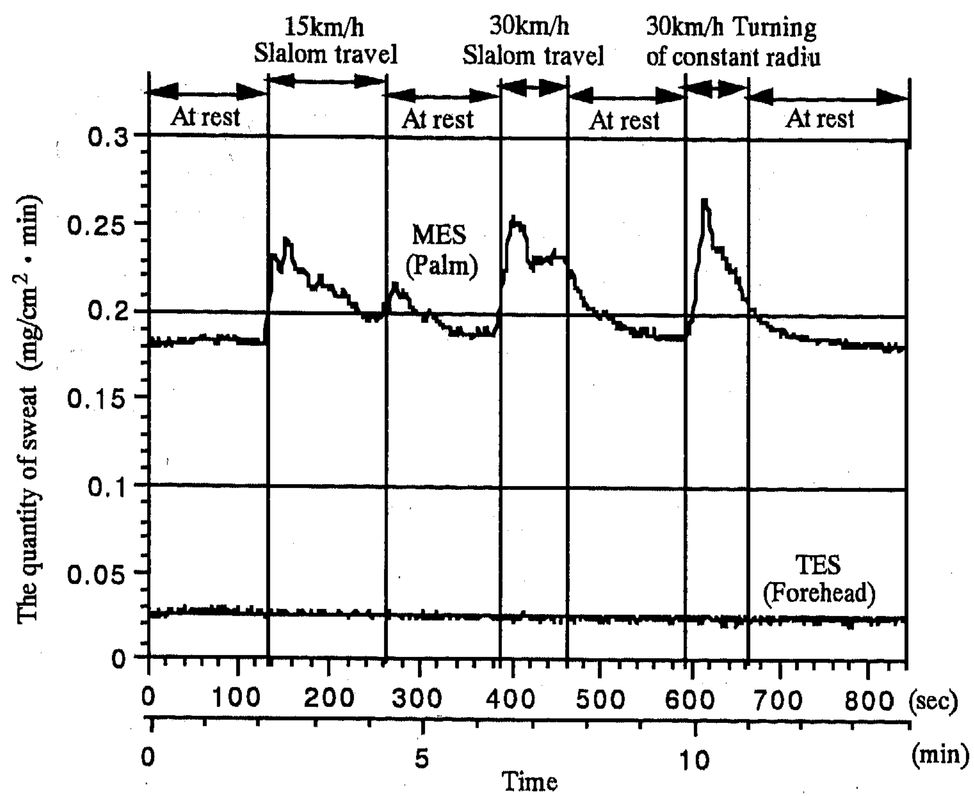


Figure 5 : The result of sweating measured in traveling vehicle

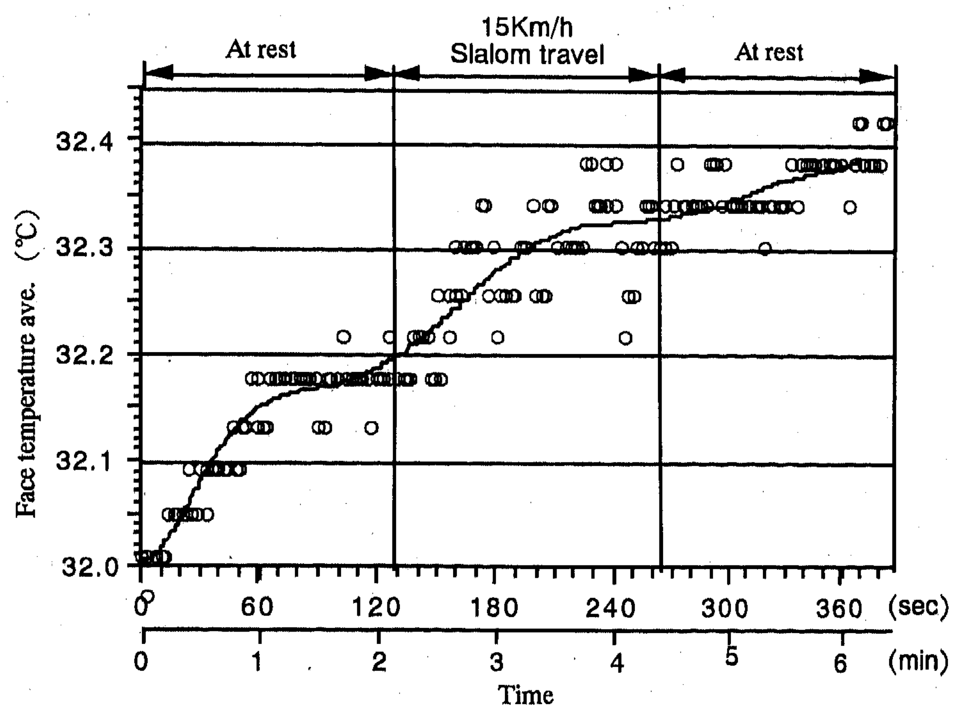


Figure 6 : The transition of face temperature while maneuvering the vehicle

Table 1: The comparison of physiological responses during the experiments on that in the rest state

* RATIO NUMBER

	HEART RATE	ECCRINE SWEAT	FACE TEMPERATURE
REST BEFORE EXPERIMENT	1	1	1
MOUNTAINS' SCENERY	1.00~1.09	0.73~1.02	0.90~1.03
MOTORCYCLES RACE	1.07~1.11	0.90~1.23	0.91~1.06
ARITHMETIC CALCULATION	1.16~1.24	1.31~1.54	1.00~1.05
SLALOM TRAVEL AT 15 Km/h	1.35~1.42	1.11~1.53	———
SLALOM TRAVEL AT 30 Km/h	1.40~1.53	1.13~1.98	———
CONSTANT RADIUS TURN TRAVEL	1.37~1.50	1.14~1.61	———

travel started at 15 km/h, and reached a peak value of 0.242 mg/cm²•min, and decreased later. The total quantity of sweat during 2 minutes was 0.217 mg/cm²•min, and the quantity of the average sweat in this period was 47.090 mg/cm². In the rest state immediately after the experiment of the slalom-travel at 15 km/h, the quantity of sweat showed a high value for the influence of that travel.

The quantity of the average sweat in the rest state that calmed down was 0.233 mg/cm²•min., and it was 0.193 mg/cm²•min. in the experiment of the slalom-travel at 30 km/h, and the total quantity of sweat for 1.5 minutes was 25.491 mg/cm². And the quantity of the average sweat in the experiment of the constant radius turn travel was 0.235 mg/cm²•min.

In next, we compare these above mentioned quantity of sweat with the average quantity of sweat in the rest state before each experiment, in the experiment of the slalom-travel at 15 km/h and 30 km/h, it was 1.11 to 1.25 times and 1.13 to 1.35 times as large as that in each rest, and in the experiment of the constant radius turn travel, it was 1.14 to 1.30 times respectively. In other words, we can propose that these numerical values show a size of the driver's mental stress. As similar as mentioned above, let compare between heart rates in the rest state and that in the experiment of the travel. In the experiment of the slalom-travel at 15 km/h and 30 km/h, it was 1.35 to 1.42 times and 1.40 to 1.53 times as large as that in each rest state respectively, and in the experiment of the constant radius turn travel, it was 1.37 to 1.50 times. The increase of a velocity in the slalom travel has a tendency to increase the mental stress of the driver, and the numerical values in the slalom travel show a high value than that in the constant radius turn travel. It shows that the driver's stress by operating a steering in the slalom travel was large.

About the change of the driver's face temperature, in the experiment of the vehicle traveling, because the driver is always moving his face during operating, It is difficult to take it by the infrared camera from the same direction angle. So, his face during the operation was taken frequently every 2 seconds by the camera, and after that his face photographs with the same direction angle only were selected out of the many face photographs.

An outline of his face in these photographs was judged individually and the temperature distribution state and the average temperature value of all of these were obtained.

In next, as shown in Figure 6, over the whole period the driver was operating, it was observed that his face temperature (a mean value) had an inclination to rise up gradually. As the reason, all of the windows of the vehicle were closed and its inside conditions were kept free, so the inside temperature in the cabin rose a little by little, from the start of the experiment (approximately 14 degrees) to the end (approximately 28 degrees). His face temperature also rose progressively by the influence of the room temperature.

Now, to analyze the influence that the operation task gave a change of the driver's face temperature, the whole transition of his face temperature accompanied with his operating progress was similar by a curved line of undulate waviness (n-powered Polynomial of Degree). It is shown with the curved line in the figure. In an early stage of the rest state, a rapid rise of the face temperature was observed because of the influence of the interception of the external fresh air caused by having closed the all windows, and it calmed down in later period.

When the experiment of the slalom-travel at 15 km/h

started, his face temperature began to rise clearly. It is considered that it was caused from the effect of the stress accompanied with the driver's operating task, as well as the above-mentioned results of changes of sweat and changes of heart rate. He entered the rest state after he finished the experiment of the slalom- travel, and the rise of his face temperature calmed down again.

Afterwards it rose slowly in accordance with the influence of the rise of the inside temperature of the cabin.

COMPARISON OF THE RESULT IN THE LABORATORY AND ON THE VEHICLE TRAVEL - We compared the experimental result of the vehicle traveling with that in the laboratory in respect with all values of the quantity of the average sweat, heart rate and the face temperature.

However there were individual differences caused by the subjects, as shown in Table 1, about a ratio of the quantity of the average sweat the former's value is 1.3 to 2.0 times as larger than the latter's one, and about changes of the heart rate, it did larger 1.2 to 1.5 times, and about changes of the face temperature, it did higher 1.0 to 2.0 times.

Accordingly, it proved that the mental stress while operating the vehicle was comparatively higher than that during watching the television and that during calculating arithmetic problems.

CONCLUSION

As the results of having analyzed the mental sweat and changes of the face temperature in respect to human physiological responses to various types of stimulation, the following results were obtained.

(1) When a mental stress burdens on a human body, his physiological responses appear such as an increase of the quantity of sweat on a palm, an increase of the heart rate and a rise of the face temperature.

(2) The human mental stress level to a stimulation can be evaluated numerically by comparing a size of his physiological responses to it.

(3 The human stress while maneuvering a vehicle is comparatively large more than that of daily occurrences in our life such as watching a TV and calculating arithmetic problems.

REFERENCES

(1) Yuzuru Matsuura: Interacting Roles of Road Environment, Vehicle and Driver for Road Safety, JSAE, Journal of the S A E of Japan Vol.45, No.10 (1993-10)

(2) Yuzuru Matsuura, Zelin Chen et al,: Study of Maneuvering Feeling via Driver's Physiological Response, SAE Transaction 931930, p2092~2100(1994-9)

(3) Yuzuru Matsuura, Zelin Chen et al: A Study of Evaluation of Maneuvering Driver's States through His Physiological Response, JSAE paper Vol.25, No.1, 9430365, p112~118(1994-1)

(4) Yuzuru Matsuura: The Difference Between Driver's Visual Experiences and His Body's Sensations Whilst Driving, Proceedings of the 12th Congress of International Ergonomics Association 1160, (1994-8)

(5) Yuzuru Matsuura, H.Matsuoka: A Study of the Motion Sickness Incidence Mechanism and Evaluation of a Passenger's Comfort Level, ISATA 27th, Road and Vehicle Safety, p365~372(1994-10)

970279

Driver Behavior Under a Collision Warning System - A Driving Simulator Study

Takamasa Suetomi and Koji Kido
Mazda Motor Corporation

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Collision warning systems are expected to be an effective countermeasure to reduce traffic accidents; however there have been relatively few studies on the effects of such warning systems on the driver's collision avoidance behavior. In this study, a driving simulator which had a large motion system was used, and 45 subjects experienced crash imminent situations in which the preceding cars suddenly decelerated while the subject looked off the road. Analyzing the subjects' collision avoidance behaviors, it was found that the braking response time and the number of simulated collisions were substantially decreased with collision warnings. Furthermore, potential reduction of rear-end collisions on the road was estimated by modeling the driver's braking response.

INTRODUCTION

Most of traffic accidents are considered to be caused by drivers' errors. A study reported that drivers' errors(recognition errors, decision errors, erratic actions, drowsy, etc.) consisted about 90% of all the causal factors of traffic accidents[1]. Therefore, active safety technologies such as collision warning systems to prevent or to cover driver's errors are expected to be an effective countermeasure against such driver caused accidents.

To develop an effective collision warning system, the human-machine interface is as important as the sensing system which detects surrounding obstacles and judges hazard levels. Therefore such system should be evaluated from the following view points.

(1) Collision avoidance potential (the number of collisions which the warning system can prevent directly)

(2) Driver's behavioral changes(the driver should not take more risks and reduce the safety margin against collisions by taking advantage of the warning system)

(3) User acceptance(The system should not irritate the driver and he/she uses the system willingly)

Therefore, first the drivers' responses to collision warnings must be carefully analyzed. Then, the performance of a collision warning system must be evaluated based on its collision avoidance potential.

There is a study which used a computer simulation and predicted that a collision warning system would reduce 60 % of rear-end collisions by assuming drivers' responses and a traffic flow[2]. However, considering drivers' behavioral changes in use of such warning systems, the man-in-the-loop analysis needs to be done carefully to predict their effectiveness in real world.

For such studies, a driving simulator which can evaluate ordinary drivers' responses under crash situations is an essential testing method since such tests are difficult in real world.

Fixed-base simulators or motion-base simulators with small motion ranges have been used to investigate drivers' behaviors while using active safety systems such as intelligent cruise control[3] and collision warning system[4]. However, to fully analyze the effects of collision warning systems on the drivers' behaviors and to evaluate collision avoidance potential, the simulator must be able to simulate crash imminent situations including drivers' collision avoidance

behavior. And to reproduce drivers' emergent maneuvers such as emergent braking or steering, an accurate motion feedback to the driver would be necessary. For this purpose, Mazda has developed a driving simulator with a large scale motion system[5].

In this paper, using this driving simulator, the responses of 45 subjects when they were faced with sudden brakings of the preceding cars were analyzed. Then, based on the results, a model of the driver's braking response was developed and potential reduction of rear-end collisions on the road was estimated.

EVALUATION OF COLLISION WARNING SYSTEM USING A DRIVING SIMULATOR

DRIVING SIMULATOR

Motion System - Figure 1 shows the mechanism of the four degrees-of-freedom motion system of the Mazda driving simulator. Consisting of a rotational mechanism that controls the three axes of roll, pitch, and yaw of the movable cabin and a linear mechanism that controls the horizontal motion of the cabin, it has a large moving range in each motion direction as shown in Table 1.

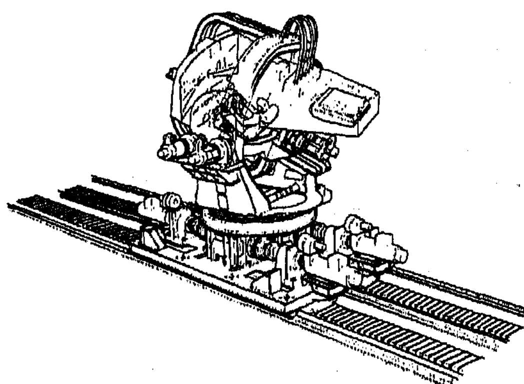


Figure 1. Driving simulator motion system

Table 1. Moving range of motion system

Coordinate	Range
Roll	+/- 40 degrees
Pitch	+/- 40 degrees
Yaw	+/- 160 degrees
Linear	+/- 3.6 meters

Rotating the moving base around the yaw axis by 90 degrees, the linear mechanism can generate either longitudinal acceleration or lateral acceleration to the driver. 0.8 G acceleration can be generated by actual linear motion, and 0.64 G acceleration can be attained by inclining pitch or roll to the maximum angle of 40 degrees. The three rotational axes pass through the driver's head where the vestibular sensors of the driver are located.

To simulate a large longitudinal acceleration in emergent braking, this motion system uses linear motion and pitch motion. The high-frequency component of acceleration is produced by the linear motion and the low-frequency component is produced as the component of gravitational acceleration by tilting the pitch of the cabin. In a short beginning stage of braking or acceleration, the linear motion system generates an actual longitudinal acceleration for a short period of time and decelerates it quickly to stay within its available displacement. And, following the stage, a constant acceleration is maintained by tilting the pitch angle.

Visual and Audio System - The visual system generates a road image onto an 80-inch screen placed 1.2 meters in front of the driver. The field of view is 68 degrees horizontally and 20 degrees up and 14 degrees down from the horizon. A full color(16.78 million colors) textured road and environment image is generated every 20 milliseconds by the road image generator. An example of road images is shown in Figure 2.

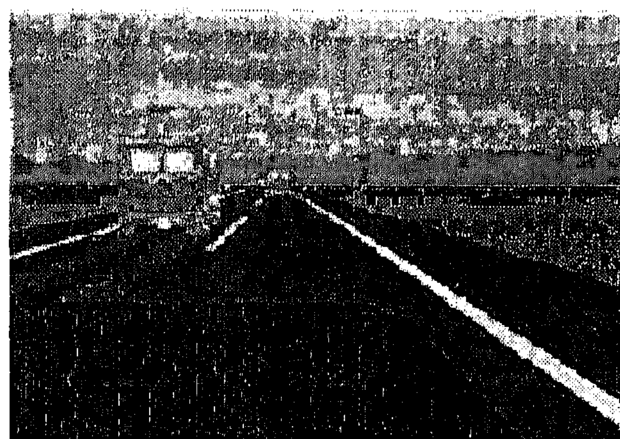


Figure 2. Road image

The road image generator which could generate road scenes and other traffics were used to produce a car following situation on a highway. Road curvatures were specified both horizontally and vertically to simulate real highway. A texture for road

surface was also used to enhance the driver's perception of the vehicle speed. Images of preceding cars on the traveling lane and adjacent lanes were also generated. The image of preceding car had brake lights so that the driver of the following car could recognize the brake operations of the preceding car. The background images such as mountains and the sky were turned according to vehicle movement.

To simulate an auditory environment of a car, sounds such as those from the engine, road, wind and tires were recorded in real environments, and they were stored in a digital sampler in the sound system. Receiving signals indicating the engine rotational speed, throttle opening level, vehicle speed and tire slipping conditions from the host computer, the sound generator reproduced the corresponding stored sounds.

Verification of Driver's Braking Behavior in the Simulator - Before evaluating collision warning systems using the simulator, whether a driver in the simulator behaves in the same way as in real world should be verified. Figure 3 shows changes of the braking headway distance with the driving speed on the simulator, an expressway, and a proving ground[6]. The headway distances were averaged over 20 km/h intervals. As shown in this figure, although the distances in the simulator were smaller at low speeds and larger at middle speeds compared with real world driving conditions, braking headway distances in the simulator were generally very close to those in real world at higher speeds. Those data proved that the driving simulator with a large longitudinal stroke is effective to reproduce driver's braking behaviors.

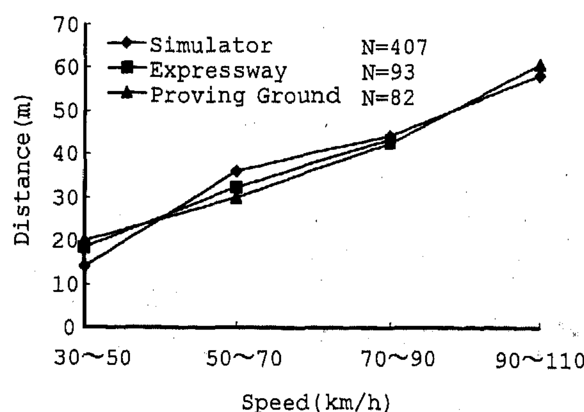


Figure 3 Changes in braking headway distance with driving speed

EXPERIMENTAL METHOD

Task - In the simulator, a subject drove on a two lane road with a lane width of 3.5 meters. The driver followed a preceding car which changed speeds from 80 km/h to 100 km/h by a sixty-second interval on the overtaking lane(right lane), while low speed vehicles(50 km/h) traveled in the running lane(left lane).

To make the subject glance off the road, he/she was asked to push a button when one of five LED lamps was lit as a secondary task. The five LED lamps were located at different places on the dashboard.

The experimenter was able to control other traffics. One hazardous scenario was preceding car's sudden braking to 30 km/h by 0.8 G, when a lamp was lit to make the subject to look at it. Another hazardous scenario was that a preceding vehicle on the adjacent lane broke into the traveling lane without blinking its turning signal just after a LED lamp was lit.

Test Procedure - At first, the subject drove until he/she became familiar with driving in the simulator and got used to the dynamic characteristics of the simulated vehicle for about 10 minutes. After a short break, in the second running, the subject drove following a preceding car, and met with a sudden braking of the preceding car or a sudden lane change of a preceding car on the adjacent lane, while executing the secondary task.

Subjects - 45 male subjects aged from 19 to 26 who had valid driving license and regular visual and auditory senses participated in the test.

COLLISION WARNING SYSTEM

The collision warning system provided a warning sound when the headway distance was shorter than the warning distance L_0 expressed by the following equation called Stopping Distance Algorithm[2].

$$L_0 = \frac{V_f^2}{2\alpha_f} - \frac{V_l^2}{2\alpha_l} + V_f \cdot t_1 \qquad \text{(Eq.1)}$$

L_0 Warning distance

V_f Following car velocity

V_l Leading car velocity

α_f Assumed following car deceleration

α_l Assumed leading car deceleration

t_1 Time margin

Table 2 shows the logic parameters used in this experiment.

Table 2. Logic parameters

α_f	5.9 m/s^2(0.6 G)
α_l	7.8 m/s^2(0.8 G)
t_1	1.0 sec

EXPERIMENTAL RESULTS

Driver's Brake Response - Figure 4 shows the relationship between the subject's brake response time and the headway distance just after the preceding car's sudden deceleration for the case of with/without warnings. The subjects who avoided collisions with the preceding car by changing the lane were excluded in this figure. Filled markers indicate that the subject couldn't avoid a collision. This result shows that the subjects who maintained small headway distances and/or responded slowly failed to avoid collisions and that the subjects' response times were generally shortened by collision warnings and also they keep longer headway distances with collision warnings.

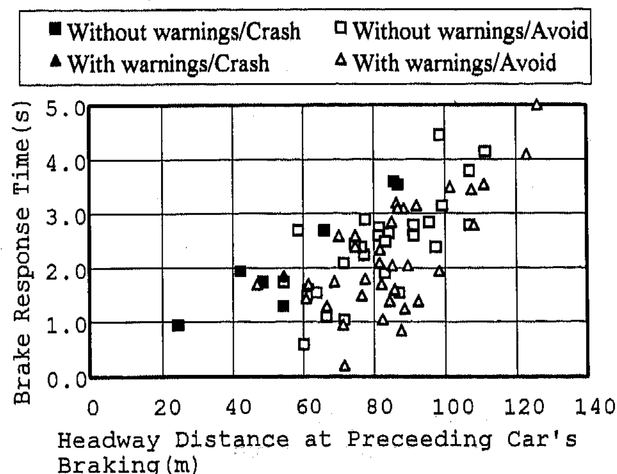


Figure 4. Braking response time with/without warnings

Before analyzing subjects' collision avoidance behavior, the subjects were grouped by the headway distance in car following situations into 3 groups: a small-headway group which kept headway less than 1 second(30%), a middle-headway group which kept headway from 1 to 2 second(52%), and a large-headway group which kept headway more than 2 second(18%).

Figure 5 shows the averaged response times of braking of each group after the preceding cars' deceleration. The average values were calculated by taking the logarithms of the response times. Table 3 shows the distribution of the logarithms of the response times with/without warnings. The results showed that the subjects in the middle-headway group responded about 0.5 seconds earlier on average with warnings than without warnings($p<0.05$), and in the other groups there were no significant differences.

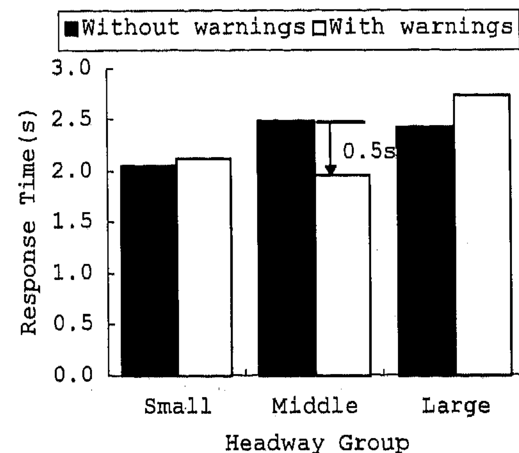


Figure 5. Averaged response time of braking for three headway groups

Table 3. Distribution of the logarithms of response times with/ without warnings

Headway Group	Average		Standard Deviation	
	without warnings	with warnings	without warnings	with warnings
Small	0.62	0.70	0.51	0.33
Middle	0.85	0.59	0.38	0.42
Large	0.83	0.95	0.35	0.36

Collision Rate - Figure 6 shows the percentage of the subjects who could not avoid collisions when the

preceding car decelerated suddenly under the conditions with and without the collision warning system. In this task, without warnings as much as 18.6 % of the subjects could not avoid collisions and the rest of them could avoid collisions successfully by braking, steering, or both braking and steering. With warnings, only 2.3 % of the subjects collided with the preceding car. Thus, the collision rate was reduced significantly (88 %)by the warning system($p < 0.05$).

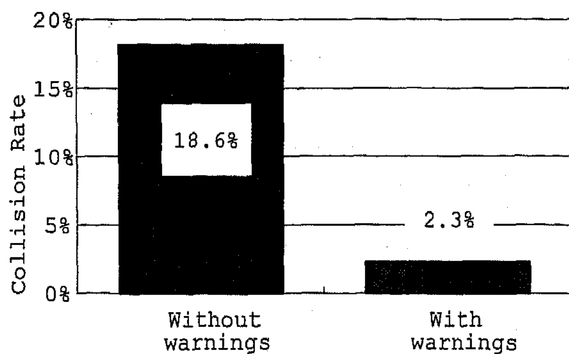


Figure 6. Collision rates with/without warnings

Figure 7 shows the collision rate for each headway group mentioned above. It shows some of the small-headway group could not avoid collision even with warnings, and the effect of warning was larger in the middle-headway group than the other groups.

Though there was no significant difference in the response time of the small-headway group between with and without warnings conditions, the collision rate of the small-headway group was much smaller with warnings than without warnings. These results indicate that the subjects in the small-headway group kept larger distances with warnings.

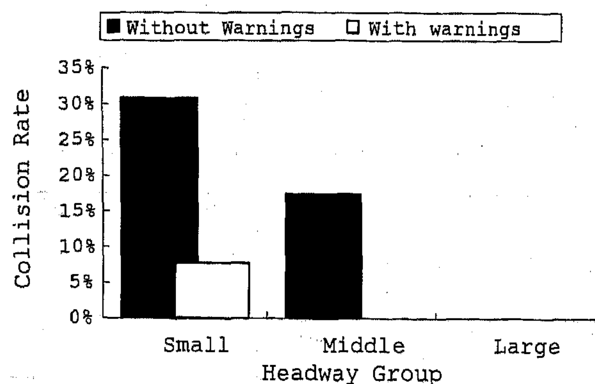


Figure 7. Collision rates for three headway groups

PREDICTION OF THE EFFECTIVENESS OF COLLISION WARNING SYSTEM ON THE ROAD

The current experiments simulated an extreme situation in which the preceding car braked at 0.8G while the driver looked off the road. Overall collision avoidance effects in general traffic situations can be predicted by assuming distributions of the deceleration and the headway distance, and by modeling the drivers' response to warnings.

Letting the initial headway distance when the preceding car starts braking at a deceleration α_l be L, and an avoidable delay time in which the following car can avoid a collision successfully by braking at a deceleration α_f after the preceding car's braking be t_a, t_a is calculated by Eq. 2.

$$t_a = \frac{L - \dfrac{V_f^2}{2\alpha_f} + \dfrac{V_l^2}{2\alpha_l}}{V_f} \qquad \text{(Eq. 2)}$$

The probabilities of collision can be calculated as the probability that the response time, which distribution is shown in Table 3, is less than the avoidable delay time t_a. Figure 8 shows the probabilities of collision when the preceding car decelerates at 0.8 G from 90 km/h(25 m/s) as the function of the initial headway distance.

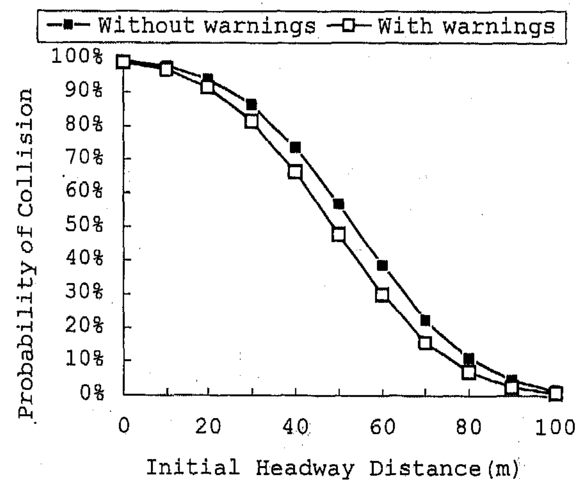


Figure 8. Probability of collision when preceding car decelerates from 90km/h at 0.8 G

To weight the initial headway distance for calculating collision probabilities, the distribution of the headway distance that the subjects kept in this

experiment shown in Figure 9 was used. And the probabilities of collision when the preceding car decelerates at 0.8 G with/without warnings were obtained as 13% and 24 % respectively. The collision reduction rate was 46 %. The estimated value for without warnings was close to 19% of the experimental result, but that for with warnings was much larger than the experimental result, 2%, and thus, the collision reduction rate was underestimated. The reason was thought that the subjects kept longer headway distances with warnings than those in this distribution.

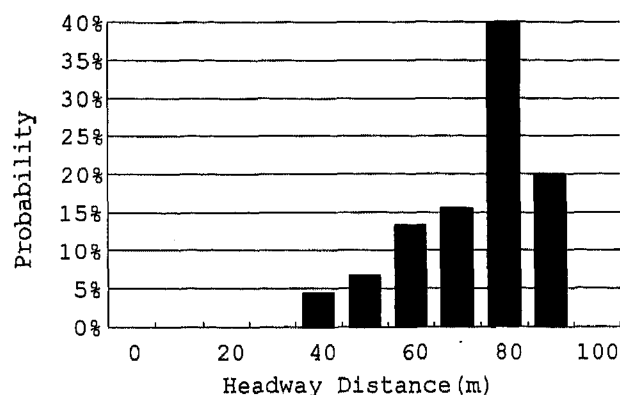


Figure 9. Distribution of the headway distance of the subjects in car following at 90km/h(25m/s)

Taking the same approach, the probabilities of collision with/without warnings for different decelerations of the preceding car were obtained as shown in Figure 10. This result indicates that as the preceding car's deceleration becomes larger, the probability of collision becomes larger for both with and without warnings and that the probabilities of collision with warnings were about a half of those without warnings.

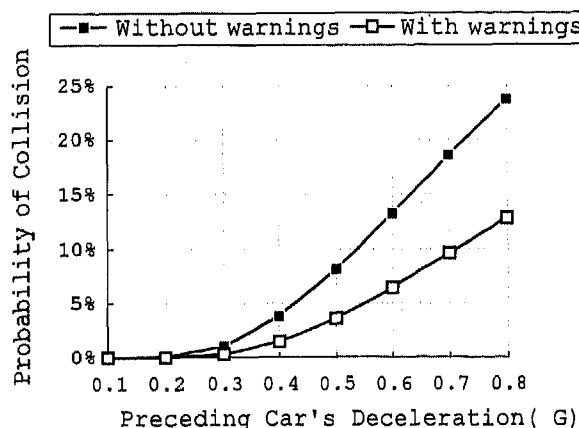


Figure 10. Relationship between probability of collision and deceleration

Using a distribution of the deceleration in a real traffic on the Metropolitan Highway in Tokyo: its average is 0.17 G and its standard deviation is 0.10G, the overall probability of collision while the driver looked off the road with/without warnings were obtained as 0.22 % and 0.58 % respectively. And it was estimated that collision warnings would reduce about 62 % of rear-end collisions. It must be mentioned here that the figures of the collision probability refer to the cases in which the driver looks off the road.

This prediction didn't consider drivers' behavioral changes; e.g. keeping longer distances with warnings than without warnings. As the longer headway distance would further reduce the probability of collision, if such behavioral changes can be modeled after obtaining enough experimental data, collision avoidance potential of a warning system would be much larger than this prediction.

CONCLUSION

With a collision warning system, the subjects were able to respond more quickly even when they looked off the preceding car; a collision reduction rate of 88 % was obtained in the simulator experiment.

However the experimental condition was an extreme case in which the preceding car decelerated at 0.8 G; therefore, assuming a distribution of deceleration in real traffic, an estimated collision reduction rate of 62 % was obtained. This value was close to that of the previous research which estimated that a collision warning system would reduce 60 % of rear-end collisions[2].

The prediction did not take into account the increase of headway distances with warnings which was observed in the simulator experiment. Therefore, if the driver's behavioral changes are modeled precisely after further experiments, the collision reduction potential of warnings would be substantially larger than the predicted value.

As the next step of research, we should optimize the warning algorithm and parameters. In this process, the collision rate will not be an effective measure, because with warnings the number of collisions will be too small to be analyzed statistically. So, a new quantitative measure to evaluate the safety margin for collision avoidance needs to be developed.

REFERENCES

[1]Najm ,W.G. et al., "Comparative Assessment of Crash Causal Factors and IVHS Countermeasures", Proceedings of the 14th International Technical Conference on the Enhanced Safety of Vehicles, p.412-421, 1944.

[2]Farber, E., "Using the REAMACS Model to Compare the Effectiveness of Alternative Rear End Collision Warning Algorithms", Proceedings of the 14th International Technical Conference on the Enhanced Safety of Vehicles,pp.400-407, 1994.

[3]Haab, K. et al., "Driver Assistance Systems for Lateral and Longitudinal Vehicle Guidance -Heading Control and Active Cruise Support-", Proceedings of Advanced Vehicle Control 94, pp.449-454,1994.

[4]Graham, R. et al., "The Effect of a Collision Avoidance System on Drivers' Braking Response", IVHS America 1994, No. 94013,1994.

[5]Suetomi, T. et al., "The Driving Simulator with Large Amplitude Motion System", SAE Paper No. 910113, 1991.

[6]Suetomi, T. et al., "A Study of Collision Warning System Using a Moving-base Driving Simulator", 2nd ITS World Congress, 1995

970286

Single Shaft Parallel Hybrid Drive System

Knut Welke, Martin Ochs, and Hans-Joachim Schmidt-Brücken
Fichtel & Sachs AG

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A joint venture of four German companies was build to develop and realize a hybrid car with a aluminium space frame. This paper summarizes the project from idea to prototype car. To find out the right drive system a comparison of alternatives was made. The result shows a clear advantage of the „single shaft parallel hybrid system with automated shift gearbox". The electric motor is a brushless permanent magnet synchronous machine with an outer rotor construction and liquid cooling system. The automated shift gearbox has a hydraulic actuator. The entire system is controlled by micro controllers.

INTRODUCTION

The goal of every automotive development project is to reach the level of mass production. A major step is to build up prototypes. A good prototype should be driveable by any person, not only by well trained specialists.

This paper summarizes the development process from an idea to a prototype car, which covers a time period of approximately 2.5 years. The car called uni1 (see figure 1) was introduced to journalists on November 6th, 1996 in Dresden by the involved German companies(2). It is planned to go ahead to mass production within 3 years.

Figure 1: The uni1 prototype car



MAIN SECTION

THE HISTORY - At the beginning of the project there was the idea to build up a car with a parallel hybrid drive system and a variable structure. The potential customers are delivery services, taxis, postal cars and other companies for urban traffic.

The target was to achieve limited zero emission capability without net fuel economy penalty. But also the possibility to drive between cities without range constraints was very important. There was no plan to develop a new combustion engine or gearbox. The fuel consumption was calculated according to the new ECE duty cycle. For the urban part (4.052 km) a power consumption of appr. 20 kW/100km was calculated, for the non urban part (6.955 km) 5.3 litre/100km was taken(VW value for Sharan TDI). With the equivalent of 5 kW per litre the overall fuel

consumption was calculated to 4.8 litre/100km. For a bigger urban part in the real duty cycle the fuel consumption will be even lower.

A joint venture of four companies was build to develop such a vehicle; it was supported by the Ministry of State for Economy and Labor of the State of Saxony.

To identify the most appropriate drive systems a comparison of eighteen alternatives, reduced to five main principles (see figure 2), was made; the systems were compared by the following factors (see table 1).

Figure 2: Five main parallel hybrid principles

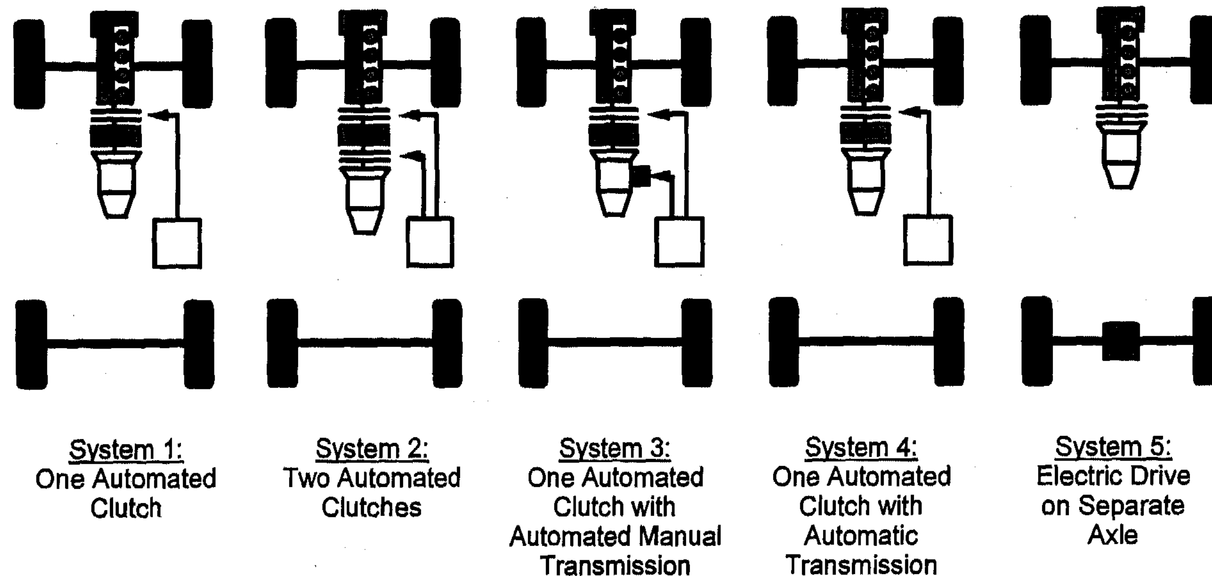


Table 1: Comparison of five main parallel hybrid principles

Evaluation Criterion's	Weight Factor	System 1: One Automated Clutch	System 2: Two Automated Clutches	System 3: One Automated Clutch with Automated Transmission	System 4: One Automated Clutch with Automatic Transmission	System 5: Electric Drive on Separate Axle
Comfort	10	1.0	3.0	3.5	3.5	1.0
Feasibility	7	1.5	2.5	2.5	2.0	3.0
Efficiency	7	2.0	2.0	3.0	2.0	2.0
Costs	5	3.0	2.0	2.0	1.0	1.0
Volume	5	3.0	1.0	2.0	2.0	1.5
Development Effort	3	3.0	2.0	1.0	2.0	3.0
Mass	1	3.0	2.0	2.0	1.0	1.0
Points		76.5	84.5	98.5	85.0	67.5
Position		4	3	1	2	5

The most important factor, which was multiplied by a factor of ten, is the comfort for the driver. The system must be easy to use for any person. Another important issue is the feasibility, which was weighted with a factor of seven. The overall efficiency is another important factor to avoid a net fuel penalty. The efficiency was roughly estimated as the possibility to allow automatic drive programs. The involved costs received a factor of five. Internal studies showed only a marginal cost difference between the various systems. Taking the costs of the batteries into consideration, the relative cost differences are negligible. Volume and mass are

only technical requirèments and their relativity depends heavily on the individual type of car.

The result of this comparison shows a clear advantage of the „single shaft parallel hybrid system with an automated shift gearbox" compared to all the other examined parallel hybrid systems (13,5 points difference). It was therefore chosen to be further developed. The reason why this system has not been developed before, is the high development effort involved in building a compact electric machine, an actuation system as well as a complex control system.

The same solution without an automated gearbox seems to be much more simple at first sight. But the synchronization rings of the transmission are unable to handle the substantially increased moment of inertia of the input shaft. Uncomfortable gear shifting (long shifting time) and shortening of the transmission life are the likely results. With the automated system the target gear is already known at the beginning of the shifting process. Therefore, the electric motor is able to synchronize the shaft speed before the gear is shifted. Due to the high acceleration capabilities of the electric machine, the shifting process can be done quickly and comfortably.

THE CAR - The potential customers for the car are delivery and postal services, taxis and urban companies. To fulfill all the customer requirements, a variable car structure and different designs are necessary. For this reason, a space frame technology was chosen (2), which allows variable axle bases and different designs (for example 1 to 7 seat van and pickup). To save weight and guarantee recycleability, aluminium was used for the hollow profiles of the space frame and for all sheet applications.

The combustion engine is the 1,9 litre VW TDI Diesel motor with a direct fuel injection system. The manual gearbox is also an unchanged high volume mass product of a big European car manufacturer.

The main data of the car is given in table 1.

Table 1: Main Car Data

Item	Value	Unit
Length	4400	mm
Weight (ready to drive with driver)	1500	kg
Maximum weight	2100	kg
Combustion engine maximum power (at 4000 rpm)	66	kW
Combustion engine maximum torque (at 1900 rpm)	202	Nm
Battery Voltage (without load)	284	V_{DC}
Battery capacity	60	Ah
Battery weight	200	kg
Maximum speed (combustion engine)	160	km/h
Maximum speed (electric motor)	100	km/h
Acceleration 0 - 100 km/h (combustion engine)	15	s
Acceleration 0 - 100 km/h (booster mode)	12	s

THE DRIVE SYSTEM - The complete drive system can be divided into four main components, which are explained in the next chapters.

The electric machine - There were three basic requirements for the electric machine. The electric motor must fit between combustion engine and manual gearbox without any rework on these mass products. The assembly of the engine-motor-gearbox system must be very easy and possible without any special tools. The electric motor must be sealed to the clutch to avoid problems with the dirt of the abrasion.

These goals are achieved with the design shown in figure 3. The electric motor is a brushless permanent magnet synchronous machine with an outer rotor construction and liquid cooling system. This machine has extremely high torque- and power-to-mass-ratios. Compared to an inner rotor construction this type of motor allows for a much smaller length, which is very important for a front drive system. The main data of the electric machine is summarized in table 2. The housing of the machine is specially designed for the mating components and can be used as an adapter for different hole pattern (e.g. if the gearbox is not designed for the used engine); the clutch is covered and sealed to the electric machine with a solid barrier.

Figure 3: The electric machine

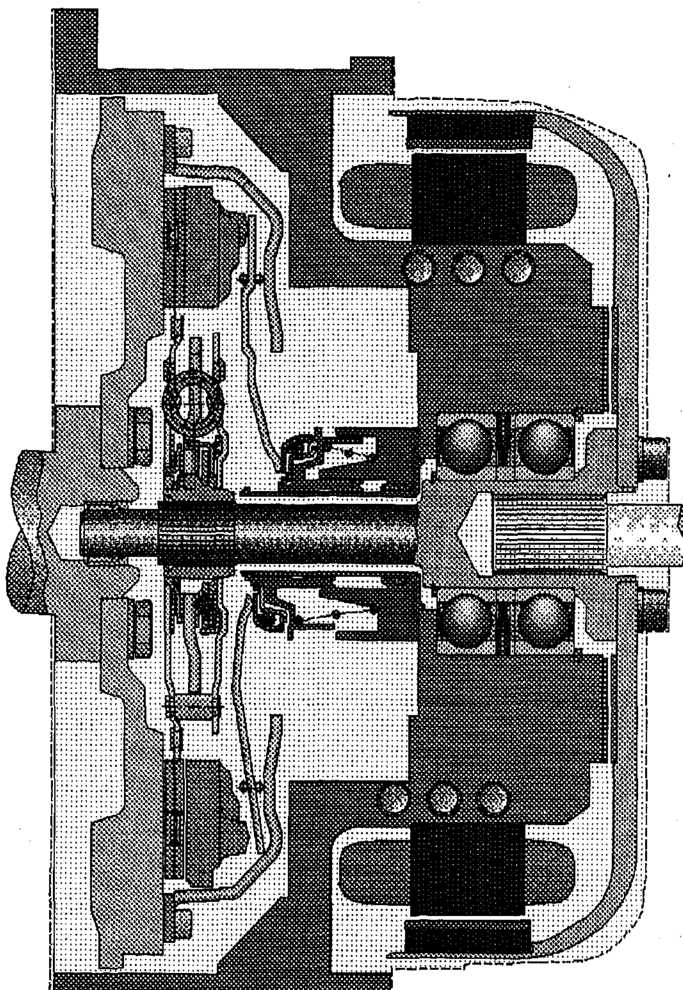


Table 2: Electric Machine Main Data

Item	Value	Unit
Maximum power	30	kW
Maximum continuous power	22	kW
Maximum torque	120	Nm
Maximum speed	5200	rpm
Maximum efficiency	93	%
Nominal voltage	200	V_{DC}
Degree of protection	IP66K	
Weight (with concentric slave cylinder)	28	kg

The rotor is fixed on a short shaft, which is guided in two ball bearings. To reduce the noise and extend the lifetime both bearings are spring loaded. Due to the forces of the permanent magnets a special device is necessary to assemble the rotor-stator unit, but in the assembled condition the motor is easy to handle. The electric motor can be guided to the rear face of block and fixed with two screws. Then the gearbox can be mounted on the electric machine and all screws can be added. The electric motor shaft is used as an extension of the gearbox input shaft and can be used as an adapter to combine different splines.

The machine is water cooled by a heat exchanging system, which is incorporated in the housing. This system is designed to minimize the pressure loss in the machine. In the prototype phase a steel pipe was wound and casted in the aluminium housing. For further motors a special casting technique is planned, which allows to make the hollow areas without additional parts; this process is based on a assembled styrofoam model, which is covered by unsolidified sand. The liquid metal melts down the foam and fills all remaining space. So a complex housing with integrated heat exchanger is possible. Using this method, the pressure loss will decrease, because the diameters inside will increase. In addition to the liquid cooling system an internal air flow system is incorporated into the structure. The „cold" air will be pumped with blades on the rotor through the gap between rotor and stator to cool down the magnets. The „hot" air will then be cooled down at the housing by flowing through the gap between rotor and housing; to complete a circle the air must flow through some holes in the rotor.

Special sintered Nd-Fe-B permanent magnets are used (3). These magnets have high energy densities and a high thermal stability. Also the target price is very attractive compared with other rare earth magnets.

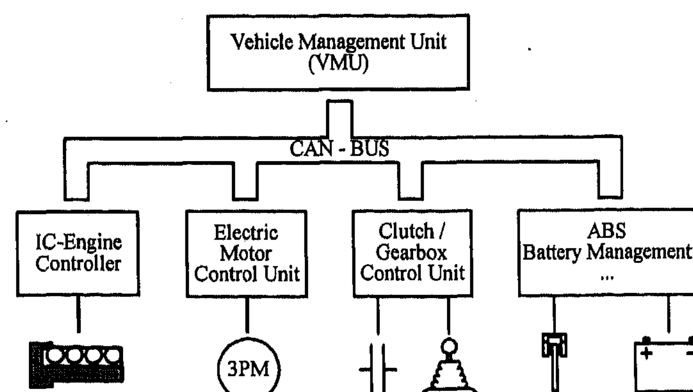
The stator has copper windings on a laminated core. All materials, including the multilayer insulation materials and the impregnation is class C. The winding temperature is sensored by three thermal couples (NTC's). Planned for the future is a assembled stator with grain oriented material at the teeth . This system allows to produce the winding of every tooth separately on a simple winding machine and assemble the stator by clicking on the tooth later. Due to the use of grain oriented material the copper area is much bigger and therefore the losses will decrease. With this manufacturing process an excellent efficiency and total automation will be possible.

The brushless resolver of the motor is also a special development for the application in automotive drives. It is made of two circuit boards with big inside diameters. The overall length of the resolver system is less than 6 mm's.

The Clutch - The clutch is a modified high volume series disk clutch. The unit of flywheel, clutch and electric motor behaviours as a dual-mass-flywheel, when the clutch is engaged. This reduces the torsional oscillations by a high degree. The change of the spring-damping-ratios is necessary to move the self-resonant frequency below the idle speed of the combustion engine. The clutch is actuated by a concentric slave cylinder, which is a modified mass product.

The automated shift gearbox (ASG) -The automated shift gearbox has a hydraulic actuator for controlling clutch, gearbox selector angle and gear shift direction (3). Hydraulic fluid from the pressure accumulator is fed through a 3/2-way proportional valve directly to the clutch slave cylinder. The clutch position is controlled by a path-sensor, which is mounted on the slave cylinder. The selector angle position is regulated by a 3/2-way proportional valve and the gear shift travel is controlled by two pressure-regulating valves. The path is sensored at the gearbox actuator in both cases.

Figure 4: The electronic management



The electronic management - The control system consists of four control units. They communicate with each other via a CAN-Bus (see figure 4) The vehicle management unit (VMU) receives the key information's, e.g. road speed, speed of combustion engine, speed of electric motor, gearbox speed, accelerator pedal position, driving mode, actual and requested gear position. It coordinates the interaction of combustion management unit, electric motor control unit, clutch and gearbox control unit. The CAN-Bus is also used to transfer data from the ABS system and the battery management or to make information available for the instrument panel. There is the possibility of combining the individual control units in a later development phase in order to simplify the whole system.

THE REALIZED FUNCTIONS - The parallel hybrid drive system with automated shift gearbox offers multiple possibilities to realize driving strategies. In the simplest mode, the driver manually selects driving mode and gear setting. This is called shift-by-wire. Additional functions like regenerative braking or charging of the batteries by the electric machine, while the vehicle is driven by the combustion engine, are possible. Another driving strategy includes automated gear shifting. This allows for an automated selection of the optimal gear setting with consideration of fuel economy, driving and shifting demands. This reduces the burden to the driver having to change his or her gear selection behavior depending on the driving mode chosen. In addition, higher fuel economy can be achieved.

Figure 5 shows an active synchronized shifting process (shifting from first to second gear in combustion engine mode). This process can be divided into 5 steps. Step 1: After receiving the shifting command both torque's must be reduced to zero. Step 2: The gearbox must be shifted to neutral by the gearbox actuator. Step 3: The input shaft must be synchronized by the electric machine. After receiving the target speed signal the electric machine is braking (or accelerating in case of a down shift) to reach this speed as fast as possible. The time to synchronize in the diagram is about 0.2 seconds. Step 4: The gear must be shifted. The actuator shifts after receiving the synchronization signal. The shifting speed is selected to guarantee a „normal" wear of the synchronization rings. Step 5: The torque must be increased. The overall time for the complete process is less than 1 second, which is a good time for a comfortable shifting process. Increasing the time lets feel the shifting as very slow and the traction interrupt is uncomfortable. Decreasing the time leads into a hard „racing" shifting process. The quick torque differences are to big to achieve a comfortable feeling.

Figure 5: Shifting process

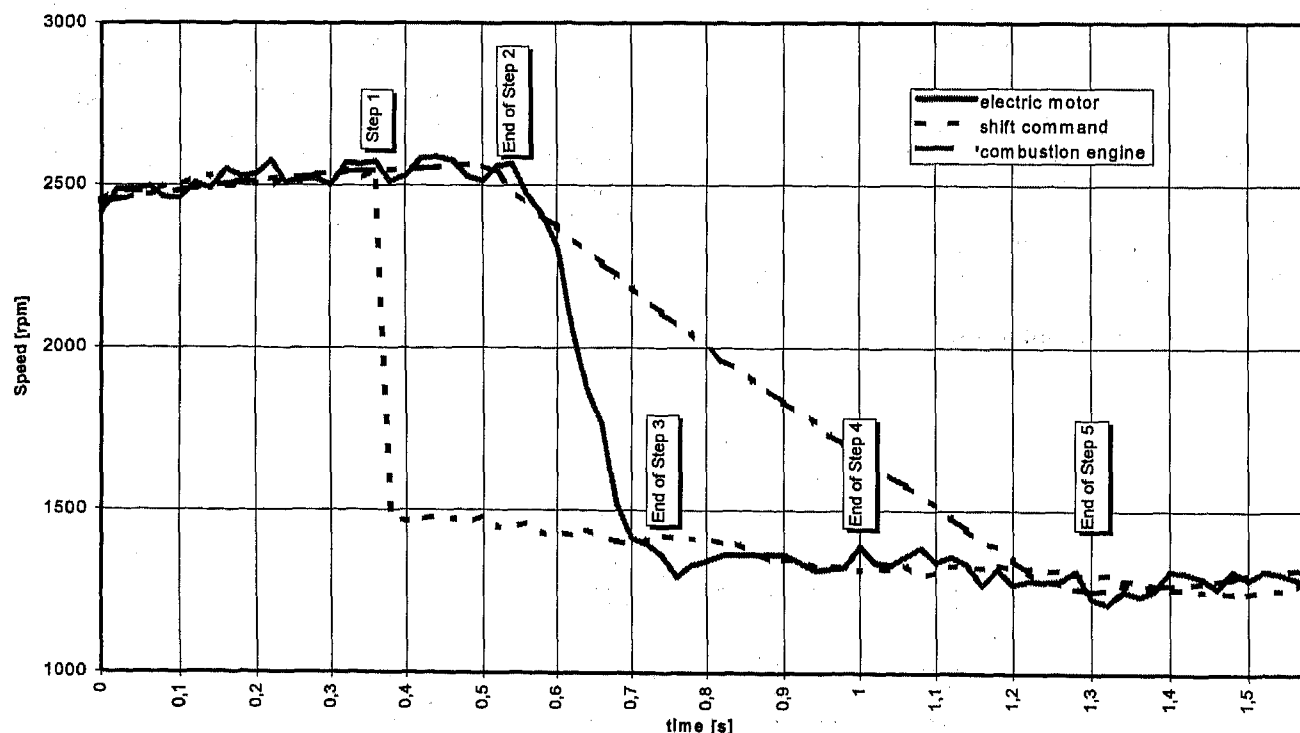
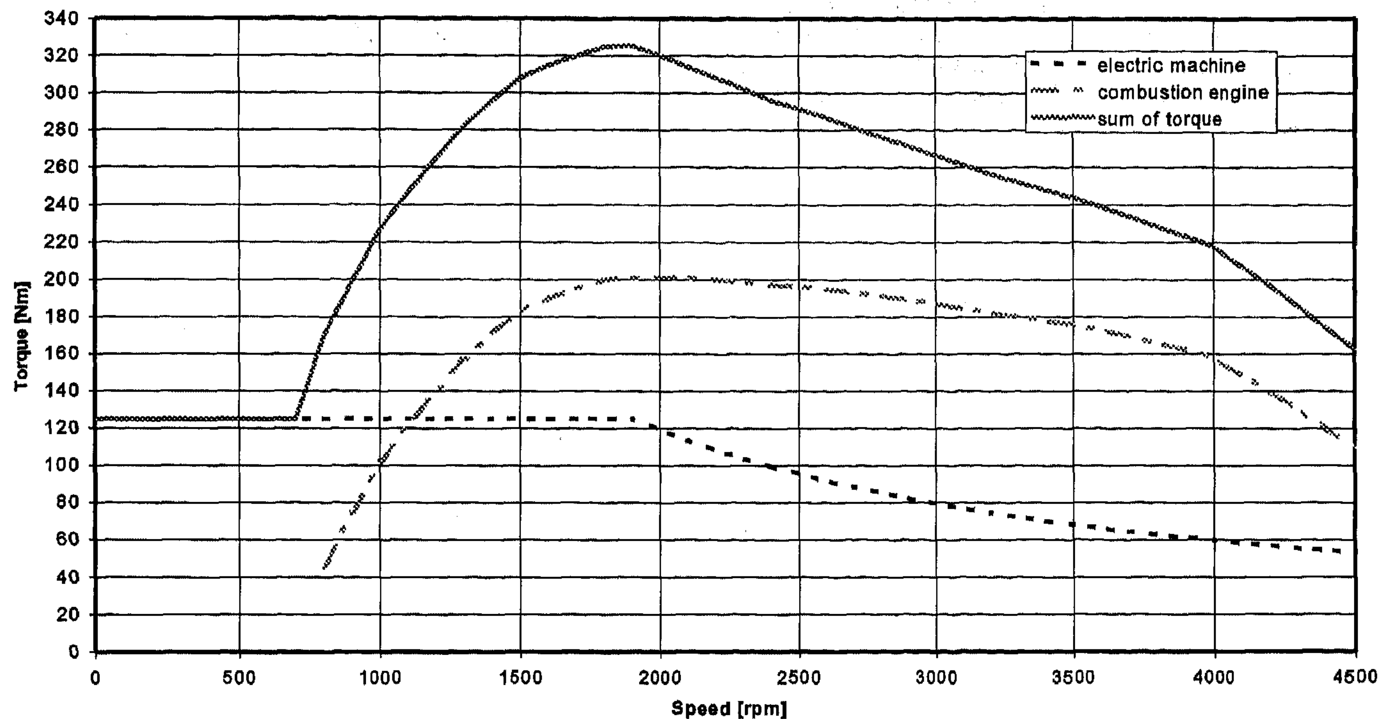


Figure 6: Torque in parallel mode



A simultaneous, parallel operation of ICE and electric motor is possible . Figure 6 shows a diagram with combustion engine torque, electric machine torque and the sum of both over the possible speed range. The maximum torque of the combustion engine (202 Nm at 1900 rpm) can be achieved in parallel mode at 900 rpm, which is only 200 rpm over idle speed. This high torque reserve of the parallel hybrid drive system allows high gear settings even at low speeds, which is desirable for fuel economy reasons. The electric motor can also support the engine during acceleration phases. The maximum torque of the system is about 320 Nm at 1900 rpm, which is comparable to a European sports car (e.g. Porsche 944 S2 has 280 Nm at 4100 rpm)

Also possible are some additional functions like a „hill-holder", which means the electric motor provide the exact amount of torque to hold the car on a hill; this is easy to realize, because all requested information is already accessible and the speed of the electric machine can be 0 rpm.

For the future it is planned to achieve a more complex strategy. It will end the strict separation of ICE-mode and pure electric mode. Depending on the demanded power, the vehicle management unit chooses the optimal drive mode. Fuel economy, driving and shifting demands, as well as frequency of switching between drive modes are being considered. The combination of automated drive mode setting as well as quick shifts into higher gears greatly increases the fuel economy.

CONCLUSION

Given the short development time, the realized prototype car is very attractive. The concept to build an alternative vehicle based on well developed series components, if they are available, is right one. The recommended variable design and the new drive system leads to an aluminium space frame technology for the car body. The hybrid drive system is also designed to use as much series components as possible. The major goals of the electric motor design are the use of combustion engine and gearbox without any machining and the easy assembly process of the drive components. The gearbox is automated by a hydraulic actuator, the clutch is actuated by a concentric slave cylinder. The entire system is controlled by micro controllers.

The realized functions and driving strategies showing the potential, what is possible with such a single shaft parallel hybrid car. The calculated fuel consumption must be verified in tests.

It is planned to go ahead with this project and realize a mass production within 3 years. The calculated number of cars is 15000 per year.

REFERENCES

(1) M. Velicescu, R. Zoller, F.J. Börgemann, D. Bauch-Panetzky, R. Sonnenburg; „Electric drives for automotive applications based on sintered Nd-Fe-B Magnets"; Proc. of 29th ISATA Florence (1996)

(2) Sachsenring Automobiltechnik; „uni1 Presseinformation" (1996)

(3) Fichtel & Sachs; „Automated shift gearbox"; (1996)

970287

Prime Movers for Series Hybrid Vehicles

F. H. Moeller
Engineering Consultant

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The known performance, emissions and fuel consumption of a VW Passat Estate car with a 1.9 liter Direct Injection Turbocharged and Intercooled Diesel Engine is compared with present and future power train concepts over a typical trip distance of 20 km at speeds between 20 and 130 km/h. (12.5 to 81 MPH)

The results are then used to discuss the proposals for future cars to change to electric transmission and to determine the power envelope for Series Hybrid Car Prime Movers. Various prime movers are also compared and the final choice is presented.

INTRODUCTION

PROBLEM: Most vehicles require their full wheel power only for very short bursts, i.e. for acceleration and hill climb. Additionally there is a human characteristic associated with cars, which is the desire to own and display more power than is actually ever required. This trend of greed for power is set to continue, in spite of more severe speed restrictions in the future. It is one of the main reasons for the fast depletion of oil reserves and for the global pollution problems we are facing.

The fact is, that all vehicles (including off road) are becoming more powerful and most of the prime mover development is concentrated on IC engines above 60 kW power output. The dilemma for the engine designer is, that an engine for a conventional vehicle is required to be powerful, fuel efficient and virtually emission free, not only at maximum power, but also at idle and anywhere in between. The prime mover has to have a virtually infinite power range.

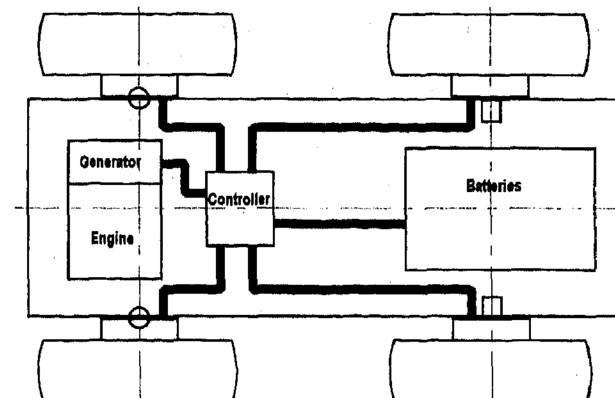
The latest Direct Injection Diesel (DDI) Engines for cars have drastically cut average fuel consumption and emissions, but with an electric transmission and some energy storage in the vehicle we can do away with this need for extreme engine flexibility and save much more energy. Also we can eliminate the typical soot cloud eminating from car exhausts, when flooring the throttle after a low power run.

However, if the prime movers were to become smaller in output and size, their combustion chambers would be less efficient and if alternatively the number of cylinders would be reduced, bad vibration and noise problems would occur with unbalanced engines. These are the topics to be dealt with in this paper.

HYBRID VEHICLE: The Series Hybrid Drive vehicle is an ideal solution, to solve the prime movers problem of limited power range. The engine only requires to produce an average power.

Its required effective power range can be reduced from infinite to around 1 to 6 and in hot climates to even less.

The required maximum engine output can be lowered to approx. one third, engine efficiency increased and harmful emissions drastically reduced.



Schematic of Series Hybrid Car with 4 wheel motors

FIGURE 1

However, as these are relatively new developments, we should at first decide on how much "average" power will actually be required and how we can design a much smaller, smooth running engine which in the end must cost less than present prime movers. Also we need to forecast the fuel savings and emission improvements that can be made.

ALTERNATIVE PRIME MOVERS FOR HYBRIDS

The present alternative power for electric cars could be:

1) Energy storage only (batteries or flywheel)
2) Fuel cells + energy storage
3) Rotating generator + energy storage, with the prime mover alternatives being:
    - 3.1) Gas turbine
    - 3.2) Gasoline or Spark Ignition engine
    - 3.3) Diesel or Compression Ignition engine

BATTERIES or flywheels on their own do not have sufficient capacity or are simply too expensive at present, to be acceptable as the only power source for a family car.

FUEL CELLS continue to make progress, but the most advanced system, the polymer fuel cell requires hydrogen as fuel. The intermediate temperature solid oxide fuel cell can operate on methanol, but this system is only just beginning to be scaled up to the multi-kilowatt size. [1]

Gas turbines, with the low power output required here, tend to have very low basic efficiencies. They suffer from the scaling down effect even more so than internal combustion engines. If additionally they have to cope with part loads down to 1/6th of full power, their average efficiency would be such, that there would be no fuel saving over cars with a conventional transmission.

Small direct injection four stroke gasoline engines could quite nicely prove the concept of Hybrid Cars in general, especially if a two cylinder opposed piston layout was used. (BMW, Citroen, Velocette, Renault)

However, at the 1996 International Seminar on the "Application of Powertrain and Fuel Technologies to meet Emission Standards" where the discussions centered mainly around present and future car engines [2], it was concluded, that with low cost options, only the Diesel engine will easily meet future emission requirements. The handling of gasoline alone creates evaporation problems, which can only be eliminated by keeping the gasoline under pressure at all times. Also there is the much disputed poisonous benzene and other HC's, of which there are significant amounts in today's lead free gasoline. These can escape from present gasoline tanks and are emitted from the exhaust during the engine warm up phase. Furthermore the efficiency of gasoline engines, even with lean burn engines, is not likely to reach the levels of well designed Diesels, using high pressure fuel injection equipment.

This leaves the compression ignition engine as the best choice. We just have to make it smaller.

Scaling down conventional engines does not work, as there would hardly be any cost saving and their combustion chambers would be much too small for good efficiency and emissions.

The alternative proposed is a Completely Balanced Single Cylinder Compound Diesel Engine with Direct Fuel Injection.

MEASUREMENTS AND COMPARISONS

The fuel consumption of a VW TDI Passat was measured for distances of 20 km, at various average speeds, using the on board trip computer. This was then used to calculate mean power requirements and engine efficiencies. At the same time comparisons with other conventional engine efficiencies were made.



FIGURE 2 Photo VW Passat Estate

They were then also compared with two theoretical models of Hybrid Cars of the same size and weight, one using a gas turbine, the other the proposed single cylinder engine.

The data of the test vehicle are as follows:

VW Passat TDI Estate

weight	1360 kg (with driver)
engine power	66 kW at 4000 RPM
torque	202 Nm at 1900 RPM
engine bore/str.	79.5 / 95.5 mm, 4 cyl.
overall transm. ratio	45 km/h per 1000 RPM
vehicle top speed	175 km/h
frontal area	2.3 m², cd = 0.3
acceleration	80 to 112 km/h in 12.4 sec
quoted fuel cons. at 90 km/h	= 4.2 l/100 km

The fuel consumption at various steady speeds was measured with a warm engine. This fuel consumption compared very well with the car manufacturers claims.

They were then also measured for typical distances of 20 km at an ambient temperature of 5 ° C, starting with a cold engine each time and then averaged out, to give a meaningful result.

The largest influence on fuel consumption for a typical trip was found to be the time taken for the engine to reach working temperature. At temperatures colder than 5° C fuel consumption increases considerably and in the summer months it was very much lower.

The second largest influence was the strength and direction of the wind. A strong head wind would increase fuel consumption by up to 20%.

Also the type of road and traffic conditions, had a big influence. The driving style adopted was with reasonably brisk acceleration, but cruising in top gear wherever possible.

EFFICIENCY CALCULATIONS: The next step was to determine the mean engine efficiencies of this VW TDI engine, at various vehicle speeds. This was simply done by calculating the required engine power, using the following formula:

$$P = 12.8\,(V/100)^3 \times Cd \times AF / TTE + W*Cd*V/360$$

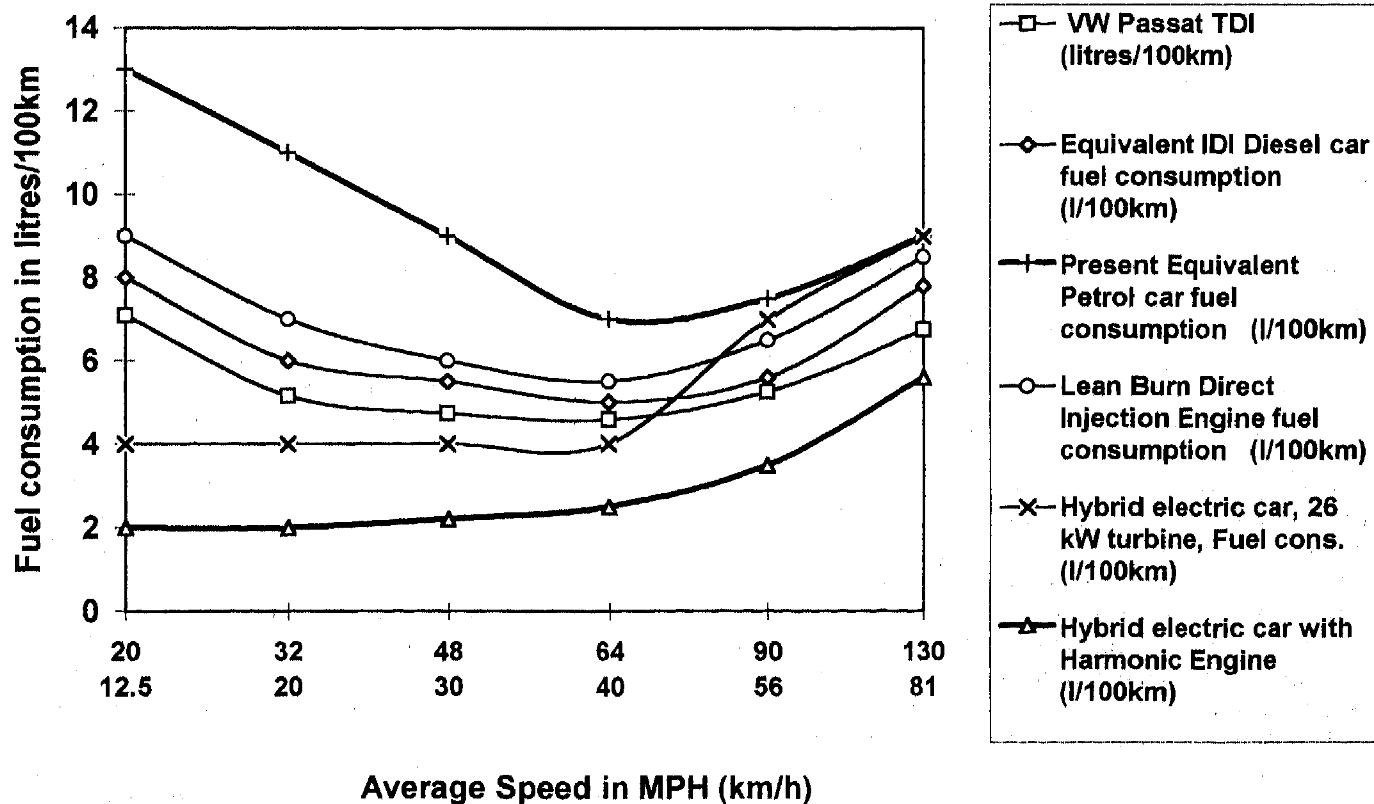
whereby
P is the power in kW
Cd the cars drag coefficient = 0.3
AF the cars frontal area = 2.3 m²
TTE the total transmission eff. = 0.86
W the weight of the car = 1360 kg
RF the rolling friction coeff. = 0.015
V the car speed in km / hour

Converting the fuel consumption to g/h and dividing this by the required engine power results in the generally quoted specific fuel consumption unit of g/kWh, which is inversely proportional to efficiency, whereby with average Diesel and Gasoline Fuels

100 % Eff. = 81.9 g/kWh

FAMILY SIZED HYBRID ELECTRIC CAR, COMPARISON DATA, USING VW PASSAT TDI AS BENCHMARK						Figure 3	
20 km trip, average speed (MPH)	12.5	20	30	40	56	81	Mean
(km/h)	20	32	48	64	90	130	64
VW Passat TDI measured fuel cons., steady speed (miles / Imp.Gall.)	58	72	78	82	68	48	67.67
(litres/100km)	4.89	3.94	3.64	3.46	4.17	5.91	4.19
For 20 km trip, (including warm up ; speed variation ; hills) (MPG)	40	55	60	62	54	42	52.17
VW Passat TDI (litres/100km)	7.09	5.16	4.73	4.58	5.25	6.76	5.44
Theoretical engine power requirement (level) (kW)	1.22	2.15	3.86	6.32	12.46	29.70	6.29
Theoretical specific fuel cons.(steady speed) (g / kWh)	676.18	492.7	380.4	294.4	252.0	216.7	385.40
Engine efficiency (steady speed) %	12.1	16.6	21.5	27.8	32.5	37.8	24.73
TDI Engine efficiency, 20 km trip (mean) %	8.4	12.7	16.6	21.0	25.8	33.1	19.59
All following data are mean and apply to 20 km trip distance							
Equivalent IDI Diesel car fuel consumption (l/100km)	8	6	5.5	5	5.6	7.8	6.3
IDI Engine efficiency %	7.4	10.9	14.2	19.3	24.2	28.6	17.45
Present Equivalent Petrol car fuel consumption (l/100km)	13	11	9	7	7.5	9	9.4
Petrol Engine efficiency, %	4.8	6.3	9.3	14.6	19.2	26.4	12.0
Lean Burn Direct Injection Engine fuel consumption (l/100km)	9	7	6	5.5	6.5	8.5	7.1
Lean Burn Direct Injection Engine efficiency %	7.0	10.0	13.9	18.6	22.2	28.0	16.0
Hybrid electric car, 26 kW turbine, Fuel cons. (l/100km)	4	4	4	4	7	9	5.33
Gas turb.Hybr.Car,engine efficiency %	14.8	16.4	19.6	24.1	19.4	24.8	19.84
Hybrid electric car with 30 kW Harmonic Engine (basic) (l/100km)	1.8	1.8	2.0	2.3	3.4	5.5	2.80
Hybrid electric car with Harmonic Engine (l/100km)							
(also allowing for heating and battery losses)	2	2	2.2	2.5	3.5	5.6	2.97
Proposed Harmonic Eng. Hybrid car, engine efficiency %	29.6	32.8	35.6	38.5	38.8	39.9	35.86
Fuel Cons. Improvement Factor, compared with VW TDI	3.55	2.58	2.15	1.83	1.50	1.21	1.83
Fuel Cons.Impr.Factor,compared with gas turbine Hybrid car	2.00	2.00	1.82	1.60	2.00	1.61	1.81
Emissions (all better than worldwide standards, Hybrid car emissions can be Zero over range of 50km at 48km/h)							
CO_2 produced per 20 km trip (kg) (3.15 kg/kg fuel consumed)							
VW Passat TDI	3.76	2.73	2.51	2.43	2.78	3.58	2.88
Gas Turbine Hybrid Car	2.12	2.12	2.12	2.12	3.71	4.77	2.83
Compound Harmonic Engine Hybrid Car	1.06	1.06	1.17	1.33	1.86	2.97	1.57
NOX produced per 20 km trip (g)							
VW Passat TDI (with retardation and oxidation catalyst)	2	2	3	4	6	8	4
Gas Turbine Hybrid Car (low pressure ratio turbine)		Probably negligible					

Fuel Consumption over average speed, 20 km trip   Figure 4



Average Speed in MPH (km/h)

Engine Efficiency over average speed, 20 km trip Figure 5

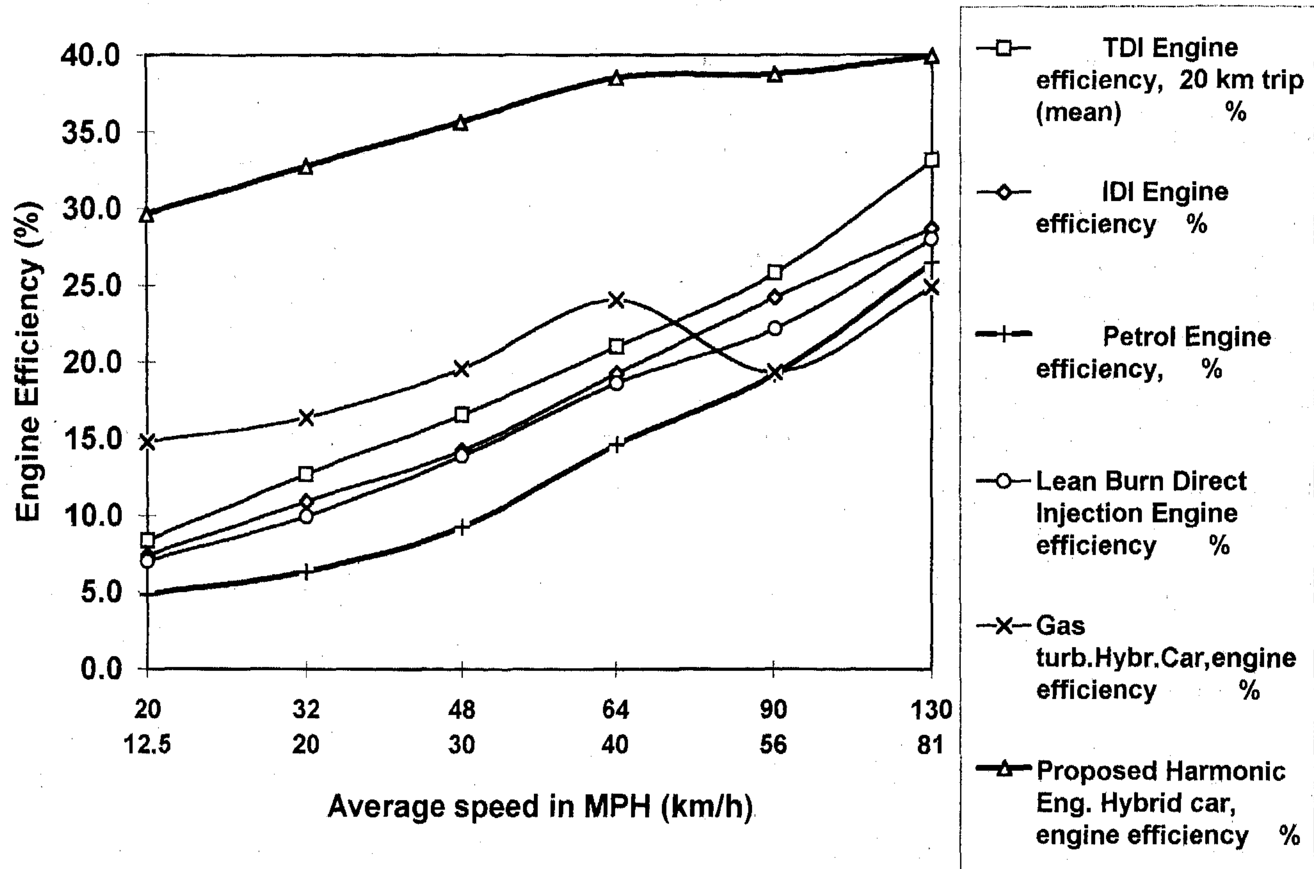


Figure 3 lists these results for the VW TDI (DDI) engine. These speak for themselves and graphs, Fig. 4 & 5 compare the fuel consumption and efficiency of all the drive lines considered. The MPG data for the alternative engines were taken from a multitude of generally available car test reports and present a reasonable average of engines presently available.

We can see, that even the DDI engine, as one of the most efficient car engines to date, has extremely low part load efficiencies. For instance, if we are driving on the level at a steady 48 km/h, when we require 3.8 kW engine power, the engine efficiency is only 21.5 %. The mean efficiency for a typical trip at an average speed of 48 km/h is lower still, at 16.6 %.

Using other prime movers, such as Diesel Engines with swirl chambers (IDI), we get 14.2% and with conventional Gasoline Engines only 9.3%.

Only the Lean Burn Gasoline engine with Direct Injection (GDI) shows reasonable results again at 13.9%, which is not much worse than the IDI Diesel. (Mitsubishi claim theirs to be equal to the IDI engine)

The point to make here is, that with all IC prime movers, especially the conventional gasoline engine, efficiency increases considerably at higher loads and if engines could be made to run at a high load rather than very small part loads, we could drastically reduce our fuel consumption, especially when traffic density limits our speeds.

This is the biggest argument for Hybrid vehicles.

ENGINE POWER FOR SERIES HYBRID VEHICLE

Power requirements are most dependent on the maximum speed the vehicle is likely to be driven at for any length of time. Long uphill gradients and head winds add to this required power.

In the end it will depend mainly on the speed restrictions which will be applied in the various countries.

If we use a prime mover power of say 30 kW and consider, that we have a series arrangement in our Hybrid Car. (Fig 1) In other words, the engine would be driving the generator, which at steady speed, would power the motors directly, with any excess power being used to charge the storage batteries (or accelerate the energy storage flywheel). The drive motors using this energy, could be mounted close to or inside at least 2 of the wheels. We then also have to assume an electrical transmission efficiency, which with the latest motors and control gear could be similar to the mechanical efficiency of a mechanical driveline, at 0.86 overall (generator 0.95, motor 0.95 and controller 0.95) (the battery is bypassed here)

Using the same car as a datum, this would then result in a continuous cruise speed of 130 km/h, or taking slight gradients and some head winds into account at least 120 km/h, which should be reasonable for future cars. Of course, short term top speeds could be very much higher and an installed wheel motor power of say 4 times 20 kW should result in very good acceleration figures and a peak top speed of 200 km/h.

PROPOSED HYBRID ELECTRIC CAR

Car size & weight, similar to VW Passat
Battery, advanced lead acid or NiCad, storage capacity 5 kWh net, max. weight = 200kg .
Two or Four drive motors with epicyclic reducers in or close to wheels,
80 kW total power, Electric Traction control, Electric Servo steering, Regenerative braking + Electric ABS
Integrated Two Stroke Diesel Engine with Brushless Generator,
Three Engine Power levels,
full load: 30 kW at 3000 RPM
part load: 12 kW at 1800 RPM
City (heating or air conditioning) : 5 kW at 1200 RPM (CHP)
(Continuous Speed and load variation should also be considered)
Max.Engine shut down period = 1h in city
Single cylinder (power) 600cc
Charge and recovery cylinder/piston 1200cc
Two Stroke Diesel cycle, max. BMEP = 12 bar
Engine in complete dynamic balance
Engine weight including generator = 80 kg

This could be the ideal car for most people. Good performance and Safety, with an infinite range, running on Diesel, Kerosene, Rape Seed or DME fuels, at a fuel consumption unheard of before and sufficiently large inside, to meet today's standards of family cars.

Regenerative braking would not require it to be especially light and in towns or cities the engine could be turned off. An additional plug in gen set (in the boot) could also make it suitable for towing trailers and caravans at reasonable speeds.

The key component for such a car is a small, smooth running Diesel Engine which should be controlled in such a way, that it either runs under load for long periods of time, or is turned off altogether.

Because the engine would generally be running at an optimum load and temperature, harmful emissions could be cut substantially. As it is smaller and running at higher specific loads, it would reach its best working temperature very much faster.

Also, depending on its utilization, it may be advisable to keep the engine warm at all times, by insulating and perhaps even electrically heating the coolant and the oil while the engine is stopped.

GAS TURBINE CONCEPT

Another theoretical model of a Series Hybrid car would be the Gas Turbine Concept, with the turbine as proposed by NoMac Energy Systems. [3]

This would be lower still as regards NOX emissions, but its fuel consumption would be very much higher.

Basic data:
Car layout and electrical concept as above
Car size & weight, similar to VW Passat.
Two Turbine Power levels assumed i.e. 26/10 kW at 96 000 RPM max, 3:1 pressure ratio, max. gas temp. 850° C, air fuel ratio 134:1, with high efficiency recuperator, weight including generator 41 kg.

The estimated fuel consumption and efficiency data of these concepts are again listed in Fig. 3, and as graphs in Fig. 4 & 5

INTERIM CONCLUSIONS:

At an assumed mean speed up to 130 km/h (81 MPH) a Hybrid Electric Transmission combined with a small prime mover and some battery or electrical flywheel energy storage could approximately half the present fuel consumption of all conventional cars, in addition to drastically cutting harmful emissions. The maximum power requirement for a Series Hybrid Car Prime Mover would be in the region of 30kW.

These two statements apply to a family sized car, regardless of the initial choice of prime mover.

DESCRIPTION OF PROPOSED ENGINE

FULLY BALANCED SINGLE CYLINDER POPPET VALVE TWO STROKE DIRECT INJECTION COMPOUND DIESEL ENGINE, WITH HARMONIC CRANKSHAFTS AND CHARGE / RECOVERY PISTON.

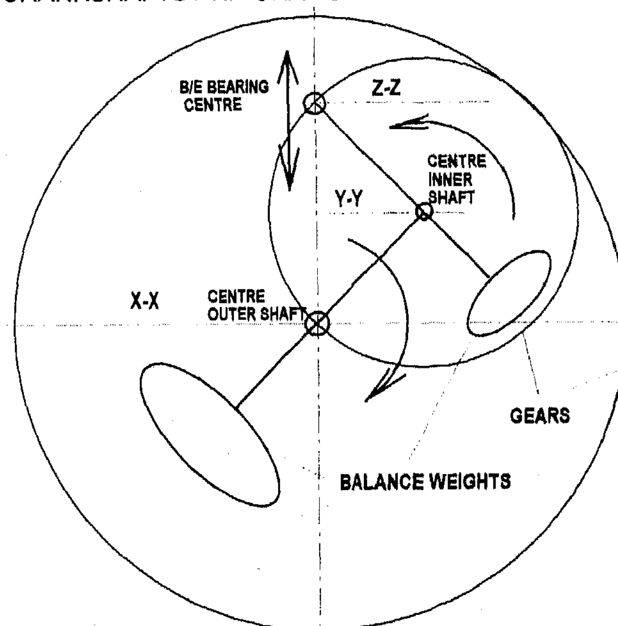
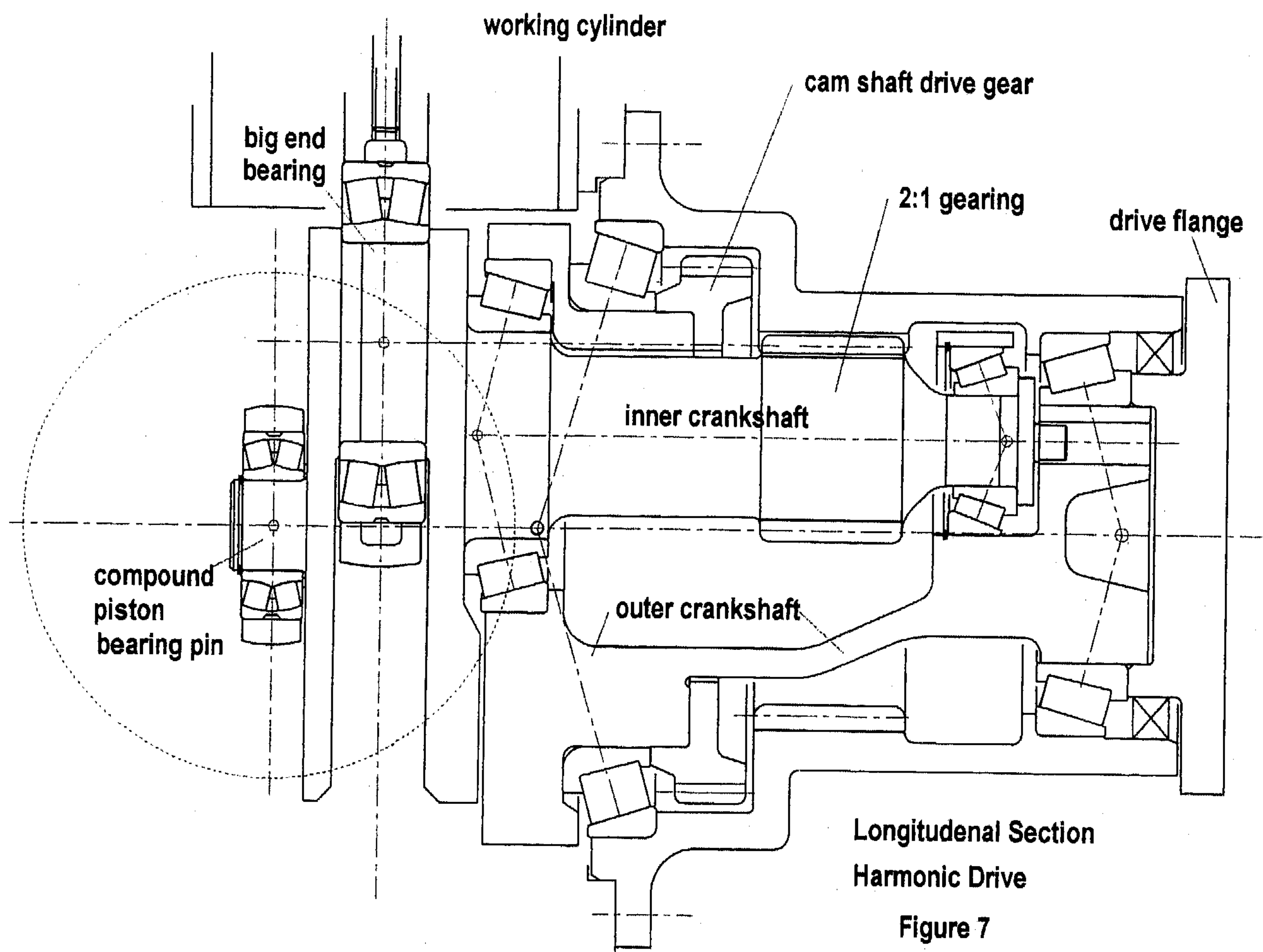


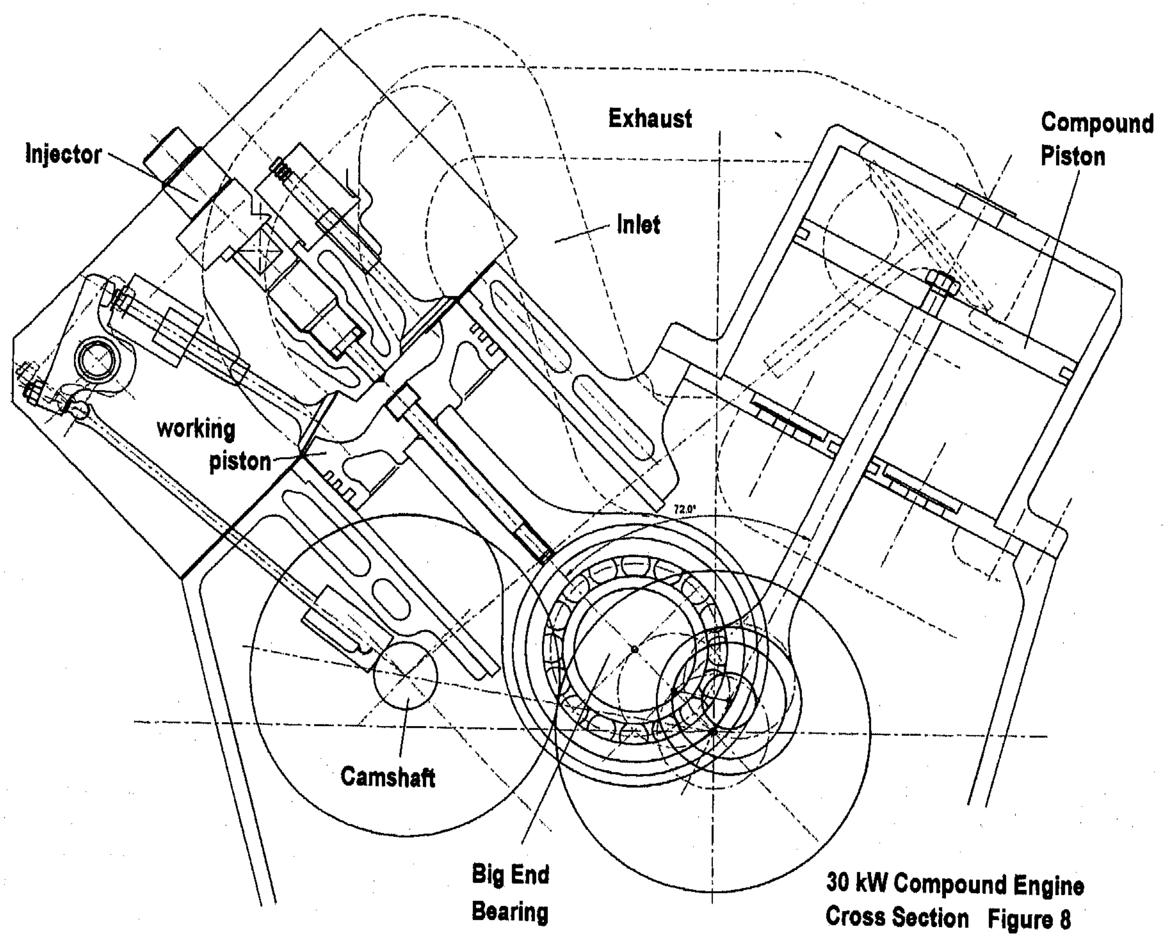
Figure 6 Basic principle of Harmonic Crank Drive

The Harmonic Crankshaft principal is not new and first proposals date back to the 19th century. There are two crankshafts mounted inside each other. They are counter rotating at the same speed and have equal crank radii, causing the big end journal, which is the crank of the inner shaft, to move in a straight line, so that no small end is required in the piston. The balance weights on the inner crankshaft counter balance the inner crank and the piston assembly. The weight on the outer crankshaft balances the whole of the inner crankshaft assembly, including the piston. The 2 : 1 ratio gears are synchronizing the two crankshafts.

Canti levering of the crankshafts and the arrangement of the balance weights are the subject of new patents. The application of these patents make a cost effective design possible. For driving the Compound Piston, a smaller canti levered pin is used, which also describes a straight line, but at a different phase angle.



Longitudenal Section
Harmonic Drive
Figure 7



30 kW Compound Engine
Cross Section Figure 8

Drawings of the complete engine are shown on Fig.7 and 8 (cross- and longitudinal section), which also provide a section through the poppet valves [4] and combustion chamber. It is here, where most of the development is taking place and the drawings do not show a final design.

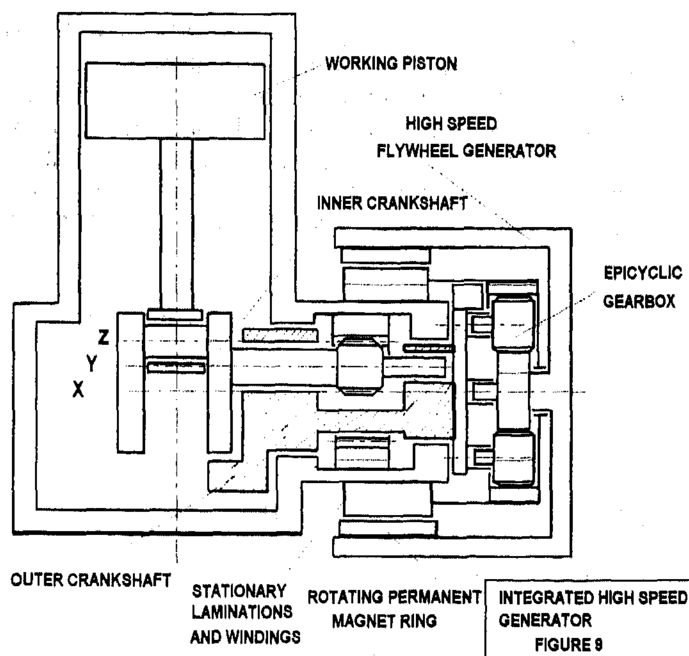


FIGURE 9

Figure 9 (schematic drawing) shows the principle of integrating a brushless generator with the engine. This drawing also shows a speed increasing gearbox, which is used to run the permanent magnets at higher speeds to generate the required current with a relatively small rotor.

PRESENT STATE OF DEVELOPMENT

There is no new technology presented here. The complete concept of a Series Hybrid car with the proposed small Diesel Engine is merely bringing present technologies together. The data presented is based on using the current level of technology on batteries, motors and generators, electronic control, shafts and gears, bearings, poppet valves, combustion chambers and fuel injection equipment. Further improvement of all these items will show better performance still.

The Cantilevered Harmonic Crank Drive has been tried in a single cylinder 4 stroke Diesel engine and has shown superior running qualities as compared with a conventional single cylinder Diesel engine. The two stroke poppet valve engine is described in a Ricardo report [4] and a Diesel version is presently under test, showing very good power levels and good efficiency. The results of these tests are partly used as a basis for the data in Figure 3.

SUMMARY OF ADVANTAGES OF SERIES HYBRID CAR WITH PROPOSED ENGINE

1) Average efficiency vastly increased, according to our predictions, we could virtually half the fuel consumption and CO2 (Fig.10) emissions of the present best Diesel cars.

2) Zero Emission range = 50 km at 48 km/h (30MPH) for local and city driving. (The possibility of Mains Charging and not running the engine, has not been taken into account)

3) Very low emissions, when the engine is running. With fuel injection, Two Stroke NOX Emissions are lower than Four Stroke. Optimum load and temperature conditions ensure very low CO, HC's and Particle emissions.

4) Infinite range as with conventional cars at average speeds up to 130 km/h (81 MPH).

5) Three engine power ranges i.e. 5, 12 & 30 kW automatically switched or varied, according to drive power or battery charge requirements and occupant heating / air conditioning, at low mean vehicle speeds.

6) Fuel can be conventional Diesel, Kerosene, Paraffin, Vegetable oil, Dimethal Ether and with Diesel pilot injection even LPG or Methane.

7) Driveability, performance and safety of vehicle is vastly improved. (ABS & Traction Control)

8) The regeneration of the cars static and kinetic energy, in brake mode, will make it less important to design for minimum total weight of the car.

9) The removal of the conventional transmission, differential, prop or drive shafts, axles and disc brakes could, on final analysis for mass production, balance the cost (and weight) of the additional batteries and drive motors. The cost of the single cylinder engine, complete with generator and control should be no higher, than present Diesel engine costs.

FINAL CONCLUSIONS

♣ The series electric Hybrid car concept, with any type of prime mover, even a relatively inefficient gas turbine, is the only concept to date, which can combine very good fuel consumption and low emissions with present and future requirements of improved drive power, traction, safety, driver and passenger comfort and driveability.

♣ Because of the relatively low power requirement of the Hybrid Car Prime Mover, propably the most cost effective and fuel efficient engine for such a Hybrid car is a single cylinder Diesel engine with an efficiency potential of more than 40%. This therefore is the best choice, if low CO2 emission levels are a prime requirement. Based on our projections, all the other emissions from such an engine could be well below future emission standards.

♣ Vibration (and noise) levels of unbalanced single cylinder engines are not acceptable and the cantilevered Harmonic Crank Engine principle could be the most cost effective way to balance such an engine in any order and plane.

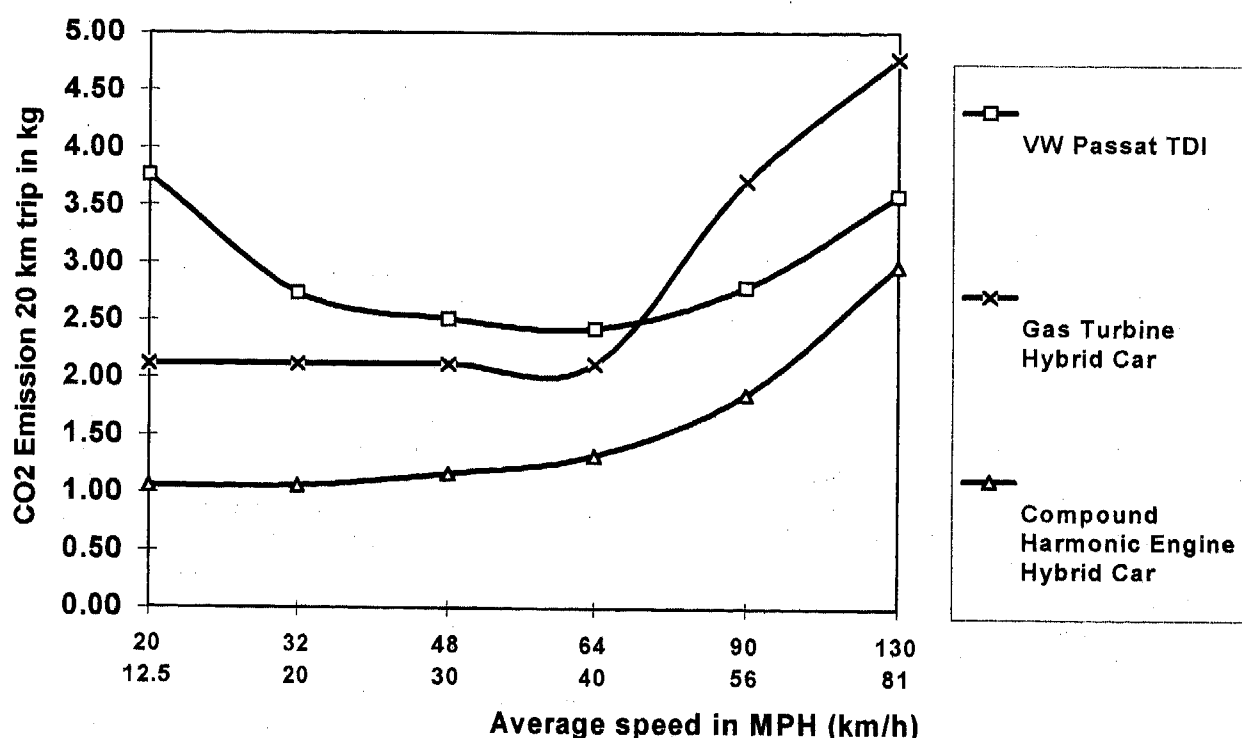
♣ The Engine proposal here also combines the high power density of the Two Stroke principal with a charge and exhaust gas energy recovery piston and cylinder, which increases the efficiency potential of the engine and considerably reduces noise levels.

This piston is added without affecting the perfect balance of the engine. No turbocharger is required to increase engine performance. The ratio of charging piston to working piston area will determine the potential output torque.

♣ The Harmonic Drive principal suits the Two Stroke Cycle, as the hitherto problematic small end in the piston is eliminated. Furthermore, there are no piston side loads, which reduces the mechanical friction and enhances the engines efficiency potential. Lubrication and cooling of the engine presents no problem, as the oil, water and combustion spaces are clearly separated.

The proposed concept makes it possible, to reduce the average fuel consumption of a family sized car to 2.5 liter per 100 km. (*)

(*) Calculated



References

1. B.C.H. Steele
Intermediate Temperature Solid Oxide Fuel Cells
Phil.Trans. Roy. Soc.A (1996), 354, 1695 - 1710

2. Powertrain and Fuel Technologies
IMECHE Seminar, London, 24-26 June 96

3. R. Mackay
Development of a 24 kW Gas Turbine-Driven Generator Set for Hybrid Vehicles
SAE 940510

4. J. Stokes, G.E. Hundleby, T.H. Lake and M.J. Christie
Development Experience of a Poppet-Valved Two-Stroke Flagship Engine
SAE 920778

970289

Analysis of the Fuel Economy Benefit of Drivetrain Hybridization

Matthew R. Cuddy and Keith B. Wipke
National Renewable Energy Lab

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Parallel- and series-configured hybrid vehicles likely feasible in next decade are defined and evaluated using NREL's flexible ADvanced VehIcle SimulatOR, ADVISOR. Fuel economies of these two diesel-powered hybrid vehicles are compared to a comparable-technology diesel-powered internal-combustion-engine vehicle. Sensitivities of these fuel economies to various vehicle and component parameters are determined and differences among them are explained. The fuel economy of the parallel hybrid defined here is 24% better than the internal-combustion-engine vehicle and 4% better than the series hybrid.

INTRODUCTION

Automobile drivetrain hybridization (using two types of energy converters rather than just one, as conventional-drivetrain vehicles do) is considered an important step to high fuel economy. The Department of Energy has established cost-shared programs with Chrysler, Ford, and General Motors under the Hybrid Vehicle Propulsion System Program to double the fuel economy of midsized automobiles, without sacrificing performance and consumer acceptability, by hybridizing their drivetrains. The government/industry Partnership for the New Generation of Vehicles (PNGV) effort has also identified hybridization as an important step toward tripling mid-sized sedan fuel economy. Recent and ongoing work seeks both to identify the likely fuel economy gains hybrid vehicles can deliver, and to ascertain the hybrid configuration that will lead to the best fuel economy [1-6].

Tamor, of Ford Motor Co., uses energy throughput spectra of current internal-combustion-engine vehicles (ICEVs) along with Ford Ecostar electric-drive data and an idealized battery model to estimate the greatest possible benefit of drivetrain hybridization to be 50% [6]. Given a 100% efficient energy storage system and ideal control stragegy, Tamor estimates a parallel hybrid will have a combined federal fuel economy of roughly 1.5 times the fuel economy of an ICEV of similar mass and engine technology. (Combined federal fuel economy is computed assuming 55% of miles are driven on the USEPA Federal Urban Drive Schedule (FUDS) and 45% on the USEPA Federal Highway Drive Schedule (FHDS).) Further, Tamor concludes that engine and road loads being equal, a parallel hybrid is more fuel-efficient than a series hybrid. Mason and Kristiansson, however, assert that series hybrids are likely to be more fuel-efficient than parallel hybrids [7]. Initial studies at NREL using current and projected component data indicated that series and parallel hybrids have similar fuel economy potential [8].

This analysis predicts the fuel economy differences among a series hybrid, a parallel hybrid, and an ICEV of similar levels of advancement and performance, using component and vehicle data adapted from current technologies. The methods of analysis and assumptions required are presented. The dependence of the fuel economy of each vehicle upon the assumptions are presented, allowing an understanding of the various projections of hybrid fuel economy made in the literature. Sensitivity coefficients, required for the fuel economy sensitivity analysis and analogous to the "influence coefficients" discussed by Sovran and Bohn are also presented [9]. These sensitivity coefficients may be used to estimate the fuel economy of derivatives of the vehicles presented.

The National Renewable Energy Laboratory's (NREL's) ADvanced VehIcle SimulatOR (ADVISOR), was used along with data from the literature and from industry contacts to define and evaluate charge-sustaining hybrids (which may operate without wall-charging, as long as there is fuel in their tank) and a ICEV for comparison. ADVISOR was also used to determine numerically the sensitivity of each vehicle's fuel economy to changes in vehicle and component parameters. These sensitivities were then used to analyze the predicted fuel economy differences among the three vehicles.

The only vehicle figure of merit being considered here is fuel economy. We recognize that there are many other important issues to be resolved in the development of a

vehicle such as cost, reliability, and emissions. Our focus here, however, is solely on the likely potential to improve fuel economy by drivetrain hybridization. With that focus, we found the series hybrid defined here is 18% more fuel-efficient than the ICEV and the parallel hybrid is 24% more fuel-efficient than the ICEV. A 10% drop in battery turnaround efficiency (from ~88% to ~80%) causes a 1.5% drop in series hybrid fuel economy (for this particular control strategy) and a 1.3% drop in parallel hybrid fuel economy. The sensitivity to regenerative braking effectiveness is likewise small: a 10% drop in regenerative braking effectiveness causes a 0.7% drop in parallel hybrid fuel economy and a 1.0% drop in series hybrid fuel economy.

BASELINE VEHICLES

The vehicles used in this study were defined using current and projected vehicle and component data. Using NREL's vehicle performance simulator, ADVISOR, (which has been benchmarked against industry simulation tools,) the components were sized to meet performance goals, and transmission and hybrid control strategies were optimized for fuel economy subject to performance constraints. ADVISOR was then used to evaluate the vehicles' fuel economy. The vehicles in this study are shown schematically in Figure 1.

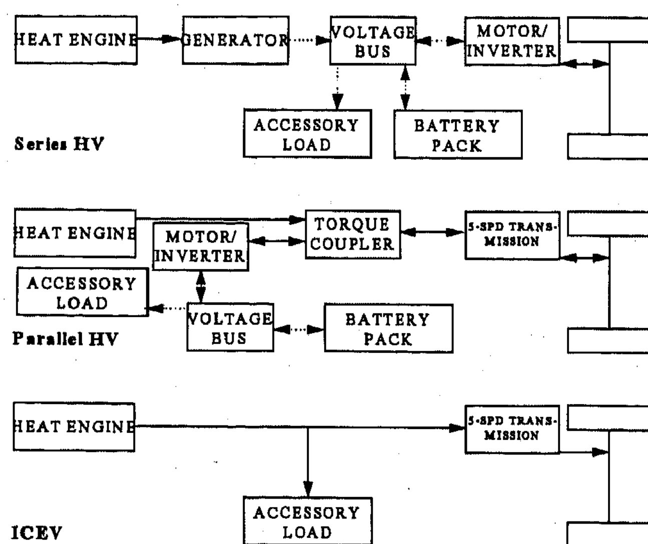


Figure 1. Energy-flow schematics of the series hybrid (top), parallel hybrid (middle), and ICEV (bottom). Solid lines indicate mechanical energy, and dashed lines indicate electrical energy.

VEHICLE SPECIFICATION

Component Specification - Table 1 lists the main component efficiencies and road load parameters assumed for this effort.

The heat engine used here is a direct-injection (DI) diesel, with a fuel-use map from the 5-cylinder, 85 kW Audi engine [10], scaled to peak efficiencies given in Table 1. See the "Scaling" section below for discussion of efficiency-versus-size-and-year considerations. The generator coupled to the diesel in the series hybrid vehicle (HV) is based on a permanent magnet motor/controller set from Unique Mobility. The traction motor/inverter set modeled here is based on the AC induction system being developed by Westinghouse. The batteries modeled here are advanced lead-acid, with characteristics adapted from Optima [11]. Vehicle drag parameters were chosen to define a Partnership for the New Generation of Vehicles (PNGV)-like vehicle with an aluminum-intensive body, heavy by Moore's standards but deemed achievable by those in industry interviewed by Duleep [2,4]. The "regenerative braking fraction" in the table is defined here as the fraction of braking energy during a given cycle that is provided to the electric drivesystem, with the balance, 60% in this case, handled by friction brakes.

Table 1. Vehicle summary data

Parameter	Series HV	Parallel HV	ICEV
Heat Engine			
type	DI Diesel	DI Diesel	DI Diesel
maximum power	32 kW	35 kW	62 kW
specific power	500 W/kg [13]	500 W/kg [13]	500 W/kg [13]
peak efficiency	43% [10]	43% [10]	46.5% [13]
avg. efficiency	40.2%	36.4%	26.0%
Generator			
type	Permanent Magnet	--	--
maximum power	32 kW	--	--
specific power	840 W/kg [14]	--	--
peak efficiency	95% [14]	--	--
avg. efficiency	90% [14]	--	--
Motor/Inverter			
type	AC induction	AC induction	--
max. continuous power	53 kW	27 kW	--
specific power	820 W/kg [15]	820 W/kg [15]	--
peak efficiency	92% [15]	92% [15]	--
avg. efficiency	86.8%	89.4%	--
Battery Pack			
type	Advanced Lead-Acid	Advanced Lead-Acid	--
maximum power	63 kW	32 kW	--
specific power @ 50% SOC (including enclosure and thermal management)	800 W/kg	800 W/kg	--
capacity	2.4 kWh	1.2 kWh	--
avg. round-trip efficiency	87.6%	87.7%	--
Transmission			
type	1-spd	man. 5-spd	man. 5-spd
avg. efficiency	98% [16]	92% [16]	92% [16]
Vehicle			
C_DA	0.4 m^2 [12]	0.4 m^2 [12]	0.4 m^2 [12]
$C_{rolling\text{-}resistance}$	0.008	0.008	0.008
accessory load	800 Welec	800 Welec	800 Wmech
regenerative braking fraction	0.4	0.4	0.4
5-6 pass. glider mass	840 kg [4]	840 kg [4]	840 kg [4]
passenger/cargo mass	136 kg	136 kg	136 kg
test mass (including driver/ passenger/cargo)	1243 kg	1218 kg	1214 kg

Table 1 also indicates propulsion component size and average energy efficiency over the combined federal cycle. Propulsion system components for each of the three vehicles were sized, using ADVISOR, in order to meet performance requirements set out by the US Consortium for Automotive Research (USCAR) for the PNGV effort [12]:

1. 0 to 96.5 km/h (0 to 60 MPH) in 12 s
2. 64.4 to 96.5 km/h (40 to 60 MPH) in 5.3 s
3. 0 to 136.8 km/h (0 to 85 MPH) in 23.4 s, and
4. 6.5% gradeability at 88.5 km/h (55 MPH).

The PNGV targets listed as 1-3 above must be attained at curb weight plus 136 kg for the driver and passenger, while the gradeability requirement is prescribed at gross vehicle weight with full accessory load for 20 minutes. We have differed from the PNGV specifications in that the gradeability requirement placed on the vehicles in this study is 6.5% at 88.5 km/h indefinitely (until the fuel runs out), with average accessory load, at curb weight plus 136 kg.

The HPU size for the hybrids is determined by the continuous gradeability requirement. Note that the HPU for the series HV is smaller than for the lighter parallel HV. This is because the 88.5 km/h requirement, for a given gear ratio, requires the parallel HV's HPU to provide adequate climbing power at a certain speed which, in this case, is not the speed at which it develops maximum power. The series HV's HPU may operate at maximum power regardless of the vehicle speed; thus, its maximum power can be set to exactly the climbing power requirement. With the HPU sized for both vehicles, the motor and batteries are sized to meet the acceleration requirements, numbered one through three above.

Scaling - In this study, the efficiency map for an 85-kW engine introduced in 1990 is used to describe the behavior of 32-, 35- and 62-kW engines [10]. For the 32- and 35-kW engines, the original map shape and peak efficiency value were maintained, while the torque axis was compressed. We acknowledge the significant technical challenge involved in achieving such high peak efficiencies with a small engine, but are encouraged by continued progress by VW/Audi (which introduced a 66 kW DI diesel with 41.8% peak efficiency two years after the 85 kW benchmark) [17]. We believe 43% peak efficiency in 32- to 35-kW diesel engines in 2005 is consistent with the Office of Advanced Automotive Technology's 2004 target of 45% peak efficiency for 40- to 55-kW engines, which is a goal at least some diesel manufacturers find reasonable [13,18]. We have assumed a peak efficiency of 46.5%, the peak efficiency of current state-of-the-art heavy-duty diesel engines, for the 62-kW engine in 2005, with its higher peak efficiency due to its larger size [13]. We expect the smaller (HV) engines to have lower specific power; the effect of changes in engine specific power can be derived using the data in Table 5.

The tractive motors in this study, at outputs of 53 and 27 kW, are significantly smaller than the 75-kW motor from which their maps come. However, motors of these lower power levels with peak efficiencies of over 92% are available now [14]. We have not attempted to scale the efficiencies up, as would likely result from further development, for lack of data.

Series Hybrid Control Strategy - The strategy chosen here was a close-power-follower strategy where the hybrid power unit (HPU) power output closely follows the tractive motor output. Figure 2 shows the behavior of the vehicle propulsion system following this strategy over the first 315 s of the FUDS. The HPU power (represented by the dots) varies directly with the tractive motor power (represented by a solid line), but is higher by a state-of-charge-dependent factor to allow for losses in the generator and battery. In this strategy, the HPU power is given by (K1*(tractive_motor _power) + K2)*(SOChi-SOC)/(SOChi-SOClo), where SOChi and SOClo are threshold SOCs.

As the third chart in the figure indicates, this control strategy leads to nearly constant battery pack state-of-charge (SOC). (See the "Results and Discussion" section for a more detailed discussion of the control strategies considered here.)

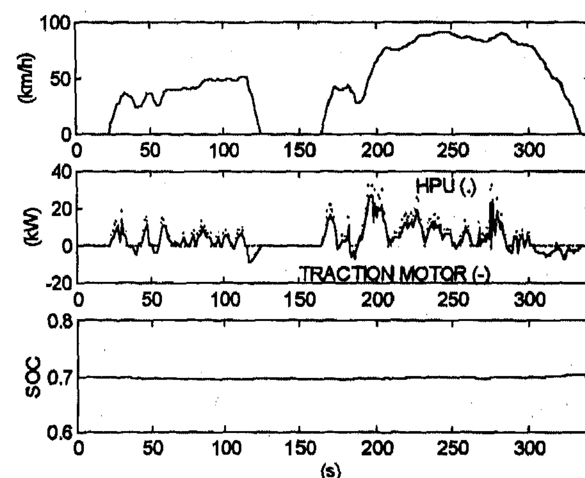


Figure 2. Close-Follower Control Strategy

We have chosen a power-follower strategy where the HPU power follows the motor power's second-by-second variation, which defines a vehicle that achieves 29.5 km/L (69.4 MPG) on the combined federal drive schedule. This strategy was chosen because:

1) it leads to the best fuel economy in the control strategy design space of power-follower approaches considered here, and

2) it requires the HPU to immediately follow tractive power requirements, as occurs in the parallel hybrid and ICEVs. This leads the fuel economy estimate of each vehicle to be overestimated roughly equally due to the consistent neglect of transient effects, minimizing the effect of transient fuel use on the differences among the three vehicles' fuel economy.

This approach likely has no emissions benefit over ICEVs, and is chosen only for its fuel economy characteristics. Alternative control strategies with potentially

better emissions characteristics are presented later in this paper.

Parallel Hybrid Control Strategy - The parallel control strategy can be defined as follows:

- The HPU does not idle (it turns off when not needed).
- The motor performs regenerative braking regardless of the batteries' SOC.
- The HPU generally provides the power necessary to meet the trace, and
- the motor helps if necessary by providing additional torque, or accepting extra torque provided by the HPU for recharging the batteries.

The parallel hybrid control strategy was defined by two parameters: a vehicle speed below which the HPU is turned off, to allow for electric launch and no idling, and minimum HPU operating torque defined as a function of engine speed, which is the lowest torque output at which the HPU would operate whenever the tractive torque requirement is positive. When the minimum-allowed HPU operating torque exceeds that required to meet the trace, the balance of torque is used to drive the motor as a generator, recharging the batteries.

Figure 3 shows a sample portion of FUDS for this parallel vehicle. Note that the motor provides power when the vehicle needs additional torque to meet the driving, as described above. During the constant velocity portion of the driving cycle (from ~90 to ~115 s), the motor power is negative as the HPU provides recharging torque to the motor. As soon as the vehicle begins decelerating, the HPU shuts off, and the SOC increases while the motor captures regenerative braking energy. Note that during the final, slow deceleration, the motor power is negative but small in magnitude, leading to only a slight rise in SOC.

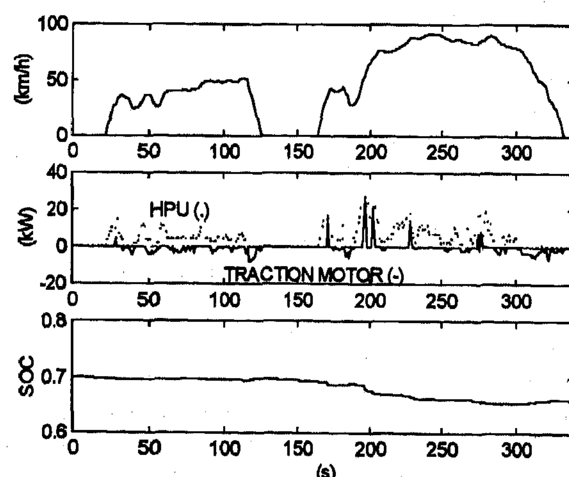


Figure 3. Parallel HV control strategy

VEHICLE EVALUATION

Vehicle Performance Simulator - ADVISOR is an empirical, physics- and map-based ADvanced Vehicle SimulatOR developed in the Simulink®/MATLAB® environment, the flexibility of which allows the development and evaluation of arbitrarily complex control strategies. ADVISOR uses fundamental equations of vehicle dynamics coupled with efficiency and/or power loss component maps to predict hybrid, internal-combustion-engine, and electric vehicle performance, range, and fuel economy. It has been compared to industry vehicle simulation programs and has been used by engineers in the hybrid program at the Chrysler Corporation. Reference eight provides more detail on this tool's application to vehicle system analysis.

Fuel Economy Calculations - The fuel economy results presented here must be understood as estimates: some uncertainty is introduced both by the model and the input data to the model. Additionally, the test procedure for determining hybrid vehicle fuel economy is a difficult problem on which debate on proper methodology continues. Due to the two energy-storage-modes of hybrid vehicles, the fuel economy calculation approach taken also may introduce uncertainty.

A Society of Automotive Engineers (SAE) task force has developed a draft hybrid vehicle test procedure which aims to properly account for changes in stored energy in charge-sustaining hybrids [19]. The unique feature of this approach is the running of two series of consecutive FUDS (for example) to determine fuel economy. The first series is started at the highest expected battery pack SOC and ends at some lower SOC, resulting in a fuel use and associated change in SOC. The second cycle is started at the lowest expected SOC and ends at some higher SOC, resulting in another fuel use and associated change in SOC. Linear interpolation is then used to predict the fuel economy estimate for the vehicle if the batteries had no net change in SOC. The benefit of using a test procedure such as this is that the uncertainty due to stored or discharged energy can be estimated by comparing the two fuel economies which bound the interpolated zero-ΔSOC fuel economy.

Because the hybrids in this study tend to reach some nearly steady SOC after a number of cycles, a different approach was used to estimate their zero-ΔSOC fuel economy. The vehicles were run over four repeated cycles (one set of FUDS and one set of FHDS) with fuel economy and SOC measured only over the last cycle, and any fuel economy measurement associated with a $|\Delta SOC| > 1\%$ (on the last cycle) was discarded. Because an exact match ($\Delta SOC=0$) is not required, some uncertainty is introduced by this method. However, we found it to be more repeatable than the SAE draft test procedure described above *for the computational evaluation of these vehicles.* It should be noted that our approach only works for hybrids whose SOC reaches some steady-state mean value over the cycle, is expected to be useful only for computational evaluation of hybrids, and is not as generally applicable as the draft SAE procedure.

RESULTS AND DISCUSSION

FUEL ECONOMY ESTIMATES - The above defined vehicles were evaluated as described, and the resulting fuel economy estimates are in Table 2. From previous comparisons with automobile manufacturers' models, we estimate the model to be accurate to ±10%, while uncertainty in the component data projections is difficult to estimate. Thus these fuel economy estimates should be treated with care. On the other hand, inasmuch as the vehicle model over- or underestimates each vehicle's fuel economy equally, the model's uncertainty contribution is bias uncertainty and does not affect the accuracy of the computed fuel economy *differences* among the vehicles.

Table 2. Estimated gasoline-energy-equivalent fuel economy estimates on USEPA cycles

Driving Cycle	Series HV	Parallel HV	ICEV
FUDS	26.6 km/L (62.6 MPG)	27.4 km/L (64.5 MPG)	21.1 km/L (49.6 MPG)
FHDS	34.1 km/L (80.2 MPG)	36.4 km/L (85.6 MPG)	32.0 km/L (75.3 MPG)
combined (55/45)	29.5 km/L (69.5 MPG)	30.8 km/L (72.5 MPG)	24.9 km/L (58.6 MPG)

DRIVE-CYCLE EFFECTS - The parallel HV is 30% more fuel-efficient than the ICEV on the urban cycle (FUDS), 14% more fuel-efficient on the highway cycle (FHDS), and 24% more fuel-efficient on the combination. The series hybrid follows the same trend, being 26% more fuel-efficient than the ICEV on the FUDS, 7% more fuel-efficient on the FHDS, and 18% more fuel-efficient on the combination. Note that both the parallel and series hybrids get similar fuel economy improvement on the FUDS, but that the parallel hybrid gets over twice the fuel economy improvement of the series on the FHDS. This is due in part to the parallel hybrid's HPU efficiently (especially on the FHDS) supplying power directly to the wheels, rather than having its output converted from mechanical to electrical and back to mechanical power, as occurs in the series hybrid.

The dependence of the fuel economy benefit of hybridization on drive-cycle is due to differences between the cycles in the number and nature of braking events, amount of idling time, and average power requirement. As braking frequency increases, so does the opportunity to recover braking energy. The more time an ICEV's engine idles in a given cycle, the more fuel may be saved by disallowing idling, as is done in both hybrid control strategies here. (Recall that the parallel vehicle powers its accessories electrically, and does not idle its heat engine). Also, the lower the average power requirement, in general, the lower the ICEV's average efficiency, and the greater the opportunity for hybridization to improve upon that efficiency [6].

The fuel economy improvement of the series and parallel hybrids over the ICEV is strongly dependent upon the driving cycle being considered. This is largely because the ICEV's fuel economy is much more sensitive to idling time and braking frequency than fuel economy of the hybrids considered here.

VEHICLE DRAG PARAMETER AND COMPONENT EFFICIENCY EFFECTS - The sensitivity of the fuel economy of each of the three baseline vehicles to various vehicle and component parameters were calculated using ADVISOR, and are presented in Table 3. The values presented indicated the percentage change in fuel economy due to a 1% increase in the given parameter (only). We took care to accurately estimate constant-performance sensitivity coefficients for each vehicle. Detailed discussion of the sensitivity coefficients is in the Appendix.

Table 3. Fuel economy sensitivity coefficients (% change in fuel economy on the combined federal cycle for a 1% increase in the given parameter)

Parameter	Series HV	Parallel HV	ICEV
average HPU eff.	1.00	1.00	1.00
average motor eff.	0.91	0.05	--
average motor-as-a-generator eff.	--	0.14	--
average battery turnaround eff.	0.15	0.13	--
mass	-0.60	-0.63	-0.75
C_DA	-0.25	-0.25	-0.22
coeff. of rolling resistance	-0.26	-0.25	-0.22
accessory load	-0.17	-0.15	-0.16
regenerative braking fraction	0.10	0.07	--

Three of the sensitivity coefficients in Table 3 require some explanation. The sensitivity coefficient for average motor efficiency indicates the percent change in fuel economy due to a 1% improvement in the tractive motor's efficiency while acting as a motor rather than as a generator. The sensitivity coefficient for the motor-as-a-generator refers to the dependence of fuel economy upon the *tractive* motor's efficiency when driven as a generator by the HPU (which occurs only in the parallel vehicle). The sensitivity coefficient labeled "regenerative braking fraction" refers to the dependence of fuel economy upon the tractive motor's efficiency when driven as a generator by the brakes.

We can use these sensitivities to estimate the effects of uncertain component efficiency assumptions. For example, the estimated 87.6% average turn-around efficiency for a high-power lead-acid battery in a hybrid vehicle may be optimistic, and we may choose a more conservative estimate of 80% average efficiency. The result of this revision is to change the fuel economy of the series vehicle by [(0.80 - 0.876) / (0.876)] x 0.15 = -0.013 = -1.3% = -0.38 km/L (-0.9 MPG), and the parallel hybrid's fuel economy changes by [(0.80 - 0.876) / (0.876)] x 0.13 = -0.011 = -1.1% = -0.34 km/L (-0.8 MPG). Of course, similar analyses can be performed with any of the parameters presented in Table 3, with good accuracy at changes of up to ±10% and reasonable accuracy at changes of up to ±20% in size.

To summarize, with the exception of mass, road load parameters have a roughly equal effect on the fuel economy of all three vehicles, and therefore do not significantly affect the differences among them. Because this series hybrid uses a close-power-follower control strategy, its fuel economy is nearly as insensitive to battery efficiency as is the parallel hybrid's fuel economy. Hybrid fuel economy is less sensitive to regenerative braking effectiveness than to all other parameters considered here except the motor efficiency for the parallel vehicle.

Comparing fuel economy estimates to previous work - Using the sensitivities presented in Table 3, we can develop an estimate of the fuel economy benefit of hybridization assuming 100% efficient components, for comparison with Tamor's estimates. We use

$$FE = FE_{table_2} + FE_{table_2} \times \sum_i \xi_i \frac{(X_{i,Tamor} - X_{i,table_1})}{X_{i,table_1}}, \quad (1)$$

where FE is the new, estimated fuel economy, FE_{table_2} is the fuel economy presented in Table 2, ξ_i are the sensitivity coefficients, and X_i are the associated vehicle parameter values. Equation 1 implies that sensitivities may be used independently, which is true to a limited extent. Because the power flows in hybrid vehicles depend upon each other and the states of the vehicle components in complicated ways, hybrid vehicle fuel economy is a *non-linear* function of drive-cycle and vehicle and component parameters. The effects of *small* vehicle-parameter changes may be combined reasonably with Equation 1, but using such an approach to combine large changes will lead to significant uncertainty in the result.

Equation 1, along with the combined federal fuel economies for the series HV and ICEV presented in Table 2, and the baseline vehicle mass, motor and battery efficiency, was used to predict the "Current" series-hybrid normalized fuel economy presented in Table 4.

Tamor used energy-throughput spectra (along with component data) to estimate the fuel economy of series and parallel hybrids relative to a lightweight Taurus using a "perfect" CVT. The hybrids' fuel economies were calculated for a range of vehicle masses and battery pack efficiencies. In this comparison, we consider one point in Tamor's range of estimates: the series HV with a 100%-efficient energy storage system and the same mass as the baseline conventional-drivetrain vehicle in his study. We choose the 100%-efficient energy storage for comparison to reduce the effect of uncertainty about Tamor's control strategy. (See the Appendix for a discussion of the battery-efficiency-sensitivity/control-strategy relationship.)

To develop a fuel-economy-improvement-from-hybridization estimate for comparison to Tamor's estimate, we begin with the series HV defined in Table 1. We then apply Equation 1 to account for the differences between the point in Tamor's space chosen above and the Table 1 vehicle. Tamor's vehicle had all braking energy done by the electric drivetrain, while in our case, only 40% of braking energy was available for regeneration. Also, the point we chose from Tamor's calculations has a 100%-efficient (round-trip) battery pack, while that of Table 1 is 87.6% efficient. Finally, our conventional-drivetrain vehicle has lower mass and a higher peak-efficiency-engine than does the series HV. The result of these modifications to the series HV computed here is shown in Table 4. We have used second-by-second simulation, component maps, and sensitivity coefficients to derive, within 5%, the same fuel economy benefit of hybridization as did Tamor using different methods, for a somewhat heavier vehicle.* This comparison is not a validation of either work, but indicates reasonable agreement using different approaches.

Table 4. Comparison of current normalized fuel economy on combined federal cycle estimate with Tamor's estimate for a *series* HV with equal mass and 100% efficient motor and battery pack [6]

	Tamor	Current
Normalized Fuel Economy (1=ICEV)	1.47	1.54

COMPONENT SPECIFIC POWER EFFECTS - It was noted above that neither hybrid vehicle's fuel economy (using the given control strategies) is particularly dependent upon battery efficiency. Let us examine the effect of battery specific power. The batteries in this comparison were assumed to have a power density of 800 W/kg, which results in a baseline battery mass of 78.4 kg for the series hybrid and 39.7 kg for the parallel hybrid. If we make the more conservative assumption of a 400 W/kg battery pack, both pack masses double. We can estimate the fuel economy effect of this change using the mass sensitivity coefficients for the two vehicles and their baseline data from Table 1. The series vehicle's mass would change by (78.4/1243)=6.3%, and with a mass sensitivity coefficient of -0.60, the series hybrid's fuel economy would change by 6.3% x (-0.60) = -3.8% = -1.12 km/L (-2.6 MPG). The parallel vehicle's fuel economy change can likewise be estimated: (39.7/1218) x (-0.63) = -2.1% = -0.65 km/L (-1.5 MPG). Thus, the fuel economy of the series vehicle defined here is significantly more sensitive to battery specific power than the parallel hybrid because of the series hybrid's larger battery pack. Similar analyses can be performed with the specific power of other components as shown in Table 5.

Table 5. Fuel economy sensitivity coefficients (% change in fuel economy on the combined federal cycle for a 1% increase in the given parameter) for component specific power

Parameter	Series HV	Parallel HV	ICEV
HPU specific power	0.03	0.04	0.08
motor/inverter specific power	0.03	0.02	--
battery pack specific power	0.04	0.02	--

These results indicate that hybrid fuel economy, and thus the fuel economy benefit of drivetrain hybridization, does not depend strongly upon any *one* of the drivetrain

* Tamor's vehicles had a mass of 1020 kg, while the ICEV mass here, to which we "corrected" the series hybrid mass using sensitivity coefficients, was 1214 kg.

components' specific power. It does indicate that the fuel economy of the series hybrid, with its more powerful motor and battery pack, is significantly more sensitive to changes in specific power of these components than is the fuel economy of the parallel hybrid.

TRANSMISSION EFFECTS

Number of gears - A 5-speed transmission was used here. It has been shown that increasing the number of gears in a transmission can improve fuel economy [4]. This is principally because the engine associated with a transmission with a greater number of gears may be able operate more of its time in good thermal efficiency regions, reducing the opportunity for hybridization, and reducing the fuel economy benefit of hybridization. The continuously-variable transmission (CVT) represents the limit of an infinite number of gears. As commercially available CVTs become more efficient, so that their losses do not negate the benefit of improved engine efficiency they offer, they may be used to make ICEV fuel economy more competitive with hybrid vehicle fuel economy.

Drivetrain configuration - A reasonable modification of the parallel hybrid analyzed here would be to move the motor closer to the wheels. That is, we might expect greater fuel economy by not having the motor transmit torque through the transmission, but rather directly to the vehicle's differential (through a single speed-reducing gear). This would reduce losses in the transmission by reducing the energy passed through it, while increasing losses in the motor by forcing its speed to be a fixed fraction of the tire speed. See Figure 4 for a diagram of this alternate configuration.

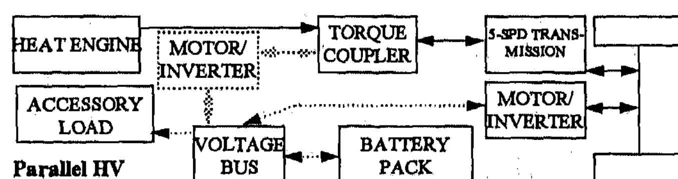


Figure 4. Parallel hybrid vehicle with motor connected directly to differential

We can estimate the fuel economy value of the motor-differential configuration over the one previously analyzed using the sensitivity coefficients presented above. Three energy-transfer modes will be affected: 1) motoring--mechanical energy provided by the motor will be subjected to fewer losses by circumventing the transmission, 2) regenerative braking--regenerative braking torque will likewise suffer fewer losses, and 3) HPU-to-motor charging--using the HPU to drive the motor as a generator will be significantly complicated by the interposition of the 5-speed transmission, but that effect notwithstanding, the transmission will incur losses on the charging torque it transmits. We assume that the parallel motor's motoring and regenerative efficiencies are unchanged from the baseline, although its torque and speed output are tied to the wheel requirements by the motor-differential setup and not by the baseline motor-transmission-differential setup. We also assume that HPU-to-motor charging through the 5-speed transmission is feasible, despite the control challenges involved.

In the motoring regime, the motor-to-wheels efficiency is improved by the ratio (1-spd transmission efficiency)/(5-spd transmission efficiency)=98%/92%=1.065, for a 6.5% improvement. Likewise, the regenerative braking, wheels-to-motor efficiency is improved by 6.5%. The HPU-to-motor efficiency is decreased by a factor of 92%/98%=0.939, for a 6.1% decrease. The combined effect of the motoring, regenerative braking, and charging efficiency changes can be computed using the sensitivities listed in Table 1 as "average motor eff.", "regenerative braking fraction", and "average motor-as-a-generator eff.", along with the changes computed above: +6.5%, +6.5%, and -6.1%. We have

$$(6.5\% \times 0.05) + (6.5\% \times 0.07) + (-6.1\% \times 0.14) = -0.9\%, \quad (2)$$

indicating a 0.9% drop from the baseline parallel vehicle fuel economy due to the change illustrated in Figure 4. This result is dependent the assumptions made about the transmission and motor efficiency and the drivetrain layout analyzed here, but indicates that the parallel hybrid fuel economy is not strongly dependent upon which of the two layouts analyzed is chosen.

EFFECT OF HPU TYPE ON FUEL ECONOMY DIFFERENCES - The size of the opportunity to avoid part-load operation depends greatly on the type of HPU used in the vehicles compared. The diesels used in this effort have relatively large operating torque-speed regions of high efficiency compared to some spark-ignition engines. The fuel economy benefit due to the hybridization of certain spark-ignition engine vehicles may therefore be greater than predicted here for the diesel engines. That is, the fuel economy *difference* between a spark-ignition hybrid and ICEV should be greater than the difference between a diesel-powered hybrid and ICEV, although both diesels should have better fuel economy than both spark-ignition engine vehicles. Although we have not quantified the difference between the fuel economy benefit of drivetrain hybridization of diesel- and spark-ignition-engine-powered ICEVs, we suggest that said quantification is perhaps not necessary. Given that one of the main aims of hybridization is high fuel economy, it seems reasonable to use a high-fuel-economy diesel as the ICEV benchmark.

Another opportunity hybrid vehicles offer is the use of alternative HPUs, such as gas turbines, fuel cells, and Stirlings in series (or perhaps parallel, in the case of the Stirling) HVs. Strictly speaking, we cannot isolate the benefit of drivetrain hybridization for vehicles using these powerplants because a conventional-drivetrain vehicle using them would undesirable, generally because of their poor dynamic response. However, we may wish to attribute the fuel economy benefits due to the use of alternative HPUs to the hybridization that makes them possible. These alternative HPUs offer the possibility of offering significantly more flexibility in fuel choice and improving vehicular emissions.

EFFECT OF HYBRID CONTROL STRATEGY - In order to compare the hybrid vehicles fairly with the ICEV simulated in this effort, significant effort was expended in developing and selecting advantageous hybrid control strategies.

Parallel HV - The parallel hybrid control strategy was defined by two parameters: a vehicle speed below which the HPU is turned off, and minimum HPU operating torque defined as a function of engine speed, which is the lowest torque output at which the HPU would operate whenever the tractive torque requirement is positive. Where used in the control strategy, these parameters were multiplied by an SOC-dependent factor ((SOChi-SOC)/(SOChi-SOClo)) designed to keep the SOC within a desired range. Changes in these parameters did not strongly affect the fuel economy of the vehicle, but rather only the steady average SOC that was attained. As stated before, significant work in control strategy development and evaluation is necessary to make hybrids reach their full fuel economy potential. This initial effort in parallel hybrid control strategy evaluation could be more fully developed, and future work is planned. In this work, the parallel hybrid fuel economy was nearly independent of the control strategy parameters used.

Series HV - A 'thermostat' hybrid control strategy (in which the HPU operates at a constant power level starting when the battery pack SOC reaches some bottom setpoint and stopping when the SOC has climbed to some upper setpoint) was evaluated along with a spectrum of power-follower control strategies. A power-follower control strategy, as discussed by Anderson and Pettit, is one in which the HPU power is dictated by the tractive motor power, but perhaps limited in its rate-of-change [20].

The spectrum of power-follower strategies that were investigated is defined by two independent variables: 'minimum HPU power' and 'averaging time'. Minimum HPU power is the power command below which the HPU is turned off. Averaging time is the size of the time window over which tractive motor power is averaged to derive the HPU power command. The HPU power command is the sum of an SOC-dependent factor and a scalar multiple of the averaged motor power. Figure 5 shows the dependence of this power-follower series hybrid fuel economy on the combined federal cycle upon these control strategy parameters.

Note that as the averaging time falls, fuel economy increases. Physically, as the averaging time falls, the HPU power level changes more quickly, following the power requirements of the tractive motor more closely. Thus, a smaller averaging time means more of the electric power generated by the HPU is used by the motor without passing through the batteries, and battery losses are decreased.

Two effects of 'minimum HPU power' are competing, particularly at low averaging times, leading to a best-fuel-economy point vertically in the center of the plot. As the minimum/cutoff power decreases, the HPU is allowed to operate at low powers which correspond to low efficiencies. As the minimum power increases, the HPU more frequently alternately charges the batteries (operating at or above its cutoff power) and shuts off, allowing the batteries to discharge. The shape of this map, then, is largely determined by battery turnaround efficiency and to a perhaps lesser extent by the HPU efficiency map.

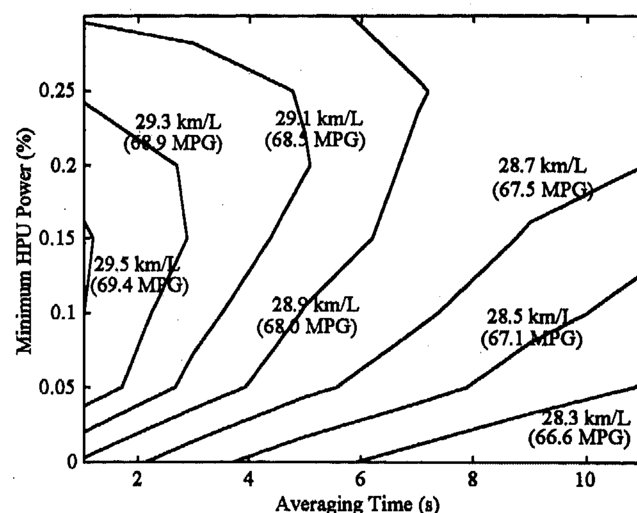


Figure 5. Fuel economy vs. control strategy parameters for series hybrid

The thermostat control strategy, using the peak HPU efficiency point and high and low SOC points of 80% and 40%, leads to a fuel economy of 31.1 km/L (73.2 MPG). Over the range in Figure 5, the power-follower strategy leads to a maximum fuel economy of 29.5 km/L (69.4 MPG) and a minimum of less than 28.3 km/L (66.6 MPG). Series hybrid fuel economy variations of at least 9%, then, can be traced to reasonable but suboptimal control strategies. Control strategy selection is an *extremely* important step in the design of hybrid vehicles for good fuel economy and emissions which we have not fully addressed. The sensitivity of fuel economy to control strategy shown here is likely a conservative estimate due to the simplifying assumptions (for example, neglect of the fuel economy effect of thermal and throttle transients).

CONCLUSIONS

With the understanding that hybrid vehicle design decisions are made for many reasons other than to maximize fuel economy, we have analyzed the dependence of hybrid fuel economy upon various vehicle and component design parameters. The main conclusions of this paper are as follows:

1. The parallel hybrid defined here achieves fuel economy that is 24% better than the ICEV while the series betters the ICEV by 18%.

2. The parallel hybrid is 4% more fuel-efficient than the series hybrid.

3. Lightweight midsized hybrid vehicles with low aerodynamic drag and rolling resistance using near- to mid-

term technology can achieve 30 km/L (70.6 MPG) (gasoline-equivalent).

4. The series hybrid control strategy strongly influences its fuel economy, and a 'thermostat' control strategy, neglecting thermal transient effects, is among the best of those examined here, *considering fuel economy only*.

5. When a series hybrid uses an HPU capable of following motor power demands on a second-by-second basis, a control strategy can be used that reduces the sensitivity of fuel economy to battery efficiency to the same level as in a parallel hybrid.

6. The parallel and series hybrid fuel economies are not strongly dependent upon component specific power, but series hybrids are more sensitive to battery and motor specific power than are parallel hybrids.

7. The series hybrid's fuel economy is less sensitive to regenerative braking effectiveness than to any other vehicle-level parameter, and the parallel hybrid's fuel economy is less sensitive only to motor efficiency than to regenerative braking effectiveness. Doubling regenerative braking capture effectiveness improves series hybrid fuel economy by 10% and parallel hybrid fuel economy by 7%.

8. This parallel hybrid's fuel economy is not significantly changed by connecting the motor directly to the differential rather than to the 5-speed transmission.

9. Control strategy development for a charge-sustaining parallel HV is challenging and warrants further investigation.

ACKNOWLEDGMENT

This work was sponsored by the Department of Energy (DOE) as part of the MRI-managed DOE Hybrid Vehicle Propulsion System Program. We would also like to gratefully acknowledge the inputs of Terry Penney, Robert Rehn, Robert Farrington, Michael Tamor, and the anonymous SAE reviewers.

REFERENCES

1. Hybrid Electric Vehicle Total Energy Cycle Analysis: Work Plan (Draft), Center for Transportation Research, Argonne National Laboratory, January 26, 1996

2. Moore, T. C., A. B. Lovins, "Vehicle Design Strategies to Meet and Exceed PNGV Goals," SAE Paper No. 951906

3. Moore, T. C., "Tools and Strategies for Hybrid-Electric Drivesystem Optimization," SAE Paper No. 961660

4. Duleep, K.G., "Automotive Technologies to Improve Fuel Economy to 2015, Draft Report," Prepared by Energy and Environmental Analysis for the Office of Technology Assessment, December 1994.

5. Duleep, K. G., "Fuel Economy Potential of Light Duty Vehicles in 2015+, Draft Final Report," Prepared by Energy and Environmental Analysis for Martin Marietta Energy Systems, April 1995

6. Tamor, M. A., "Hybrid Vehicle Modeling," presented at ENV '96, February 1996

7. Mason, W. T., U. Kristiansson, "Hybrid EVs versus Pure EVs: Which Gives Greater Benefits?" SAE Paper No. 94C017

8. Wipke, K., M. Cuddy, "Using an Advanced Vehicle Simulator (ADVISOR) to Guide Hybrid Vehicle Propulsion System Development," Presented at the NESEA 1996 Sustainable Transportation Conference, New York, NY, September 16, 1996

9. Sovran, G., M. S. Bohn, "Formulae for the Tractive-Energy Requirements of Vehicles Driving the EPA Schedules," SAE Paper No. 81084

10. Stock, D., R. Bauder, "The New Audi 5-Cylinder Turbo Diesel Engine: The First Passenger Car Diesel Engine with Second Generation Direct Injection," SAE Paper No. 900648, 1990

11. Deep Cycle Battery Specification and Technical Information, Optima Batteries, Inc., September 6, 1996

12. Swiatek, R., of Chrysler Corp., Personal Communication, October 3, 1995

13. Draft National Plan for Advanced Automotive Technologies, Office of Advanced Automotive Technologies, August 16, 1996

14. Product Information on the CaliberEV 32 Permanent Magnet Traction Drive System, Unique Mobility, June 3, 1996

15. Lesster, L.E., et al., "An Induction Motor Power Train for EVs--The Right Power at the Right Price," Advanced Components for Electric and Hybrid Electric Vehicles: Workshop Proceedings, October 27-28, 1993, Gaithersburg, Maryland

16. "Efficiency guidelines for future manual transmissions," Automotive Engineering, January 1996, pp. 45-49

17. Neumann, K. H., M. Kuhlmeyer, J. Pohle, "The new 1,9 L TDI diesel engine with low fuel consumption and low emission from Volkswagen and Audi," Société des Ingenieurs de l'Automobile Paper No. 92038

18. Reynolds, S., of Detroit Diesel Corp., Personal Communication, December 5, 1996

19. Draft SAE J1711 Recommended Practice, "Measuring the Electric Energy Consumption, All Electric Range, Fuel Economy, and Exhaust Emissions of Hybrid Electric Vehicles," February 24, 1995

20. Anderson, C., E. Pettit, "The Effects of APU Characteristics on the Design of Hybrid Control Strategies for Hybrid Electric Vehicles," SAE Paper No. 950493

APPENDIX

Fuel economy sensitivity coefficients for the vehicles were computed using the following equation:

$$\xi_i = \frac{FE_{110\%} - FE_{90\%}}{FE_{baseline(100\%)}} \times \frac{1}{1.10 - 0.90}, \quad \text{(A1)}$$

where i indicates the parameter being considered, $FE_{baseline(100\%)}$ is the fuel economy for the parameter at its baseline value, and $FE_{110\%}$ and $FE_{90\%}$ refer to fuel economies corresponding to parameter values of 110% and 90% of the baseline value.

Sensitivities to road load parameters other than mass (C_DA, coefficient of rolling resistance, and regenerative braking fraction) were computed using constant-efficiency vehicle representations. This approach was used to minimize the effect of the road load parameters on driveline component efficiencies. For example, if the coefficient of rolling resistance increases, the force required to propel the vehicle at any time over the cycle (regardless of speed) increases, leading to higher torque requirements of the transmission and motor and/or heat engine. Generally, increasing the torque output of a component increases its efficiency. Thus, the effect of increasing the coefficient of rolling resistance in a vehicle whose drivetrain component efficiencies depend upon torque (and speed) would be to increase those average component efficiencies. This must be avoided to isolate the effect of the coefficient-of-rolling-resistance increase.

To compute the fuel economy sensitivity to mass, drivetrain components for the 10% more massive and 10% less massive vehicles were sized to provide 0 to 96.5 km/h times equivalent to that of the baseline vehicle.

Fuel economy sensitivities to drivetrain component (HPU, motor, and battery) efficiency were computed using vehicle representations that included power loss maps. This approach, in contrast to the constant efficiency approach, was used to preserve the baseline vehicle's time history of power flows among the components. These flows depend upon the battery SOC at a given time as well as motor power demands, both of which are dependent upon the drivetrain component efficiency history.

Note that the sensitivity coefficient for the HPU for all three vehicles is unity, implying that for a 1% increase in engine efficiency (for example, from 40% to 40.4%), there will be a 1% increase in fuel economy. Another way of stating this is that fuel economy is directly proportional to HPU efficiency. This is as one would expect: all energy used on-board the vehicle (assuming it is charge-sustaining, and the fuel economy test is corrected to zero change-in-SOC) comes from fuel conversion occurring in the HPU.

The sensitivity coefficient for the motor for the series hybrid (0.91) is over 18 times that of the parallel hybrid (0.05). This difference is due to the much smaller amount of motor energy used by the parallel hybrid than by the series. In the parallel vehicle, most tractive energy is supplied directly by the HPU, while in the series, all tractive energy must pass through the motor. Note that the motor and HPU sensitivity coefficients differ for the series hybrid because while all energy used by the vehicle passes through the HPU, the same is not true of the motor: accessory draw is independent of the motor.

Table A1 (reprise of Table 3). Fuel economy sensitivity coefficients (% change in fuel economy on the combined federal cycle for a 1% increase in the given parameter)

Parameter	Series HV	Parallel HV	ICEV
average HPU eff.	1.00	1.00	1.00
average motor eff.	0.91	0.05	--
average motor-as-a-generator eff.	--	0.14	--
average battery turnaround eff.	0.15	0.13	--
mass	-0.60	-0.63	-0.75
C_DA	-0.25	-0.25	-0.22
coeff. of rolling resistance	-0.26	-0.25	-0.22
accessory load	-0.17	-0.15	-0.16
regenerative braking fraction	0.10	0.07	--

The vehicle fuel economy sensitivities to the remaining parameters in Table A1 are nearly independent of vehicle type. For the battery efficiency, the closeness of the series and parallel hybrid sensitivities is somewhat coincidental, as this sensitivity is highly dependent upon control strategy. However, the control strategies used in the two vehicles are (intentionally) similar in that they minimize battery energy throughput. Changes in battery efficiency become more important as more energy is passed through the battery. The sensitivity of a thermostat-control-strategy series hybrid's fuel economy to battery efficiency would be much higher than shown here, since the effect of the thermostat control strategy is to alternately fill and empty the batteries.

The small differences among the sensitivities to road load parameters, mass, $C_{rolling\text{-}resistance}$, and C_DA (the aerodynamic drag product) are to be expected. The relative values of these sensitivities depend only upon the relative values of the road load parameters themselves. Road load parameters are identical for all vehicles, with the exception of mass, which differs by less than 2.4% among the three. Accessory load sensitivities are likewise dependent upon road load parameters, and are nearly identical for the three vehicles. Vehicle-to-vehicle differences in any one of these parameters can be ascribed to uncertainty in the calculation.

Sovran and Bohn use the standard vehicle tractive force equation, $F=(1/2)\rho C_DAv^2+mg(C_{rolling\text{-}resistance}+C_{rr1}v)+ma$, along with driving cycle data to analytically determine sensitivity coefficients or "influence coefficients," as they are called in the paper [9]. We can use the equations presented there to develop sensitivity coefficients against which to compare the coefficients we developed here, for each vehicle.

The sensitivity coefficients developed using Sovran's methods are presented in Table A2.

Comparing Tables A1 and A2, we can see that all the predicted sensitivities to aerodynamic drag product (C_DA) and coefficient of rolling resistance agree to within 9.1%. The mass sensitivity coefficients for the ICEV agree to within 2.7%. Using Sovran and Bohn approach for the sensitivity of fuel economy to mass is inappropriate, however, as evidenced by the difference between the two sets of estimates (-0.60 vs. -0.76 for the series and -0.63 vs. -0.76 for the parallel). Sovran and Bohn's equations were developed for a ICEV, without regenerative braking. The hybrids in this study employ regenerative braking; therefore they recover some of the energy expended to accelerate the vehicle, reducing the effect mass has on the total tractive energy required, and thus, fuel required for a given cycle.

Table A2. Fuel economy sensitivity coefficients developed using the methods of Sovran and Bohn (% change in fuel economy over the combined federal cycle for a 1% increase in the given parameter). Note that Sovran and Bohn's "influence coefficients" were positive; their signs were changed for consistency with previous sensitivity coefficients presented.

Parameter	Series HV	Parallel HV	ICEV
mass	-0.76	-0.76	-0.77
C_DA	-0.24	-0.24	-0.23
coeff. of rolling resistance	-0.26	-0.24	-0.24

970290

Hybrid and Conventional Hydrogen Engine Vehicles that Meet EZEV Emissions

Salvador M. Aceves and J. Ray Smith
Lawrence Livermore National Lab

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Hydrogen-fueled, spark-ignited, homogeneous-charge engines offer an alternative for providing Equivalent Zero Emission Vehicle (EZEV) levels, along with a range and performance comparable to today's automobiles. Hydrogen in a spark-ignited engine can be burned at very low equivalence ratios, so that NO_x emissions can be reduced to less than 10 ppm without a catalytic converter or EGR. HC and CO emissions may result from oxidation of engine oil, but by proper design are negligible (a few ppm). Lean operation also results in increased indicated efficiency due to the thermodynamic properties of the gaseous mixture contained in the cylinder and due to reduced heat transfer. The high effective octane number of hydrogen allows the use of a high compression ratio, further increasing engine efficiency.

In this paper, a time-dependent engine model is used for predicting hydrogen engine efficiency and emissions. The model uses basic thermodynamic equations for the compression and expansion processes, along with an empirical correlation for heat transfer, to predict engine indicated efficiency. A friction correlation and a supercharger/turbocharger model are then used to calculate brake thermal efficiency. The model is validated with many experimental points obtained in a recent evaluation of a hydrogen research engine.

The validated engine model is then used to calculate fuel economy and emissions for three hydrogen-fueled vehicles: a conventional, a parallel hybrid, and a series hybrid. All vehicles use liquid hydrogen as a fuel. The hybrid vehicles use a flywheel for energy storage. Comparable ultra capacitor or battery energy storage performance would give similar results. This paper analyzes the engine and flywheel sizing requirements for obtaining a desired level of performance. The results indicate that hydrogen lean-burn spark-ignited engines can provide a high fuel economy and Equivalent Zero Emission Vehicle (EZEV) levels in the three vehicle configurations being analyzed.

INTRODUCTION

Hydrogen has very special properties, including a very high laminar flame speed, a high effective octane number, and no toxicity or ozone-forming potential [1]. Homogeneous-charge spark-ignited piston engines can be designed to take advantage of these characteristics. The high laminar flame speed allows the use of very low equivalence ratios (as low as 0.2), reducing NO_x emissions to near-zero levels without requiring a catalytic converter, that may deteriorate with time. The use of low equivalence ratios also increases the indicated efficiency, and reduces the need for throttled operation [2]. The engine can have a high compression ratio, due to the high octane number of hydrogen.

This paper analyzes the applicability of hydrogen homogeneous-charge spark-ignited piston engines to Equivalent Zero Emission Vehicles (EZEV).[1] The analysis uses an engine model that is calibrated to match the data obtained in a recent experiment [4], and a vehicle simulation code that can be applied to calculating fuel economy and emissions [5]. The engine model is used to generate optimized engine performance maps for supercharged operation, with constraints of limited NO_x emissions, limited manifold pressure and maximum mean piston speed.

The engine maps are applied for predicting fuel economy and emissions for a conventional vehicle, a parallel hybrid vehicle, and a series hybrid vehicle. All these vehicles use liquid hydrogen as a fuel, and hybrid vehicles use a flywheel for energy storage. Although liquid hydrogen is not a common fuel, its safe

[1] Equivalent Zero Emission Vehicles are defined here as those that generate less emissions when operating inside the Los Angeles Basin than the power plant emissions generated as a result of electric car operation. These emission levels are being considered for approval by CARB [3]. Proposed EZEV levels are: 0.0025 g/km (0.004 g/mi) NMHC; 0.11 g/km (0.17 g/mi) CO; and 0.012 g/km (0.02 g/mi) NO_x. These values are equal to one tenth of CARB ULEV standards.

use has been demonstrated by BMW for a number of years [6]. The parallel hybrid operates almost identically as a conventional car, with a flywheel for regenerative braking and for complementing the power of the engine during sudden accelerations. The series hybrid vehicle engine operates in an on-off mode to keep the flywheel charged. When the engine is off, the flywheel provides all the energy for transportation and accessories.

ENGINE MODEL

The engine model uses first principles and correlations to predict piston engine efficiency and power output. The engine model is a lumped (zero-dimensional), time-dependent model which solves the basic differential equations for the compression and power strokes. An empirical equation [7] is used for calculating heat release. The heat release equation has three free parameters which determine the shape and duration of the heat release function as a function of crank angle. These three parameters are determined for each experimental run by using an optimizer [8]. The optimizer finds the combination of the three parameters that minimizes the differences between the experimental pressure trace and the pressure trace calculated by the model. The results have been very satisfactory. The relative errors in matching the pressure traces at MBT timing have been of the order of 0.5%, with a maximum error of 1% over all engine speeds, equivalence ratios and manifold pressures.

The engine model uses Woschni's correlation [9] to estimate engine heat transfer. It was found during the analysis that the heat transfer correlation underpredicts heat transfer losses. Woschni's correlation includes two constants, C_1 and C_2, which determine the effect of fluid flow on heat transfer. To obtain a better match with the experimental data, the original values of these two constants were multiplied by 1.8. The increased heat transfer coefficient reduces the predicted engine efficiency by about 1% with respect to the engine model with the unmodified Woschni's correlation. It is noted that Woschni's correlation has been applied in the past to diesel and gasoline engines, so some adjustment to the coefficient is to be expected as a result of using a different fuel. It is encouraging that the adjustments to those coefficients, which are not directly proportional to heat transfer, have a small effect on engine efficiency.

The engine model includes a friction model and a supercharger/turbocharger model to predict brake thermal efficiency. The friction model uses a detailed correlation [10]. Supercharger and turbocharger performance are calculated by using a thermodynamic model and assuming a constant (0.7) isentropic efficiency for both the turbine and the compressor. Selection of a supercharger to optimally increase the power as required to meet the power demands is outside the current experience of the authors. However, a detailed supercharger map could be incorporated into the model if further refinement is desired. A water-cooled intercooler is assumed with a thermal effectiveness of 0.7. Based on reported data for typical engines [2], volumetric efficiency is assumed to vary linearly from 85% at low engine speeds, to a maximum of 95% at 4000 rpm, down to 90% at 5000 rpm.

The engine model is validated by comparing the calculated results with the experimental results obtained in a recent engine evaluation [4]. The engine used in the experimental evaluation is an Onan engine which was modified by incorporating a head containing two spark plugs, along with the original two valves. The intake valve was modified with a 1.5 mm high, 180 degree shrowd to add some swirl which resulted in improved efficiency. The combustion chamber is a simple right circular cylinder with no squish and a flat top piston. This geometry has been shown to be the most efficient shape for reducing heat transfer losses in lean-burn engines [11]. Engine characteristics are listed in Table 1, along with the range of conditions used in the experiment.

Table 1. Modified Onan engine characteristics and experimental conditions.

Bore, mm	82.55
Stroke, mm	92.08
Displacement, cm^3	493.0
Geometric compression ratio	14.0
Experimental range for equivalence ratio	0.2-0.5
Experimental range for engine speed, rpm	1200-1800
Experimental range for volumetric efficiency, %	90-215

Figure 1 shows a comparison between experimental and calculated indicated efficiencies, as a function of equivalence ratio, for all the experimental points at MBT timing. Engine speeds are indicated with different symbols. A 0.39 equivalence ratio was selected for most supercharged runs. Figure 1 shows gross indicated efficiency, defined as including only work done on the piston during the compression and expansion strokes [2]. The experimental engine was supercharged by directly supplying the fuel and air from pressurized sources. Pumping work is accounted for as a part of the engine friction model.

Figure 1 shows that the model predicts absolute values as well as trends with good accuracy for engine indicated efficiency, over the whole range of operating conditions, with the maximum error of the order of 1%.

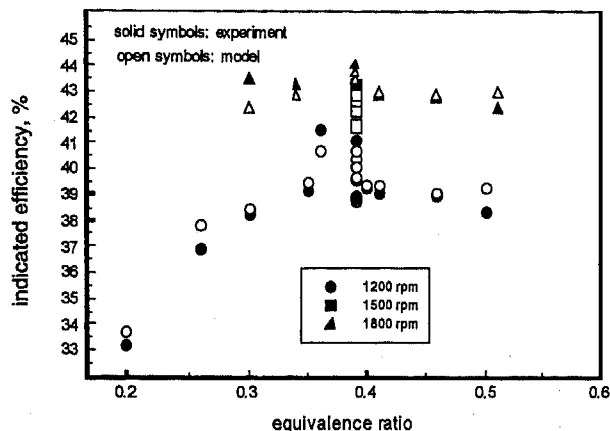


Figure 1. Indicated efficiency for the Onan engine as a function of equivalence ratio. The figure includes both the experimental results and the model predictions.

No validation is done for brake thermal efficiency, because the Onan engine used in the experiment has substantially more friction per cylinder than a current automotive engine, for which the friction correlation applies.

Figure 1 shows that indicated efficiency increases as a function of engine speed, as a consequence of reduced heat transfer losses. The variation of indicated efficiency with equivalence ratio is best observed for 1200 rpm operation, for which the greatest fuel/air range was used. Indicated efficiency reaches a maximum near a 0.40 equivalence ratio. Increasing the equivalence ratio from this point results in a decreased indicated efficiency, due to increased heat transfer and a decreased specific heat ratio ($\gamma = c_p/c_v$) for the gas products in the cylinder [2]. Decreasing the equivalence ratio from the optimum point increases the timing losses, due to slower heat release, thereby reducing the indicated efficiency. Supercharged operation results in small indicated efficiency gains due to slightly lower heat transfer losses per unit mass of fuel at the higher densities. Supercharged operation has a larger effect on brake thermal efficiency by increasing the output work relative to the frictional work.

The engine model also includes a correlation for calculation of NO_x emissions. Experimental results (Figure 2) have indicated that emissions of NO_x are mainly a function of engine equivalence ratio, being very insensitive to engine speed or intake pressure. The correlation of NO_x as a function of equivalence ratio predicts emissions accurately. The model uses a correction for supercharged and turbocharged operation, to take into account the higher intake temperature resulting from the compression process

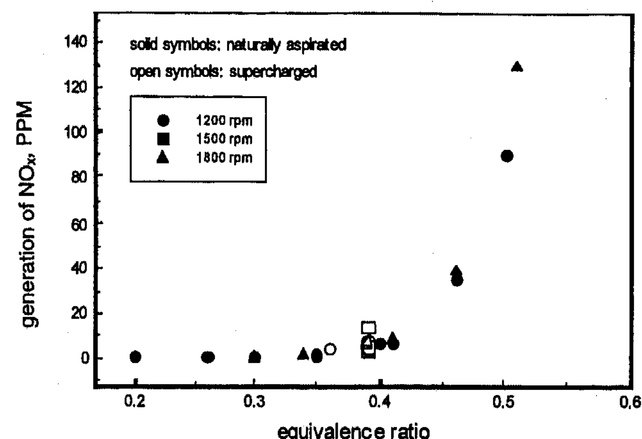


Figure 2. Emission of NO_x in parts per million for the Onan engine as a function of equivalence ratio, for all the experimental points at MBT timing obtained in the analysis. Engine speeds and supercharged operation are indicated with different symbols.

and the less than perfect effectiveness of the intercooler.

The engine model is then applied to predicting the engine performance that results if a 4-cylinder (1.97 liter) and a 3-cylinder (1.48 liter) hydrogen engines are built with the same cylinder characteristics of the Onan engine. The applicability of the 4-cylinder and 3-cylinder engines to conventional and hybrid vehicles is discussed later in this paper. The geometry and dimensions of the engine cylinders are not changed in the analysis, because small changes in geometry may result in significant changes in efficiency. It is expected, however, that larger engine cylinders will improve engine efficiency.

Using the engine model for predicting vehicle performance requires extrapolating from the engine speeds used in the experiment (1200-1800 rpm) to engine speeds that are required for vehicle operation. A maximum engine speed of 5000 rpm is assumed. Extrapolation is done by assuming that the heat release correlations developed from the experimental data apply throughout the desired range of engine speeds, along with the previously cited volumetric efficiency assumptions. It is recognized that this extrapolation may result in errors. Errors are due to turbulence variations with engine speed which influence heat release rate, thus changing both timing losses and heat transfer losses. However, Figure 1 shows that the model predicts the efficiency trends with good accuracy for the range in which experimental data exist, and it is considered that the model can do a reasonable job at predicting efficiency

for high and low engine speeds. In addition to this, engines in conventional and hybrid vehicles are most often operated at low to moderate speeds, which is the range for which the model has been validated. For the conventional and parallel hybrid vehicles analyzed in a later section of this paper the mean engine speed is 1500 rpm for the urban cycle and 2700 rpm for the highway cycle (see Figure 8). Maximum engine speed for the driving cycles is 3300 rpm. The series hybrid vehicle is set to operate at a constant 2400 rpm during the driving cycles.

Engine brake thermal efficiency is required for applying the engine code to vehicle calculations, and brake thermal efficiency is calculated with a friction and a supercharger/turbocharger model that have not been validated for this particular application. While the model cannot replace experimental runs, it is considered that the brake thermal efficiency calculated with the model gives a good idea of the performance that can be obtained with such an engine. Engine emissions are very insensitive to engine operating conditions other than maximum temperature within the cylinder, and it is therefore expected that the engine NO_x model can provide accurate predictions for emissions levels throughout the operating range.

The engine model is applied to generate engine emissions and performance maps, necessary for predicting vehicle performance. A gasoline engine has only one degree of freedom for controlling the output torque at any given speed: the inlet manifold pressure. This is due to the use of three-way catalysts that require near-stoichiometric operation for high conversion efficiency. A hydrogen engine has two degrees of freedom, because equivalence ratio can also be varied. Generating an engine map therefore requires determining a control strategy that specifies how to adjust these two parameters to obtain the desired torque for any given engine speed. In this analysis, an optimizer [8] is used to determine the combination of equivalence ratio and inlet manifold pressure that satisfies the torque requirement while providing the maximum engine brake thermal efficiency. A constraint is used in the optimization: engine NO_x emissions are less than 10 ppm under all operating conditions. An engine generating 10 ppm of NO_x is well below the EZEV standards (a tenth of ULEV), provided that it is installed in a vehicle with a fuel economy of 40 mpg or higher. While emission levels for a car are set in grams per mile and not in ppm, setting a maximum allowable ppm limit guarantees that the vehicle is intrinsically an EZEV, regardless of how the vehicle is driven.

Lean operation results in low power output, and therefore turbocharged or supercharged operation is required for providing adequate power output for the chosen displacement. Both supercharged and turbocharged operation have been considered for generating the engine performance map. The performance maps for both cases are very similar, with turbocharged operation having a slight efficiency advantage over supercharged operation. Only the results for supercharged operation are shown in this paper. Supercharged operation is preferred to turbocharged operation due to the lag time that may exist in turbocharged operation for the conventional and parallel hybrid vehicles.

Figures 3 and 4 show the predictions for engine efficiency and emissions maps, for the 4-cylinder supercharged engine. Performance maps for the 3-cylinder engines can be obtained to within a good approximation by multiplying the torque scale by 3/4. Not everything scales by a factor of 3/4 when the number of cylinders is decreased (for example, the number of bearings). However, the effect of these changes is small, and the scaled maps give a very good approximation for the 3-cylinder engine maps.

Figure 3 shows lines of constant brake thermal efficiency (in percent) as a function of engine speed and engine torque. The figure also shows a dotted line corresponding to the conditions at which the engine generates 10 ppm of NO_x, and a dashed line that indicates the maximum torque that can be obtained within the upper bounds of equivalence ratio (0.5) and inlet pressure (2 bar) used in the analysis. The 10 ppm NO_x curve is the lower of the two, and therefore sets the limit on the maximum torque and power that can be obtained from the engine. The maximum power approaches 60 kW at 5000 rpm. The contour lines in this and the following figures spread beyond the 10 ppm line, showing the potential power gains obtained by relaxing this restriction. A square in the figure indicates the approximate range of experimental conditions covered.

Figure 3 shows that the engine is predicted to have a broad area of high efficiency, for intermediate speeds and high torques. The efficiency drops for lower speeds due to increased heat transfer losses, and for higher speeds due to increased friction. As expected, the efficiency drops to zero as the load is reduced. However, the drop occurs more slowly than in conventional engines, because the equivalence ratio can be reduced as the load is reduced, resulting in lower throttling losses.

The predicted engine map is similar in shape to engine maps for recent gasoline engines (see, for example, [12]). However, engine maps are most often drawn with efficiency as a function of speed and power, so the maps presented in this paper have to be transformed before a direct comparison is possible. Current gasoline engine maps show that efficiency drops with increasing speed at low torque conditions,

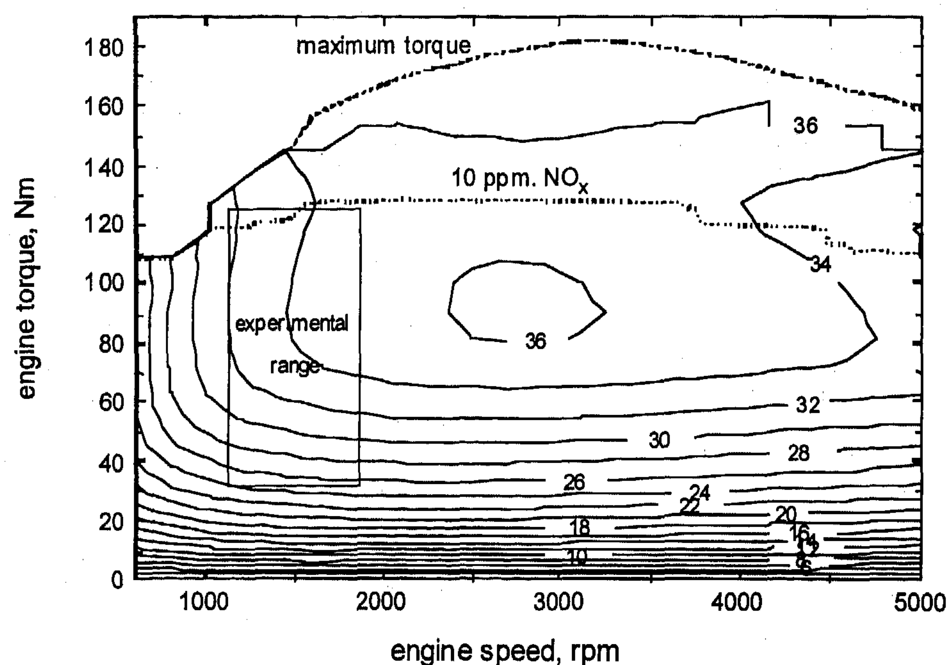


Figure 3. Contour lines of constant brake thermal efficiency (in percent) as a function of engine speed and engine torque. The dotted line corresponds to the conditions at which the engine generates 10 ppm of NO_x; the dashed line indicates the maximum engine torque that can be obtained within the constraints of maximum equivalence ratio (0.5) and inlet pressure (2 bar); and the square indicates the approximate area in which the experimental data were taken.

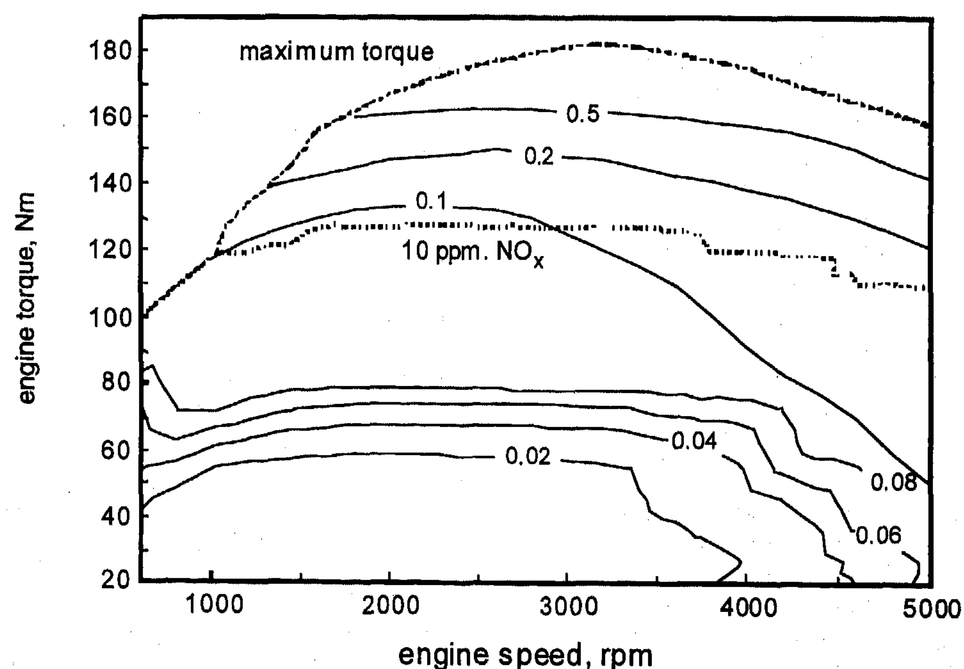


Figure 4. Contour lines of NOx emissions in g/kWh as a function of engine speed and torque. The figure includes a dotted line for the 10 ppm NO_x limit, and a dashed line for the maximum engine torque that can be obtained within the constraints of maximum equivalence ratio (0.5) and inlet pressure (2 bar).

while Figure 3 shows that efficiency is almost independent of speed at these conditions (lower right corner of the map). These differences may be due to lower heat transfer and pumping losses in the lean-burn engine compared to the gasoline engine.

Figure 4 shows contours of NO_x emissions in g/kWh as a function of engine speed and torque. Emissions are less than 0.1 g/kWh for most of the map, with higher emissions obtained at high speed and high torque. The line of 0.1 g/kWh roughly corresponds to the limit of 10 ppm at intermediate speeds. Emission levels are near zero (<0.02 g/kWh) over the low load range, which is the range at which the engine is operated most of the time during city and highway driving in conventional and parallel hybrid cars (see Figure 8). Emissions increase slowly as the torque increases, until the restriction of 10 ppm is approached. When this restriction is approached, the operating conditions in the engine are adjusted so that the 10 ppm line is pushed as high as possible by increasing the manifold pressure without further increases in equivalence ratio, at the cost of some losses in efficiency. This explains the great distance between the 0.08 and the 0.1 g/kWh curves shown in the figure. Emission levels shown in the figure are expected to be valid over the lifetime of the engine, since no catalytic converter, which may deteriorate with time, is used to control emissions.

Intake pressure and equivalence ratio are selected by the optimizer to obtain the optimum brake thermal efficiency at every operating point. At low loads, the equivalence ratio is as low as 0.25 and the intake pressure is 0.5 bar. The equivalence ratio is reduced to keep a relatively high manifold pressure, therefore reducing pumping losses. Full information about the control strategy, including maps of equivalence ratio and intake pressure as a function of engine torque and speed, can be obtained from a previous publication [13].

VEHICLE ANALYSIS

The engine efficiency and NO_x maps presented in the previous section are now used in predicting vehicle fuel economy and emissions for a conventional and parallel and series hybrid vehicles. Vehicle power train components are also selected to meet minimum performance requirements (acceleration, hill climb and range). This is accomplished by incorporating the engine maps into an existing vehicle simulation code [5].

The main characteristics of the three vehicles are listed in Table 2. All the vehicles have good aerodynamics and low rolling friction, and use a cryogenic liquid hydrogen storage. The conventional and the parallel hybrid cars use a 5-speed transmission, and the series hybrid uses a single-speed transmission. Transmission efficiency is considered independent of driving conditions. All the vehicles are required to meet the following minimum performance specifications: acceleration from 0 to 97 km/h (0-60 mph) in no more than 10 seconds; climbing a slope of at least 6% at a constant 97 km/h (60 mph) speed with a 272 kg payload; and minimum range of 483 km (300 miles).

Both hybrid vehicles use a flywheel for energy storage. A flywheel is selected because it has the potential of providing a high efficiency and power output. A detailed model of a flywheel has been developed. This model includes a performance map of efficiency as a function of state of charge and power output, as well as a correlation of stand-by losses as a function of state of charge.

A brief description of the operating strategies for the parallel and series vehicles is as follows: The parallel hybrid vehicle operates very similarly to a conventional car, except that a flywheel and a traction motor are used for complementing the power of the engine during sudden accelerations, and for regenerative braking. The use of the flywheel for power peaking allows a reduction in the size of the engine, which may therefore operate more efficiently than the engine used in the conventional car at the low-power conditions that constitute most of the urban and highway driving cycles. The engine is sized to provide the required performance during long hill climbs, for which the flywheel cannot provide the required energy. A low-capacity, low-power flywheel is enough for this application. The control strategy used for the flywheel consists of keeping it near 50% state of charge. In this way, the flywheel is always ready to provide energy for a sudden acceleration, and to absorb energy during regenerative braking. It is noted that there are many possible parallel hybrid control strategies. The 50% state of charge for the flywheel is chosen as a simple strategy that accomplishes both engine downsizing and regenerative braking. It is likely that better parallel strategies may exist.

The series hybrid vehicle operates with an engine in an on-off mode, with no mechanical link between the engine and the wheels. An electric motor provides all tractive power. When the engine is running, it drives a generator that supplies electricity to both the electric motor and an energy storage system. A flywheel is used for energy storage. When the storage system is fully charged, the engine is turned off, and the storage system provides all the energy required for traction and accessories. Series hybrid vehicles have high fuel economy because the engine is operated at a high efficiency condition without ever idling. Spark ignition engines can be run unthrottled thus avoiding pumping

Table 2. Main parameters for hydrogen-fueled conventional, parallel hybrid and series hybrid vehicles.

Vehicle parameter	conventional	parallel	series hybrid
frontal area, m^2	2.04	2.04	2.04
aerodynamic drag coefficient	0.24	0.24	0.24
coefficient of rolling friction	0.007	0.007	0.007
transmission efficiency	0.94	0.94	0.95
transmission gears	5	5	1
accessory load, W	1000	1000	1000
engine idling speed, rpm	600	600	-
launch engine RPM, maximum effort acceleration	3600	3600	-
regenerative braking	no	yes	yes
fraction of available energy recovered by regen. braking, %	-	70	70
maximum one-way flywheel efficiency, %	-	96	96
generator type	-	permanent magnet	permanent magnet
motor type	-	AC induction	AC induction
energy storage device	-	flywheel	flywheel
hydrogen storage	liquid	liquid	liquid

Table 3. Results of the analysis for the conventional, parallel hybrid and series hybrid hydrogen-fueled vehicles.

Vehicle parameter	conventional	parallel hybrid	series hybrid
test weight, (empty weight + 136 kg)	1136	1127	1238
number of engine cylinders	4	3	3
engine displacement, liters	1.97	1.479	1.479
hydrogen tank storage capacity, kg	7.2	5.7	4.8
motor maximum short-term torque, Nm	-	48	190
motor maximum speed, rpm	-	11000	11000
flywheel energy storage, kWh	-	0.2	1
flywheel maximum power, kW	-	20	100
average engine efficiency, urban cycle, %	21.4	21.9	36
average engine efficiency, highway cycle, %	22.7	25.7	36
fuel economy[1], urban cycle, km/liter (mpg)	15.2 (35.8)	20.0 (47.0)	24.4(57.4)
fuel economy[1], highway cycle, km/liter (mpg)	22.0 (51.8)	26.5 (62.4)	29.8(70.0)
fuel economy[1], combined cycle, km/liter (mpg)	17.7 (41.6)	22.5 (52.9)	26.6(62.5)
NO_x emissions, urban, 10^{-3}g/km (10^{-3}g/mile)	2.08 (3.35)	2.24 (3.61)	12.4(20.0)
time for 0-97 km/h (0-60 mph), s	10.0	10.0	10.0
max. climbing slope at 97 km/h (60 mph), %	12.8	9.0	6.0
vehicle range, combined cycle, km (miles)	483 (300)	483 (300)	483 (300)

losses. Engines in series hybrids operate most of the time at a low power, high efficiency condition, at constant speed and load. When additional power is required during long hill climbs, the engine can be operated at a high power level, possibly at a lower efficiency. The flywheel provides the power for sudden accelerations, and the engine is sized for providing the required performance for long hill climbs. On-off engine operation results in decreased engine efficiency, due to the repeated engine cold starts. This reduction in efficiency is taken into account in this analysis by introducing an energy penalty (10 Wh) every time the engine is restarted. This is done to account for increased friction due to engine cold operation. Emissions are not expected to increase considerably due to on-off engine operation because no catalytic converter is used to control emissions.

The results of the analysis are listed in Table 3. Empty weight is set to 1000 kg for the conventional car, which is used as base case. This low weight is chosen because research programs such as PNGV (Partnership for a New Generation of Vehicles) are likely to result in weight reductions for future cars.

Weight for the other two vehicles is calculated by replacing components, and adding or subtracting the weight of the components. A 30% structural penalty is added to the difference in power train weight, to take into account the need for a heavier structure that results from a heavier power train. The parallel hybrid car has almost the same weight as the conventional car. The weight of the added components (flywheel and motor) is compensated by the reductions in weight that result by downsizing the engine, transmission and hydrogen tank. The series hybrid vehicle weighs about 100 kg more than the other two vehicles, primarily due to its need for a flywheel with greater energy storage capacity.

The conventional vehicle requires a 4-cylinder engine to provide the required acceleration performance. This engine is therefore oversized for providing the required hill climbing performance (12.8% instead of 6%). The hybrid vehicles only require a 3-cylinder engine, since the flywheel can be used for providing the power for maximum effort acceleration. Engines for the hybrid vehicles are sized for the hill climb requirement. However, the previously imposed restriction of keeping the cylinder size and geometry constant in the analysis results in an oversized engine for the parallel hybrid. It would be necessary to reduce the size of the individual cylinders to exactly obtain the specified hill climb performance. The mass of hydrogen in the tank (and therefore the mass and volume of the hydrogen tank) is adjusted to meet the desired range. The internal tank volume for the conventional car is about 100 liters (27 gal) [14]. The parallel hybrid only requires a 20 kW, 0.2 kWh flywheel for power augmentation, compared to 100 kW and 1 kWh for the series hybrid.

Fuel economy is listed in the table in gasoline-equivalent units. Fuel economy for the series hybrid is highest because the engine always operates at peak efficiency (36%). The difference in fuel economy between the parallel hybrid and the conventional car is due in part to the higher average engine efficiency that results from downsizing the engine, and in part due to the regenerative braking that can be done in the hybrid.

While all vehicles have a high fuel economy, the fuel economy of the series hybrid vehicle is lower than the 33.5 km/l (79 mpg) predicted for a hydrogen series hybrid vehicle in a previous work by the authors [5]. The results are different because the engine model used in this analysis is based on experimental data for a particular engine (Onan), while the previous work indicates improvements that are likely to be obtained in a future optimized hydrogen engine.

The most desirable feature of the vehicles analyzed in this paper is their low emissions. Emissions are measured only during the urban cycle for EPA certification. Therefore only these values are reported in Table 3. The conventional car has the lowest emission levels. Emissions for the conventional car are projected to be a factor of 60 lower than the CARB ULEV requirements, and therefore a factor of 6 lower than EZEV. The reason for the emissions to be so low is that the engine is operated most of the time at low torque, generating much less than the 10 ppm maximum allowable NO_x (Figure 8). Emissions out of the parallel hybrid vehicle are slightly higher than the conventional car, because the downsized engine has to operate at a higher load. The series hybrid vehicle operates at peak efficiency, which occurs near the limit of 10 ppm of NO_x. Emissions out of the series hybrid engine are therefore higher than for the other two vehicles, exactly at the EZEV limits. Emissions out of the series hybrid could be reduced by changing the engine operating conditions, at the cost of a slight loss in efficiency, to allow for a margin of error during the certification process.

Two important facts should be emphasized about the emissions generated by these vehicles. First, no catalytic converter is required, and therefore emissions do not increase during the life of the car due to catalytic converter or emissions control system failure. Second, emissions are intrinsically low. No matter how aggressive a driver may drive, emissions are always under 10 ppm NO_x, eliminating high off-cycle emissions that are common in gasoline cars.

Figures 5, 6, 7 and 8 are used to illustrate some of the results. Figure 5 shows flywheel state of charge for the series hybrid and the parallel hybrid vehicles along the

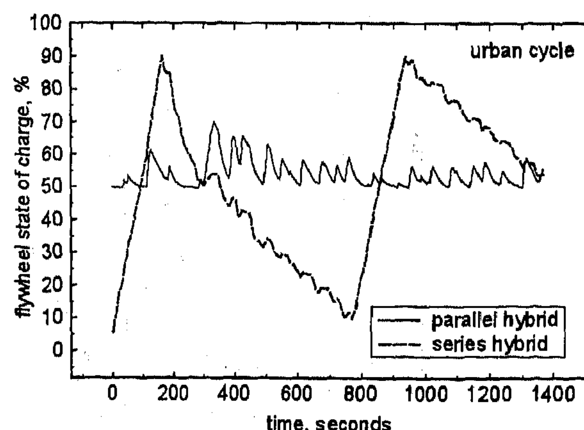


Figure 5. Flywheel state of charge for the series hybrid and the parallel hybrid vehicles along the urban driving cycle. For the series hybrid, the engine runs at constant power until it fully charges the flywheel. For the parallel hybrid, the flywheel is kept at about 50% state of charge, ready to store energy during regenerative braking, or provide energy for a sudden acceleration.

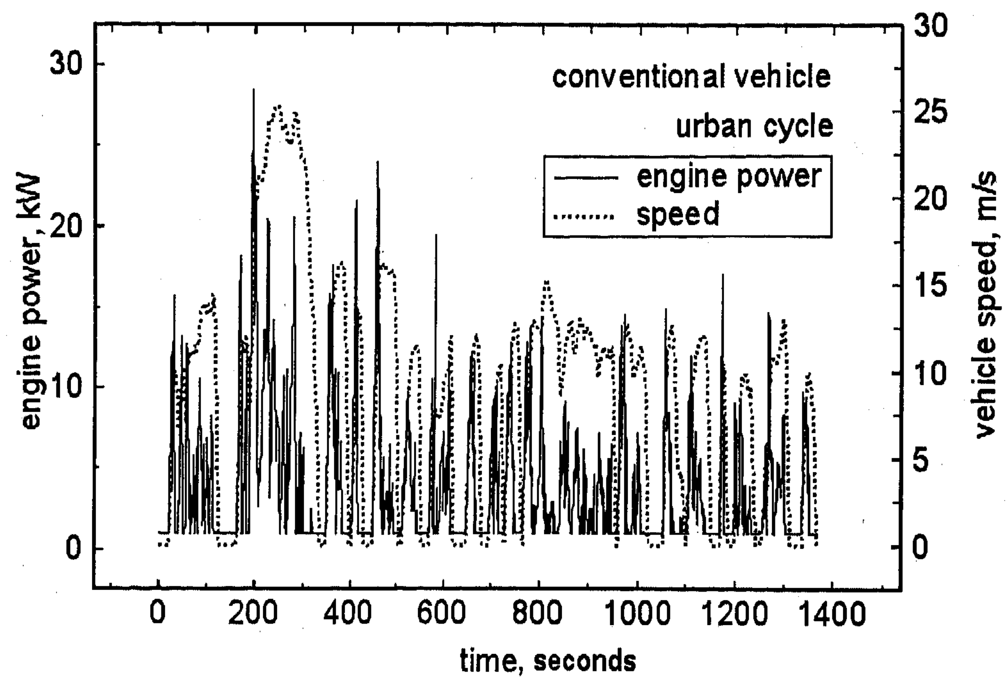


Figure 6. Engine power for the conventional vehicle during the urban cycle. The driving cycle speed is indicated by a dotted line.

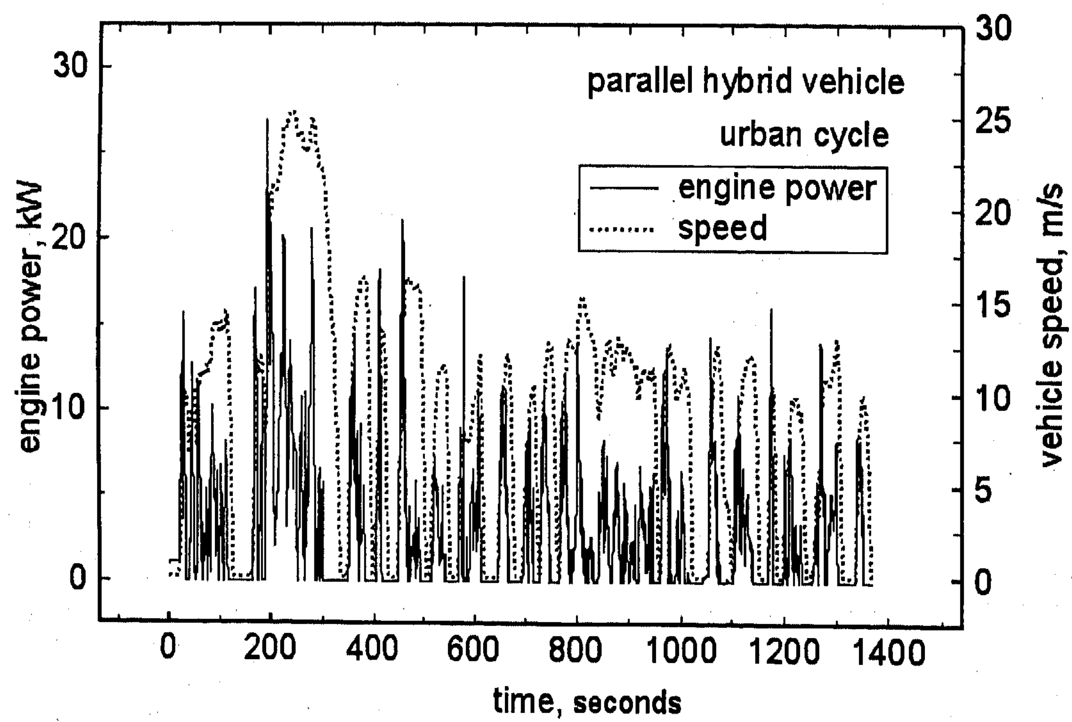


Figure 7. Engine power output for the parallel hybrid vehicle. The driving cycle speed is indicated by a dotted line.

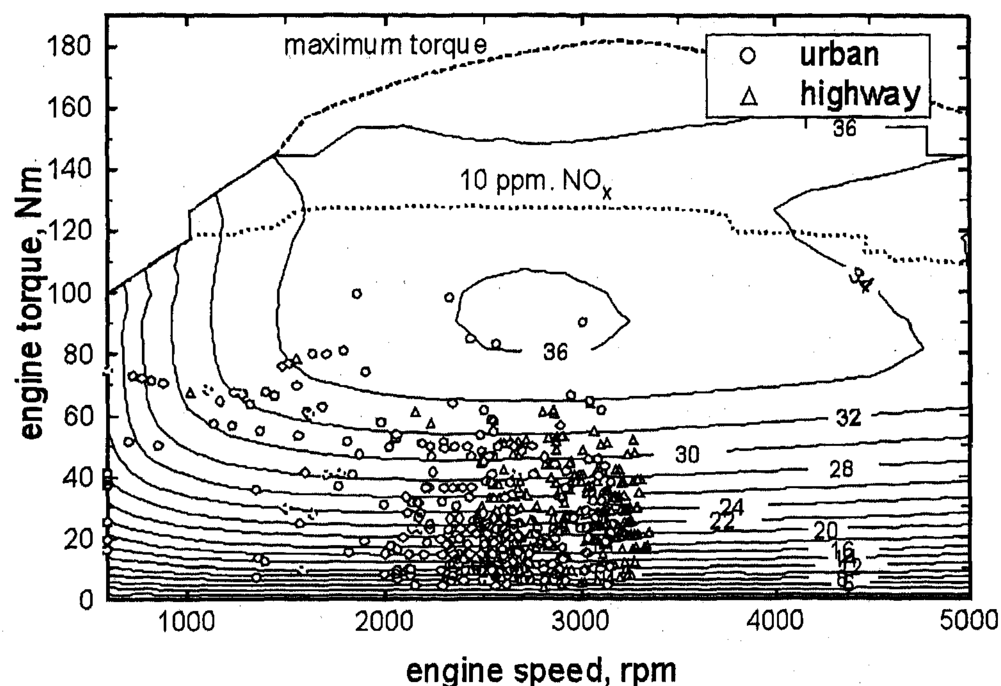


Figure 8. A representative set of engine operating points for the conventional car, for both the urban and highway cycles, superimposed on the engine efficiency map (repeated from Figure 3).

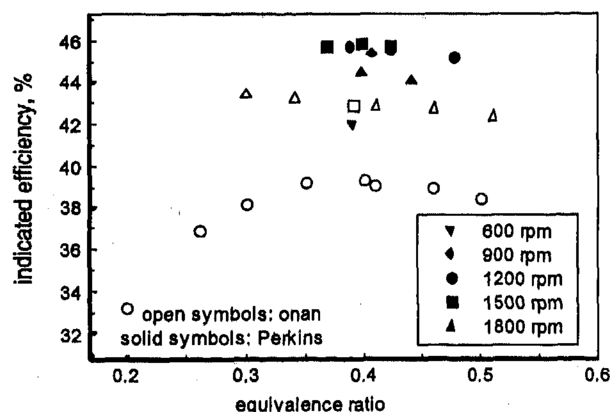


Figure 9. Indicated efficiency as a function of equivalence ratio and engine speed for the Perkins engine (solid symbols) and the Onan engine (open symbols, repeated from Figure 1 for comparison).

urban driving cycle. For the series hybrid, the engine runs at constant power until it fully charges the flywheel, which takes about 180 seconds. The flywheel then provides all the energy requirements, discharging in about 600 seconds. The engine is then turned on again to repeat the cycle. For the parallel hybrid, the flywheel is kept at about 50% state of charge, ready to store energy during regenerative braking, or provide energy for a sudden acceleration. The flywheel state of charge increases due to regenerative braking, and drops when power is extracted from the flywheel. Flywheel stand-by losses are projected to be of the order of 80 W and also tend to reduce the state of charge.

Figure 6 shows engine power for the conventional vehicle during the urban driving cycle. The driving cycle speed is indicated by a dotted line. Minimum engine power output is 1000 W, due to the need to provide for accessories. Figure 7 shows the nearly identical engine power output for the parallel hybrid vehicle. Power out of the two engines is very similar, except that peak power out of the parallel hybrid is reduced by 1 to 3 kW due to the power provided by the flywheel. Another difference between the output powers is that minimum power out of the parallel hybrid engine is zero, since the flywheel can now provide the accessory loads.

Figure 8 shows a representative set of engine operating points for the conventional car, for both the urban and highway cycles, superimposed on the engine efficiency map (repeated from Figure 3). The

points indicate that the engine operates at low torque during most of the time, therefore generating very low emissions. Efficiency could be improved by relaxing the 10 ppm NO_x emission limit, and this limit could be relaxed while still achieving EZEV levels. However, a vehicle that can emit more than 10 ppm of NO_x would not be intrinsically an EZEV, and would generate high NO_x if driven aggressively at high torque.

EXPERIMENTAL HYDROGEN ENGINE EVALUATION UPDATE

An experimental evaluation of a new hydrogen engine (Perkins) has been recently done [4]. The main characteristics of this engine are listed in Table 4. The cylinder is bigger than for the Onan engine (834 cm^3 compared to 493 cm^3), therefore reducing heat transfer losses. Preliminary results are shown in Figure 9, which gives indicated efficiency as a function of equivalence ratio and engine speed for the Perkins engine and the Onan engine (repeated from Figure 1 for comparison). The indicated efficiency for the Perkins engine is higher than for the Onan engine, and it is very insensitive to equivalence ratio and engine speed. Future runs will attempt to increase the engine efficiency by optimizing the turbulence within the cylinder to minimize heat transfer and timing losses.

Table 4. Perkins engine characteristics and experimental conditions.

Bore, mm	91.44
Stroke, mm	127.0
Displacement, cm^3	834.0
Geometric compression ratio	14.04:1
Experimental range for equivalence ratio	0.36-0.47
Experimental range for engine speed, rpm	600-1800
Experimental range for volumetric efficiency, %	90-110

CONCLUSIONS

This paper presents the development and validation of a time-dependent engine model, and its application to conventional and hybrid vehicles. The engine model is applied to a hydrogen engine which has been experimentally tested. The model predicts accurately engine efficiency and NO_x emissions over the full range of experimental operating conditions. The validated model is then used to generate engine performance and emission maps for supercharged engine operation. The performance maps are then incorporated into a vehicle evaluation code to obtain performance and emissions for hydrogen-fueled conventional, parallel and series hybrid vehicles. All vehicles are specified to meet minimum acceleration, hill climb and range requirements that make them comparable to current automobiles. Analysis of these vehicles yields the following results:

1. Emissions out of the conventional car are projected to be a factor of 60 lower than the CARB ULEV requirements, and therefore a factor of 6 lower than EZEV. The engine control strategy presented in this paper guarantees that the conventional vehicle achieves EZEV emissions levels regardless of how the car is driven. The conventional vehicle also has a high fuel economy.

2. The parallel hybrid vehicle uses a downsized engine and a small flywheel for power augmentation and regenerative braking. Emissions are also well within EZEV range. Fuel economy falls between the conventional and the series hybrid cars.

3. The series hybrid vehicle has the highest fuel economy. Emissions are higher than for conventional cars, at the EZEV limits. Emissions in a series hybrid vehicle are intrinsically independent of driver's input.

These results indicate that lean-burn hydrogen spark-ignited engines are an alternative to providing EZEV emissions, while at the same time providing a range, acceleration and hill climbing performance comparable to conventional cars. These vehicles are intrinsically EZEV, since emissions are not controlled with a catalytic converter that may degrade, and are also independent of the driving habits of the driver.

REFERENCES

1. Smith, J.R., 1994, "Optimized Hydrogen Piston Engines," Proceedings of the 1994 International Congress on Transportation Electronics, Convergence 1994, SAE, pp. 161-166.

2. Heywood, J.B., 1988, Internal Combustion Engine Fundamentals, McGraw-Hill, New York.

3. California Air Resources Board, Mobile Sources Division, 1995, "Proposed Amendments to the Low-Emission Vehicle Regulations to Add an Equivalent Zero-Emission Vehicle (EZEV) Standard and Allow Zero-Emission Vehicle Credit for Hybrid Electric Vehicles," Preliminary Draft Staff Report, CARB, El Monte, CA, July 14.

4. Van Blarigan, P., 1996, "Development of a Hydrogen-Fueled Internal Combustion Engine Designed for Single Speed/Power Operation," Proceedings of the 1996 SAE Future

Transportation technology Conference and Exposition, Vancouver, BC.

5. Aceves, S.M., and Smith, J.R., 1995, "A Hybrid Vehicle Evaluation Code and Its Application to Vehicle Design," SAE paper 950491.

6. Pehr, K., 1996b, "Aspects of Safety and Acceptance of LH2 Tank Systems in Passenger Cars," International Journal of Hydrogen Energy, Vol. 21, pp. 387-395.

7. Ferguson, C.R., 1986, "Internal Combustion Engines, Applied Termosciences," John Wiley and Sons, New York.

8. Haney, S.W., Barr, W.L., Crotinger, J.A., Perkins, L.J., Solomon, C.J., Chaniotakis, E.A., Freidberg, J.P., Wei, J., Galambos, J.D., and Mandrekas, J., 1992, "A SUPERCODE for Systems Analysis of Tokamak Experiments and Reactors," Fusion Technology, Vol. 21, p. 1749.

9. Woschni, G., 1967, "Universally Applicable Equation for the Instantaneous Heat Transfer Coefficient in the Internal Combustion Engine," SAE Paper 670931.

10. Patton, K.J., Nitschke, R.G., and Heywood, J.B., 1989, "Development and Evaluation of a Friction Model for Spark-Ignition Engines," SAE Paper 890836.

11. Olsson K., and Johansson, B., 1995, "Combustion Chambers for National Gas SI Engines Part 2: Combustion and Emissions," SAE paper 950517.

12. Thomson, M.W., Frelund, A.R., Pallas, M., and Miller, K.D., 1987, "General Motors 2.3L Quad 4 Engine," SAE Paper 870353.

13. Aceves, S.M., and Smith, J.R., 1996, "Lean-Burn Hydrogen Spark-Ignited Engines: The Mechanical Equivalent to the Fuel Cell," Proceedings of the 18th Annual Fall Technical Conference of the ASME Internal Combustion Engine Division, ICE-Vol. 27-3, Alternative Fuels, Edited by J.A. Caton, pp.23-32.

14. Peschka, W., 1998, "Liquid Hydrogen, Fuel of the Future," Springer-Verlag, Vienna, Austria.

970294

Parametric Design of the Drive Train of an Electrically Peaking Hybrid (ELPH) Vehicle

Yimin Gao[1], Khwaja M. Rahman, and Mehrdad Ehsani
Texas A&M Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The operation of an electrically peaking hybrid vehicle (ELPH) can be divided into two basic modes. • Constant or cruising speed mode in which a small internal combustion engine (ICE) is used to power the vehicle. • Peak power mode in which the combination of an electric motor and ICE is used to supply peak power for acceleration and limited-duration steep hill climbing of the vehicle.

A method, by which the engine size and the speed reduction ratio from the engine to drivewheels can be developed based on the cruising mode, is presented in this paper. The electric motor power rating and the motor gear ratio to the drive wheels can then be determined, based on the acceleration and gradeability. The results show that a simple single-gear transmission would be a good selection for overall performance.

INTRODUCTION

Petroleum fueled, internal combustion engine powered vehicles are the most popular means of transportation, because of their high energy density and relatively high power density. However, in recent years, increasing concern over air pollution, caused by tailpipe emissions of the petroleum-based vehicles, and the dwindling petroleum resources have caused the automotive engineers and automakers to re-evaluate the designs of the conventional internal combustion engine powered vehicles. The conventional automotive drive train suffers from a number of disadvantages :

1. The inherently mismatched speed-torque characteristic of the engine and the vehicle necessitates a complicated transmission, with its associated losses and inflexibility.

2. In order to have ample power for acceleration and gradeability with a limited number of transmission gears, the engine must be oversized to roughly ten times that required for cruising at *100 Km/h* on a level road and three or four times that required for maintaining *100Km/h* on a *6%* grade.[1]

3. Today's internal combustion engines show a significant difference in specific fuel consumption between partial load and the optimal operating point, which is close to full load as shown in *Fig. 1*. An oversized engine moves the cruising operating point away from the optimal operation point. Consequently, ICE vehicles have low efficiency [2].

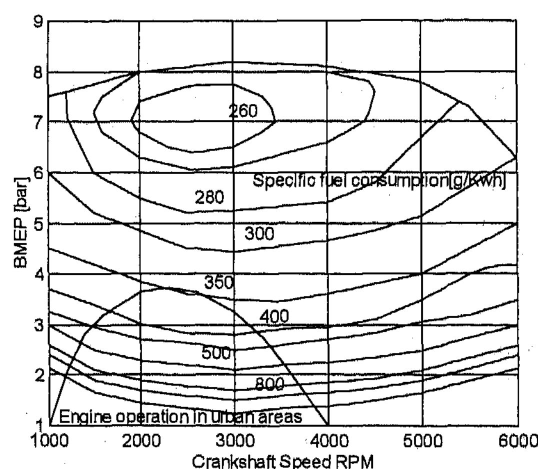


Fig. 1. Typical SI-engine SFC Map with Actual Operation Point[2]

The electric motor, in addition to its clean, quiet and efficient operation, has inherent flexibility in its speed-torque control. *Fig. 2* shows a typical torque-speed characteristics of the electric motor with power electronic drive. The motor can operate anywhere within this torque-speed boundary. This

[1] Visiting scholar from Jilin University of Technology, China

allows the motor to meet the vehicle torque-speed requirements at all times, without a variable gear transmission. However, pure electric vehicles (Ev's) suffer from other disadvantages[9].

1. The heavy and bulky battery pack, with its relatively small energy storage, makes the EV limited in range, and load carrying capacity.

2. Long charging time limits the EV's practicality.

Therefore, commercial success of the EV depends strongly on the development of advanced batteries. However, progress in batteries over the past several decades has not been adequate.

The hybrid configurations, in which two power sources are applied to propel the vehicle, show significant promise[7,8]. The hybrid electric-internal combustion engine drive train, if properly configured, can combine many of the advantages of both EV and ICE vehicles [9].

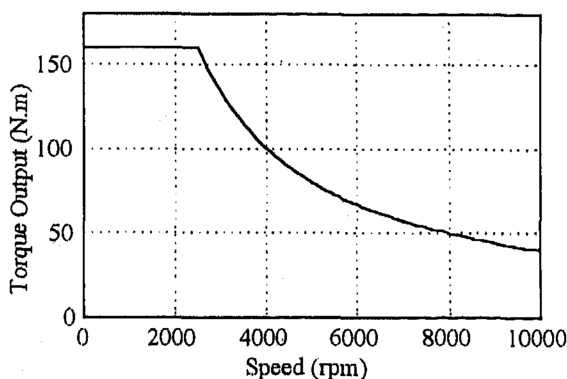


Fig. 2. The Speed-torque Characteristic of Electric Motor

ELECTRICALLY PEAKING HYBRID CONFIGURATION

The operation of passenger vehicles can be divided into two basic modes: constant speed (cruising), and acceleration (peak power). In cruising mode, relatively low power is required from the drive train. However, large amount of energy is consumed in a long trip. In acceleration mode, high peak power is required but not much energy is consumed, due to its brief and transient occurrence. Proper parallel combination of ICE, for cruising operation, and electric motor with a small battery pack, for acceleration, can satisfy these requirements in a viable drive train. The motor with a small chemical battery pack optimized for high power would have acceptable power capability and power density for acceleration[3]. A small ICE with a small fuel tank would have the energy density for cruising and recharging the battery for the next acceleration.

A configuration of such an electrically peaking hybrid (ELPH) vehicle is shown in *Fig. 3* in which the base power unit is the engine and peaking power unit (or load leveling device) is the electric motor.

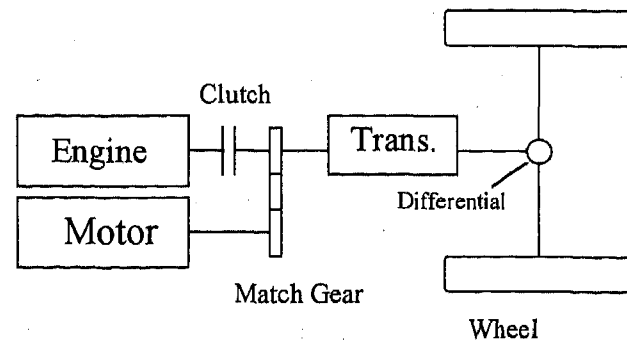


Fig. 3 The Configuration of an Electrically Peaking Hybrid (ELPH) Vehicle

The ELPH configuration, in addition to satisfying the requirements of acceleration and gradeability, has the ability to recover braking energy with the electric motor functioning in regenerating mode. Furthermore, when the vehicle operates with light load, such as at a low constant cruising speed or going down a slight hill, the engine can recharge the battery pack to maintain adequate state-of-charge. More beneficially, this enhances the engine load, for operation close to its optimal point. A well designed ELPH vehicle may never use an external source to charge its battery pack and can achieve an excellent fuel economy.

In this paper, the following specification of the Electrically Peaking Hybrid (ELPH) vehicle prototype, which is being developed at Texas A&M University, will be taken as the example.

Curb weight	*1700 kg,*
Rolling resistance coefficient of tire	*0.013,*
Aero-dynamic drag coefficient	*0.29,*
Front area	*2.13 m^2,*
Wheel radius	*0.2794 m.*

CRUISING MODE

As described above, when the vehicle operates on level ground with a constant cruising speed, the engine alone delivers power to the drivewheels. Consequently, the performance of the vehicle is determined only by the engine power and speed reduction ratio from the engine to the drivewheels.

ENGINE POWER REQUIREMENT- The determination of the engine power plays a very important role in the design of ELPH vehicle. An oversized engine would lose its fuel-saving advantage as in the conventional vehicles. On the other hand, an undersized engine can not deliver enough power to meet the requirements of the vehicle. Therefore, a large battery pack and a large motor would be needed to compensate for the shortage of engine power. Consequently, the ELPH vehicle would suffer from the same problems as an EV.

When the vehicle operates on a level ground with a constant cruising speed, the engine must deliver enough power to overcome the vehicle load power which includes the rolling resistance power and the aero-dynamic drag power. Thus, the required engine power output is

$$Pe = \frac{v}{1000\eta_{et}}\left(mgf + \frac{1}{2}\rho C_D A v^2\right) \qquad (1)$$

where, v = speed of vehicle, *m/s*,

η_{et} = efficiency of the transmission form engine to drivewheels, *0.9*

m = vehicle mass, *Kg*,

f = rolling resistance coefficient,

g = gravity acceleration, *9.81 m/s²*,

C_D = aero-dynamic drag coefficient,

A = front area of the vehicle, m^2,

ρ = air density, *1.228 Kg/m³*.

Fig. 4. shows the vehicle load power as a function of the vehicle speed in cruising mode This figure indicates that not much power is needed to maintain the vehicle operation in cruising mode. About *30 Kw* of power capacity will be plenty to meet the requirement of the vehicle at a speed of *130 Km/h (81.5 mph)*. This *30 Kw* of engine power capacity is quite small in comparison to the average *50* to *70 Kw/1000Kg* specific power of conventional cars More precise selection of the engine power capacity will be made with consideration of recharging of the peaking battery pack

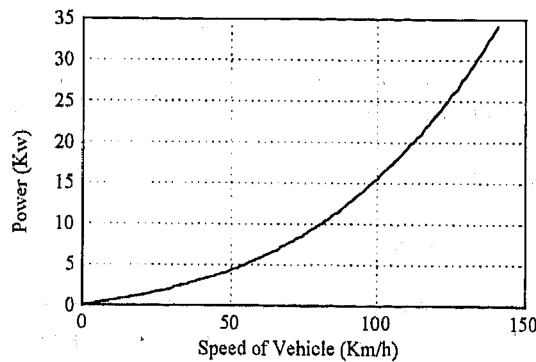


Fig. 4. Required Engine Power Versus Constant Speed on A Level Road

DIFFERENTIAL GEAR RATIO-*Fig. 5.* shows the speed-power characteristics of a typical gasoline SI engine. Equation (2) gives the vehicle speed for a single-gear transmission or a top gear operation of a multi-gear transmission. By equation (2), *Fig. 5* can be transformed into a vehicle speed-power profile, as shown in *Fig. 6*,

$$v = \frac{\pi r N}{30 r_d} \quad (m/s) \qquad (2)$$

where, r_d = differential gear ratio drivewheels,

r = radius of the drivewheel, *m*

N = speed of engine, *rpm*.

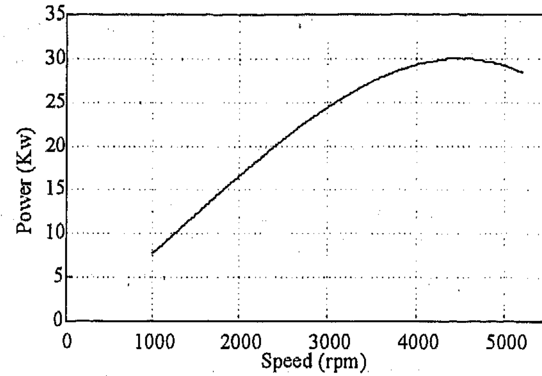


Fig. 5. Typical Characteristics of SI Engine

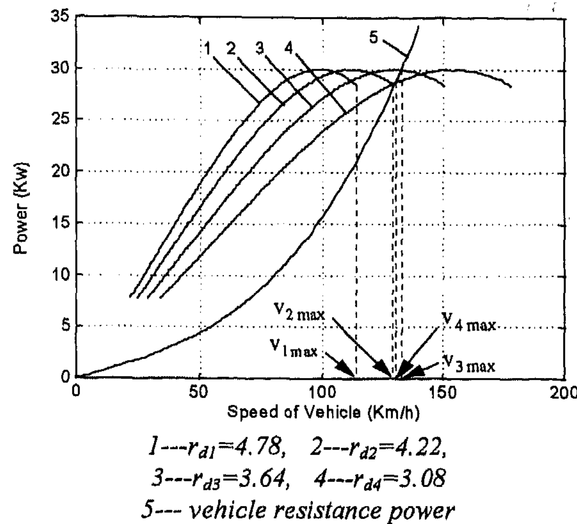
In *Fig. 6.* several engine power curves with different differential gear ratios are shown along with the vehicle resistance power curve. The top speeds of the vehicle with different differential gear ratios are represented by the crossing points of engine power curves with resistance curve (such as v_{2max}, v_{3amx} and v_{4max} corresponding to r_{d2} , r_{d3}, and r_{d4}) or by the maximum rpm of the engine (such as v_{1max} corresponding to r_{d1}). *Fig. 6* indicates that, with r_{d3}, the vehicle has a highest top speed which corresponds to the peak power of the engine. However, with r_{d2}, the vehicle has more remaining power, defined as the difference between the engine power curve and vehicle resistance power curve only a slight top speed reduction. The remaining power can be used for acceleration or grade climbing or battery charging . As shown in *Fig. 6*, with a gear ratio of r_{d2} and at a speed of *100 Km/h*, the engine has about half of its power remaining. A large gear ratio, such as r_{d1}, will allow the vehicle to have more remaining, but the top speed will suffer a significant reduction. In contrast, too small a gear ratio, such as r_{d4}, will allow the vehicle to have a small remaining power. For a small engine with the addition of battery charging load, the differential gear ratio of r_{d2} is a suitable selection.

ACCELERATION MODE

When the vehicle experiences acceleration or steep hill climbing, high traction peak power is needed. This high transient power (peaking power) is supplied by both the engine and the motor working together. In the ELPH vehicle, the contribution of the electric motor drive to the peak power should be dominant. Thus, the acceleration and maximum gradeability of the ELPH drive train is mainly determined by the electric motor drive and battery power rating.

ACCELERATION-The ideal characteristic of a power unit for vehicular application is constant power output over the full speed range[4,5,6,]. A vehicle with this kind of power unit does not require a multi-gear transmission to enhance its acceleration and gradeability performance. A well controlled electric motor has speed-torque characteristics that are close

to this ideal, as shown in *Fig. 2*. Here the motor has a constant power output over a large range of speed and a constant torque output over the low speed range. Therefore, the acceleration performance of the ELPH vehicle will be determined by the power rating of the electric motor and is only slightly influenced by the gear number and gear ratios of the transmission.



1---r_{d1}=4.78, 2---r_{d2}=4.22,
3---r_{d3}=3.64, 4---r_{d4}=3.08
5--- vehicle resistance power

Fig. 6. Engine Power Output And Resistance Power Versus Vehicle Speed with Different Differential Gear Ratios

The vehicular acceleration on a level ground can be expressed as:

$$\frac{dv}{dt} = \frac{F_t - F_w - F_f}{\delta m} \quad (m/s^2) \qquad (3)$$

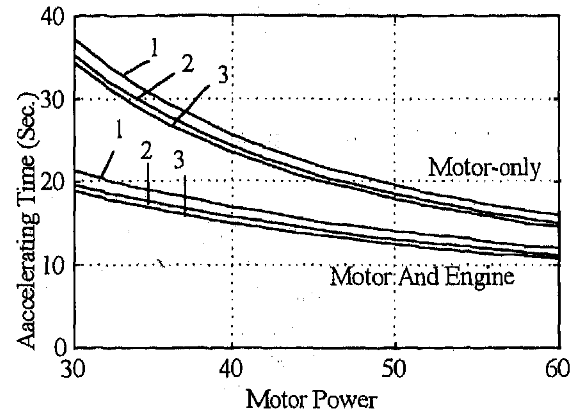
where, $F_t = (T_e + T_m r_m) r_t r_d \eta_t / r$, the thrust force of ground acting on the drivewheels (T_e and T_m are engine torque and motor torque respectively; r_t and r_d are the transmission and differential gear ratios, respectively; r_m is gear ratio from motor to driveshaft.), $F_w = 0.5\rho C_D A v^2$, the aero-dynamic drug, and $F_f = mgf$, the rolling resistance. δ is the rotational inertia coefficient.

The accelerating time from zero to a desired speed , v^*, can be expressed as:

$$t = \int_0^{v^*} \left(\frac{dt}{dv}\right) dv = \int_0^{v^*} \frac{\delta m}{F_t - F_w - F_f} dv \qquad (4)$$

The calculation of the vehicle acceleration time is somewhat difficult due to the speed-torque characteristic of the engine which, generally, can not be expressed by an explicit formula. However, it may be completed by numerical methods with a computer.

Fig. 7 shows a plots of acceleration times from zero to *96 Km/h* vs. electric motor power with different gear ratios from the motor to the driveshaft, r_m . The assumptions are a *30 Kw* engine, motor characteristics as shown in *Fig. 2*, a differential gear ratio of r_d = *4.23* and transmission gear ratio of r_t = *1.0*.

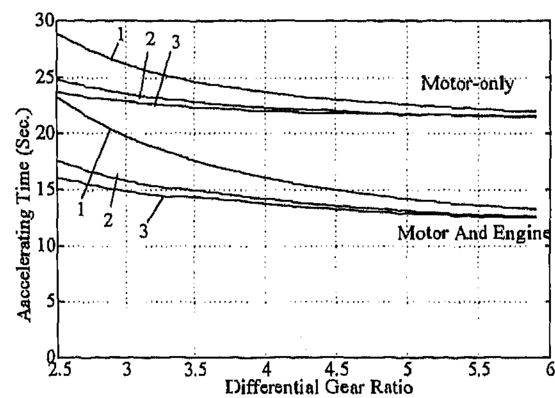


1---r_m=1.5, 2---r_m=1.9, 3---r_m=2.3

Fig. 7. The Accelerating Time Versus Electric Motor Power

As we expected, the gear ratio from the motor to the driveshaft has a small influence on the acceleration time and, the electric motor power plays a dominant role in the acceleration of the vehicle.

Using a computer program, the acceleration time of the vehicle with a multi-gear transmission can also be calculated, as shown in *Fig. 8*. The results indicate that the number of gears in the transmission has a small influence on the acceleration of the vehicle when the differential gear ratio is greater than *4*. This result is desirable because a single-gear transmission can greatly simplify the drive train and the control system.



1- gear transmission, gear ratio, 1
2--2-gear transmission , gear ratios , 1.6,1
3--3-gear transmission , gear ratios, 2.54, 1.6 and 1
Engine power = 30 Kw,
Electric motor power = 42 Kw
Gear ratio from motor to driveshaft, r_m=1.7

Fig. 8. Accelerating Time with Different Number of Gears and Different Differential Gear Ratios

GRADEABILITY-The gradeability of the vehicle is determined by the maximum thrust forces acting on the drivewheels. The gradeability of a vehicle can be expressed by

$$\sin\alpha = \frac{F_t - F_w - F_f}{mg} \quad (5)$$

where α is the road angle in degrees.

Due to the small aero-dynamic drag at low speeds, the gradeability of the vehicle is mostly determined by the maximum torque of the motor and the engine as well as the gear ratios of the transmission. *Fig. 9* shows the gradeability of the vehicle with a three-gear transmission in which the parameters used are the same as in *Fig. 8*. *Fig. 9* indicates that the gradeability will be greatly enhanced by using a multi-gear transmission. However, in real applications, the vehicle would seldom use the first or second gear for climbing a grade, because, normally, such large grades are seldom encountered in highway and urban driving. This also means that a single-gear transmission would serve the gradeability well.

Fig. 10 shows gradeability versus different power outputs with a single-gear transmission in which the parameters are the same as in *Fig. 8*. Actually, the gradeability can be enhanced, without a need for increasing the gear ratio or applying a multi-gear transmission, by decreasing the base speed of the electric motor. This is accomplished in the low speed, constant torque region of the motor operation

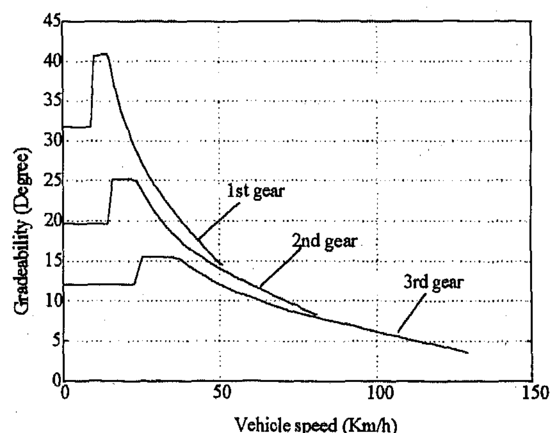


Fig. 9. Gradeability with A Multi-gear Transmission

CONCLUSION

The ELPH vehicle can combine many advantages of conventional and electric vehicles. A well designed ELPH vehicle would have a comparable performance to the conventional vehicle and excellent fuel economy and emission characteristics.

Using the method presented in this paper, the power capacity of a small internal combustion engine and a single speed reduction gear can be used to achieve a satisfactory performance and a excellent fuel economy at cruising speed. Properly selected electric motor power and speed reduction gear can eliminate application of the complicated multi-gear transmission with satisfactory acceleration and hill climbing performance.

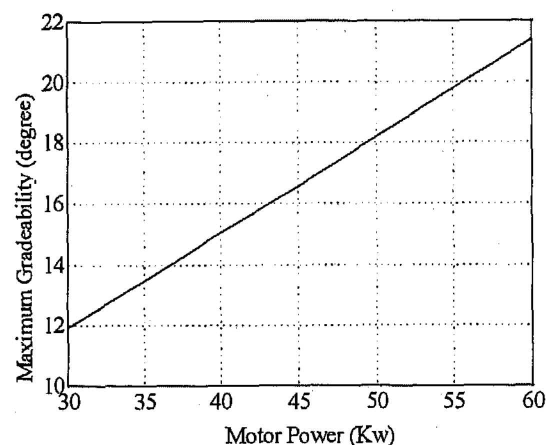


Fig. 10. Gradeability of the Vehicle with Different Motor Power Output And A Single-gear Transmission

ACKNOWLEDGMENT

The financial support of Texas Higher Education Coordinating Board and Texas Transportation Institute for the ELPH project is gratefully acknowledged.

REFERENCE

[1] Timothy C. More and Armory B. Loind, *Vehicle Design Strategies to Meet and Exceed PNGV Goals,* SAE 951906.

[2] Martin Endar and Philipp Dietrich, *Duty Cycle Operation as a Possibility to Enhance the Fuel Economy of an SI Engine at Part Load,* SAE 960229.

[3] A.F. Burke, *Battery Availability for Near Term (1998) Electric Vehicle,* SAE 911914.

[4] J.Y. Wong , *The Theory of Ground Vehicle* , John Wiley & Sons Inc. Press, 1978.

[5] M.Ehsani, K. M. Rahman, and H. Toliyat, *Propulsion System Design of Electric & Hybrid Vehicle,* accepted for publication in the IEEE Tran. on Industrial Electronics.

[6] H.A. Toliyat, K. M. Rahman, and M. Ehsani, *Electric Machine in Electric & Hybrid Vehicle Application,* Proceeding of ICPE, 95, Seoul, pp 627--635.

[7] M.Ehsani, *Electrically Peaking Hybrid System And Method,* U.S. Patent pending.

[8] J. Howze, M. Ehsani and D. Buntin, *Optimizing Torque Controller for a Parallel Hybrid Electric Vehicle,* U.S. Patent pending.
[9] Proceedings of the ELPH Conference, Texas A&M University, College Station, Texas, Oct. 1994.

970378

Theoretical Consideration of Relation of Rear-Wheel Skid to Steering Inputs

Hideki Sakai
Toyota Motor Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper examines the steering inputs that cause rear-wheel skid. First, we compare the front and rear side force at maximum lateral acceleration and thus distinguish rear-wheel skid from front-wheel skid. Next, we examine the affect of forward velocity on rear-wheel skid, and show that rear-wheel skid cannot occur below a given forward velocity. We use this velocity as an index of rear-wheel skid tendency and formulate it. This formula shows that the difference between front and rear axle trajectories strongly affects rear-wheel skid.

INTRODUCTION

This paper investigates the kind of steering inputs that make vehicles skid on their rear wheels, extending the Force and Moment method[1] to the frequency domain.

As shown in Figure 1, the stability factors of almost all passenger cars are always positive even at maximum lateral acceleration. In other words, they tend to front-wheel skid in steady-state turning (i.e. understeer) at all times. However, these vehicles also often tend to rear-wheel skid with rapid steering inputs.

From the Force and Moment method, in front-wheel skid, the front axle generates its maximum side force at the maximum lateral acceleration, while the rear axle does not. This paper applies this concept to behavior with sinusoidal steering inputs in order to investigate rear-wheel skids with rapid steering inputs.

VEHICLE BEHAVIOR WITH SINUSOIDAL STEERING INPUTS

MODEL - Figure 2(A) shows the vehicle model[1][2]. Transforming Fig. 2(A) into Fig. 2(B) as described in the particle system, we investigate vehicle behavior. Using Fig. 2(B), we will focus on the difference between the particles lateral motion in the discussion section. The motion equations of this vehicle model are in the form

$$m\,u\left(\frac{d\beta}{dt}+r\right)=F_f+F_r \tag{1}$$

$$I_d\, m\, l_f\, l_r\,\frac{dr}{dt}=l_f\,F_f-l_r\,F_r \tag{2}$$

$$F_i=CF_i\;F_{zi} \tag{3}$$

Figure 3 shows the tire model[2]. We assume dynamic properties of the tires based on Reference 3.

The motion equations of the tire model are in the form

$$CF_i=C_i\left[1-\frac{4}{27}\left(\frac{C_i}{\mu_i}\right)^2\alpha_{\text{ground }i}{}^2\right]\alpha_{\text{ground }i} \tag{4}$$

$$\alpha_{\text{ground }i}=\alpha_i-\frac{v_{\text{ground }i}}{u} \tag{5}$$

$$v_{\text{ground},i}=\frac{d}{dt}\,y_{\text{tire }i} \tag{6}$$

$$y_{\text{tire }i}=\frac{\frac{1}{2}F_i}{k_{\text{tire}i}}=\frac{CF_i}{ck_{\text{tire}i}} \tag{7}$$

$$\alpha_f=\beta+\frac{l_f\,r}{u}-\delta \tag{8}$$

$$\alpha_r=\beta-\frac{l_r\,r}{u} \tag{9}$$

$$ck_{\text{tire}i} = \frac{k_{\text{tire}i}}{\frac{1}{2}F_{zi}} \qquad (10)$$

where $\alpha_{groundi}$ is the slip angle at the contact region of the tire, $y_{groundi}$ is lateral deformation of the tire, and k_{tirei} is tire lateral stiffness. In addition, we term C_i *effective cornering coefficient* and μ_i *effective friction coefficient.*

The front wheels skid when $CF_f = \mu_f$ and conversely the rear-wheel skid when $CF_r = \mu_r$. Thus the comparison of μ_i implies the skid tendency in steady state turning as follows[1]:

1) If $\mu_r > \mu_f$, the vehicle tends to front-wheel skid because the lateral acceleration of the vehicle cannot exceed μ_f, which is maximum lateral acceleration.

2) If $\mu_f > \mu_r$, the vehicle tends to rear-wheel skid.

This paper assumes $\mu_r > \mu_f$ and $C_r > C_f$; this condition represents almost all current passenger cars (Appendix A).

INPUT - This paper adopts a sinusoidal steering input at various frequencies as the input into the vehicle model. This steering angle is set to encourage the maximum lateral acceleration. This steering angle is expressed by

$$\delta = \delta_0(\omega) \sin \omega t \qquad (11)$$

where $\delta_0(\omega)$ is the steering angle that promotes the maximum lateral acceleration at ω.

OUTPUT - The tire model has nonlinearity, and consequently the vehicle response contains harmonics. Hence, we extract the fundamental component from the output of the calculation using a Fourier series.

DEFINITION OF SKIDDING IN THE FREQUENCY DOMAIN - We can extend the definition of skidding defined in the Force and Moment Method to the frequency domain as follows:

1) If $CF_f = \mu_f$, the front wheel skids.

2) If $CF_r = \mu_r$, the rear wheel skids.

RESULTS - Figure 4(A) shows calculations of the side forces coefficients and maximum lateral acceleration ($A_{max}(\omega)$) for both axles. These calculations were made to detect the optimum $\delta_0(\omega)$ through many trials.

In Fig. 4(A), we can divide the frequency band as follows:

1)Frequency where $CF_f = \mu^*_f$, namely, the front wheel skids.

2)Frequency where $CF_r = \mu^*_r$, namely, the rear wheel skids.

At the rear-wheel skid frequencies, the rear particle moves in the lateral direction more than the front particle, as shown in Fig. 4(B). In particular, when I_d=1, the ratio of the lateral displacement of the rear particle to that of the front exceeds μ^*_r/μ^*_f (>1) at these frequencies.

In addition, μ^*_i in Fig. 4(A) disagrees with μ_i in Fig. 3; this phenomenon is caused by one of the propertie of the filter, which will be described in the discussion section.

SLALOM DIAGRAM - Figure 5 is transformed from Fig. 4(A). We call this figure the *slalom diagram.* The slalom diagram enables us to understand the skid tendency. $A_i(\omega)$ is the "extended" maximum lateral acceleration at the center of gravity assuming that the other axle of axle i does not reach effective friction coefficient; $A_i(\omega)$ is defined by

$$A_i(\omega) = \frac{\mu^*_i}{|CF_i(\omega)|}|A_{\max}(\omega)| \qquad (12).$$

The larger one of $A_i(\omega)$ cannot be realized and the smaller one equals $A_{max}(\omega)$. This figure shows the skid tendency as follows:

1) If $A_r(\omega)>A_f(\omega)$, the front wheel skids.

2) If $A_f(\omega)>A_r(\omega)$, the rear wheel skids.

This agrees with the definition of front/rear-wheel skid tendency in the frequency domain.

To clarify this point, we will focus on the response, for instance, at about 1.1 Hz. As shown in Fig. 4, $CF_r(\omega)$ corresponds to μ_r^* at about this frequency. By substituting $CF_r(\omega)=\mu_r^*$ into EQ (12), EQ (12) becomes $A_r(\omega)=A_{max}(\omega)$. On the other hand, since $CF_f(\omega) < \mu_f^*$, $A_f(\omega)>A_{max}(\omega)$ from EQ (12). Since $A_f(\omega)>A_r(\omega)$ (=$A_{max}(\omega)$), the vehicle tends to rear-wheel skid at this frequency. The slalom diagram thus shows $A_{max}(\omega)$ and skid tendency.

LINEAR CALCULATION FOR SLALOM DIAGRAM

This section describes a linear calculation method for the slalom diagram. The reasons for the linearization are as follows:

1) In the simulation method in the previous chapter, we need much time for many calculations to detect the optimum steering angle that promotes maximum lateral acceleration at each frequency.

2) This numerical simulation is unsuitable for theoretical interpretations of $A_i(\omega)$ because $A_i(\omega)$ is not formulated in the simulation.

Motion Equations - Since we adopt sinusoidal inputs, we can describe the motion equation using the $j\omega$ operator form. Linearizing CF_i (APPENDIX B), we can reduce EQS (1) through (9) to

$$m\,u\,(\beta j\,\omega + r) = CF_f\,F_{zf} + CF_r\,F_{zr} \qquad (13)$$

$$I_d\,l_f\,l_r\,m\,r\,j\,\omega = l_f\,CF_f\,F_{zf} - l_r\,CF_r\,F_{zr} \qquad (14)$$

$$CF_f = -\left(1 - \frac{l^*_{df}}{u} j\,\omega\right) C^*_f\,\alpha_f \qquad (15)$$

$$CF_r = -\left(1 - \frac{l^*_{dr}}{u} j\,\omega\right) C^*_r\,\alpha_r \qquad (16)$$

$$\alpha_f = \beta + \frac{l_f\,r}{u} - \delta \qquad (17)$$

$$\alpha_r = \beta - \frac{l_r r}{u} \quad (18)$$

where

$$\frac{C^*_i}{C_i} = \frac{3}{4} \quad (19)$$

$$\frac{l^*_{di}}{l_{di}} = \frac{1}{\sqrt{3}} \quad (20)$$

$$l_{di} = \frac{C^*_i}{ck_{tirei}} \quad (21).$$

In these equations, l^*_{di} is the tire relaxation length. We treat this length as negative time-constant of a first-order-lead-system based on the "Complex Cornering Stiffness" concept[4][5]. Solving EQS (13) through (21), we obtain

$$\frac{A_f(\omega)}{g} = \left| \frac{\frac{(j\omega)^2}{\omega_s^2} + \frac{2\zeta_s j\omega}{\omega_s} + 1}{\frac{(j\omega)^2}{\omega_{skid}^2} + \frac{2\zeta_{skid} j\omega}{\omega_{skid}} + 1} \right| \mu^*_f \quad (22)$$

$$\frac{A_r(\omega)}{g} = \left| \frac{\frac{(j\omega)^2}{\omega_s^2} + \frac{2\zeta_s j\omega}{\omega_s} + 1}{T_{skid}\, j\omega + 1} \right| \mu^*_r \quad (23)$$

$$\frac{1}{\omega_{skid}^2} = \frac{I_d}{C^*_r}\frac{l}{g} - cic_r \frac{l^*_{dr}}{l}\left(\frac{l}{u}\right)^2$$

$$\frac{2\zeta_{skid}}{\omega_{skid}} = \left(cic_r - \frac{l^*_{dr}}{l}\right)\frac{l}{u}$$

$$\frac{1}{\omega_s^2} = \frac{I_d}{C^*_r}\frac{l_r}{g} - \frac{l_r}{l}\frac{l^*_{dr}}{l}\left(\frac{l}{u}\right)^2$$

$$\frac{2\zeta_s}{\omega_s} = \left(\frac{l_r}{l} - \frac{l^*_{dr}}{l}\right)\frac{l}{u}$$

$$T_{skid} = (I_d - 1)\frac{l_r}{u}$$

$$cic_r = \frac{I_d\, l_f + l_r}{l}$$

(Appendix C).

Figure 6 shows the calculation results of EQS (22) and (23). The comparison between the results of the linear calculation and those of the numerical simulation shows that we can use this linearized model for slalom diagram prediction.

INDEX FOR REAR-WHEEL SKID

This section proposes and formulates an index for rear-wheel skid to steering inputs.

INDEX - Figure 7 shows the rear-wheel skid tendency at various speeds. This figure illustrates that rear-wheel skid never occurs below a given speed for sinusoidal steering. We call this speed the *skid critical speed* (V_{skid}). V_{skid} is adopted as the index of rear- wheel skid in this paper. V_{skid} is described in the form

$$V_{skid}^2 = F_s V_{tt}^2$$

(when I_d=1 and l^*_{dr}=0) (24)

$$V_{skid}^2 = \frac{C^*_r}{C_r}\left(1 - \frac{l^*_{dr}}{l}\right)^2 \left[F_s + 2\frac{R^4}{\left(1 - \frac{l^*_{dr}}{l}\right)^2}\frac{l^*_{dr}}{l}\right] V_{tt}^2$$

(when I_d=1) (25)

$$V_{skid}^2 = \frac{C^*_r}{C_r}\frac{R^2}{I_d}\left[\frac{cic_r^2}{R^2} - (I_d - 1)^2\left(\frac{l_r}{l}\right)^2\right] F_s V_{tt}^2$$

(when l^*_{dr}=0) (26)

$$V_{skid}^2 = \frac{C^*_r}{C_r}\left[1 + \frac{l_r - l_f}{l}(1 - I_d) + 2\left(\frac{R^4}{F_s} - 1\right)\frac{l^*_{dr}}{l}\right] F_s V_{tt}^2$$

when $I_d \doteqdot 1$ and $l^*_{dr}/L \ll 1$) (27)

where

$$V_{tt}^2 = C_r \frac{l}{2} g \quad (28)$$

$$F_s = \left(1 + \sqrt{1 - \frac{1}{R^2}}\right) R^2 \quad (29)$$

$$R = \frac{\mu^*_r}{\mu^*_f} \quad (30)$$

(Appendix D), where

$$\mu^*_i = \frac{9}{8}\mu_i \quad (31)$$

(Appendix B)

$$cic_r = \frac{I_d\, l_f + l_r}{l} \quad (32)$$

$$I_d = \frac{I_Z}{m\, l_f\, l_r} \quad (33).$$

In EQ (24), we define the first term as *statics factor* (F_s) and the second as *track-on-track speed* (V_{tt}). F_s is similar to the "yawing moment coefficient," which is defined by the force and moment method[1]. That is, F_s and the yawing moment coefficient indicate the effect of maximum side forces on rear-wheel skid in statics. Thus F_s is closely related to the yawing moment coefficient. On the other hand, V_{tt} implies the effect of attitude angle on rear-wheel skid in dynamics. This is examined in detail in the discussion. In addition, we call *R friction balance ratio*.

SENSITIVITY OF V_{skid} - Figure 8 shows the sensitivity of V_{skid} to different parameters. We define this sensitivity as

$$sensitivity = \frac{\partial V_{skid}}{\partial x}\frac{x}{V_{skid}} \quad (34)$$

where x is the vehicle parameters in EQ (27).

EXPERIMENTS FOR VERIFICATION

Figure 9 compares the rear-wheel skid region of the linear calculations and that of the actual vehicle experiments. Since they agree closely, we can infer that the above studies are valid.

DISCUSSION

FRICTION BALANCE RATIO - Friction balance ratio (R) belongs to statics as well as "yawing moment coefficient[1]." Figure 8 shows that R enhances V_{skid}, which corresponds to the force and moment method[1]. However, the increase of R, i.e., the increase of F_s, decreases $A_{max}(\omega)$ in steady state turning ($=A_{max}(0)$) when the tires are not changed, as shown in Fig. 10.

Now R affects V_{skid} due to the statics factor (F_s). It follows from EQ (34) that the sensitivity of V_{skid} to F_s is approximately 1/2 because V_{skid} is roughly proportional to the square root of F_s. In particular, when I_d=1 and l^*_{dr}=0, this sensitivity is exactly 1/2 because V_{skid} is exactly proportional to the square root of F_s, as shown in EQ (34).

TRACK-ON-TRACK SPEED - Suppose $C^*_r=C_r$, I_d=1, l^*_{dr}=0 and R ->1; then V_{skid} approaches V_{tt} and ω_p (Appendix D) approaches 0 Hz, which is steady-state turning. Hence, we can examine rear-wheel skid during steady-state turning at V_{tt}.

Figure 11 illustrates V_{tt}; at V_{tt}, the track of the rear axle, on which one of particles lies, follows that of the front axle, on which the other lies. Thus, V_{tt} represents the attitude angle properties. Inward attitude increases the lateral displacement of the rear particle during a sinusoidal input, as shown in Fig. 4(B), and consequently generates the rear-wheel skid tendency. Thus, the attitude angle strongly affects the rear-wheel skid due to steering inputs.

DYNAMIC INDEX TO PLANE MOTION - This section discusses the affect of I_d on V_{skid}. This affect depends on the longitudinal distribution of mass (l_r/l). Namely, with decreasing I_d, V_{skid} increases when the front end is heavy (l_r>1/2) but decreases when the rear end is heavy. This is illustrated in Fig.12, which shows the attitude angles during steady-state turning at the speed at which the trajectories of two particles correspond. We define this speed as the *characteristic speed of particle model* (V_{par}). V_{par} can be expressed as

$$V_{par}^2 = \left[1+\frac{l_r - l_f}{l}\left(1-\sqrt{I_d}\right)\right] V_{tt}^2 \quad (35).$$

V_{par} is the general case of V_{tt}. That is, V_{par} corresponds to V_{tt} when I_d=1 and l^*_{dr}=0. Hence, we can deduce that V_{skid} increases with increasing V_{par} as well as increasing V_{tt}.

To verify this point, we will examine the V_{skid} formula that contains V_{par}. By substituting l^*_{dr}=0 into EQ (27), it becomes

$$V_{skid}^2 = \left[1+\frac{l_r - l_f}{l}\left(1-I_d\right)\right] F_s V_{tt}^2 \quad (27').$$

If we make EQ (35) a Taylor's expansion around I_d=1, we obtain

$$V_{par}^2 = \left[1+\frac{1}{2}\frac{l_r - l_f}{l}\left(1-I_d\right)\right] V_{tt}^2 \quad (35').$$

We eliminate

$$\frac{l_r - l_f}{l}\left(1-I_d\right)$$

from EQ (27') and EQ (35'), and we thus obtain

$$V_{skid}^2 = \left[1+2\left(\frac{V_{par}^2}{V_{tt}^2} - 1\right)\right] F_s V_{tt}^2 \quad (36).$$

Therefore, the effect of I_d on V_{skid}^2 is exactly twice as large as the effect of I_d on V_{tt}^2.

EFFECT OF ACTIVE REAR STEERING - The previous section noted that an outward attitude increases V_{skid}. It is well known that an inward attitude is corrected by active rear steering (ARS)[7][8]. Table 1 shows the effect of ARS on V_{skid}. If we can set the control gain to G*, V_{skid} increases to infinity. Thus, ARS can reduce the rear-wheel skid tendency due to steering inputs.

REAR SUSPENSION PROPERTIES - V_{skid} increases with increasing C^*_r and decreasing l^*_{dr} as shown in Fig. 8. C^*_r and

l^*_{dr} can be controlled not only by tire characteristics but also by rear suspension properties[4][5]. This is summarized in Table 2.

EFFECT OF FRONT AXLE ON V_{skid} - The effective cornering coefficient of the front axle (C^*_f) is quite independent of the forms of $A_f(\omega)$ and $A_r(\omega)$ as shown in Fig. 16 for the following reason. We obtained EQS (18) through (21) assuming that

$$\alpha_f = \alpha_{max\,f} \sin \omega t$$

(Appendix B).

Hence the side force is given by

$$CF_f = C^*_f \, \alpha_{max\,f} \sin \omega t.$$

Hence

$$CF_f = \mu^*_f \sin \omega t \quad (37).$$

Thus, the input of this linear calculation is equivalent to a sinusoidal maximum side force. Consequently, C^*_f is automatically eliminated(Fig.13).

RELATIONSHIP BETWEEN SLALOM DIAGRAM AND FORCE AND MOMENT METHOD - The difference between $A_f(\omega)$-$A_r(\omega)$ is equivalent to the *"yawing moment coefficient"* (C_n), which is defined by the Force and Moment Method[1], at maximum lateral acceleration. We can describe the difference with the form

$$A_f(\omega) - A_r(\omega) = \frac{l_f F_f(\omega) - l_r F_r(\omega)}{l_f l_r m g} \quad (38).$$

The yawing moment coefficient is defined by the form

$$C_N = \frac{l_f F_f(0) - l_r F_r(0)}{l m g} \quad (39).$$

Hence

$$C_N = \frac{l_f l_r}{l}\left[A_f(0) - A_r(0)\right] \quad (40).$$

holds. Therefore, the slalom diagram is an extension of the Force and Moment Method to the frequency domain.

EFFECT OF STEERING INPUTS ON REAR-WHEEL SKID - Rear-wheel skid as a function of steering inputs fundamentally differs from rear-wheel skid with a fixed steering angle. From EQ (28), we find that C^*_r strongly affects rear-wheel skid. In other words, the relation of rear-wheel skid to steering inputs depends on only α_r. However, rear-wheel skid tendency with a fixed steering angle depends on α_r - α_f [6]. Therefore, the two kinds of skid are different.

SUMMARY

The results of this paper are summarized as follows.

1) We extended the definition of front/rear-wheel skids during steady state turning to the frequency domain. Moreover, we found that vehicles, which tend to front-wheel skid in steady-state turning, tend to rear-wheel skid with sinusoidal steering inputs. This is because the rear wheels move in the lateral direction more than the front wheels do.

2) We proposed the *slalom diagram*, a "force and moment diagram [1]" that is extended to the frequency domain. Using this diagram, we can predict rear-wheel skid tendency as a function of steering inputs.

3) We proposed *skid critical speed* (V_{skid}) as an index of rear-wheel skid; below V_{skid}, rear-wheel skid is never caused by sinusoidal steering inputs. Furthermore we established a formula for this index.

4) Interpreting the formula of V_{skid}, we found that the rear-wheel skid tendency is strongly affected by the attitude angle during steady-state turning as well as the friction balance between the front and rear axles. Therefore, we can reduce the rear-wheel skid tendency by increasing wheel base and cornering coefficient as well as by active rear steering.

REFERENCES

1. Milliken, Jr., William. F, Dell'Amico and Roy S. Rice, "The Static Directional Stability and Control of the Automobile," SAE760712, 1976

2. Johnson, D. B. and J. C. Huston, "Nonlinear Lateral Stability Analysis of Road Vehicles Using Liapunov's Second Method," SAE paper 841057, International Off-Highway & Power Plant Congress & Exposition, Milwaukee, Wisconsin, 1984

3. Kusaka, K., M. Higuchi, K. Shibusawa, N. Muraoka, and M. Tukagoshi, "Transient Tire Model for Handling Analysis and Prediction," Transaction of IPC-8, 1995, Yokohama, Japan, 1995

4. Sakai, Hideki and Y.Satoh, "The Theory of Automobile Stability and Control for Transient State," Proceedings of FISITA'94 945075, Beijing, China, 1994

5. Sakai, Hideki and Y.Satoh, "A Theoretical Study of Vehicle Dynamic Behavior with Complex Cornering Stiffness," Proceedings of IPC-8 9531011, Yokohama, Japan, 1995

6. Pacejka, H. B., "Principles of Plane Motions of Automobiles", Proceedings of IUTAM Symposium on The Dynamics of Vehicles on Roads pp.33-59, Swets and Zeitlinger, Amsterdam, The Netherlands, 1976

7. Yamamoto, Masaki, H. Harada, and Y. Matsuo, "A Study on Active Controlled Chassis System for Vehicle Dynamics," 11th IAVSD Symposium, August 1989

8. Fukunaga, Y., "Improved Handling and Stability Using Four-Wheel Steering," 11th International Technical Conference on Experimental Safety Vehicles, Washington D. C., 1995

9. Ellis, J. R., Vehicle Dynamics, London Business Book, Ltd.,London, 1969

NOMENCLATURE

cic_i : coefficient of instantaneous center of plane motion to the side force of axle i

ck_{tirei} : coefficient of tire lateral stiffness

f : frequency

g : gravity
h : height of center of mass
j : unit of imaginary number
k_i : tire lateral stiffness
k_{susi} : suspension lateral stiffness
l : wheelbase
l_f : length from C.G. to front axle
l_r : length from C.G. to rear axle
l_{di} : relaxation length
l^*_{di} : linearized relaxation length
m : vehicle mass
r : yaw velocity
ck_{tirei} : coefficient of tire lateral stiffness
r_{chi} : roll center height
t : time
u : forward velocity
$v_{groundi}$: lateral velocity of tire contact area
$y_{groundi}$: lateral deformation at tire contact area
$A_i(\omega)$: side force coefficient representing maximum lateral acceleration when the limit of the other axle force is neglected
$A_{max}(\omega)$: maximum lateral acceleration at ω
C_i :effective cornering coefficient (equivalent cornering stiffness / normal load)
C^*_i : cornering coefficient of linearized effective axle characteristics
CF_i :side force coefficient (F_i/F_{zi})
F_i :side force
F_{zi} : normal force
F_s : statics factor
G^* :control gain of active rear steer
I_z : yaw moment of inertia
I_d :dynamic index of lateral motion ($I_d=I_z/(m\, l_f\, l_r)$)
R :friction balance ratio (μ^*_r / μ^*_f)
V_{skid} : critical speed of rear-wheel skid
V_{tt} :track-on-track speed
V_{par} : speed where trajectories of two particles correspond during steady-state turning
X,Y : components in Earth-fixed axes
α_i : slip angle
$\alpha_{ground\,i}$: slip angle of tire contact area
β : side slip angle at C.G.
δ : steer angle
δ_r : controlled steer angle of rear wheel
μ_i : effective friction ratio of axle i
μ^*_i : friction ratio of linearized effective axle characteristics
ω : angular frequency
ω_p : frequency where skid function is maximized

Suffix
i :i=f or r
f : front
r : rear
L : left
R : right

APPENDIX - A
TIRE CHARACTERISTICS

Comparing the two curves in Fig. 3, we can gain the following insights into vehicle behavior.

1) The slopes of the two curves determine the stability factor[9]. Namely, if the rear slope is larger than the front slope at a given side force coefficient, i.e. lateral acceleration, then the vehicle tends to understeer at this lateral acceleration[6]. Conversely, if the rear slope is smaller than the front slope, the vehicle tends to oversteer.

2) The two μ_i values determine the skid tendency during steady-state turning[1]. Namely, if $\mu_r < \mu_f$, the vehicle tends to rear-wheel skid. Conversely, if $\mu_r > \mu_f$, the vehicle tends to front-wheel skid.

Table A1 illustrates various vehicle behaviors when the steering wheels are fixed.

From Fig. 1, we can see that Case 1 in table A1 represents almost all current passenger vehicles. Thus, this paper focuses on Case1 rear-wheel skid to steering inputs.

APPENDIX B
LINEARIZATION FOR EFFECTIVE AXLE CHARACTERISTICS

We linearize the tire model, which is effective axle characteristics[6], using the harmonic balance method. Figure B1 outlines this procedure.

Namely, we assume the slip angles are

$$\alpha_i = \alpha_{max\,i} \sin \omega t = \frac{3}{2}\frac{\mu_i}{C_i} \sin \omega t \qquad \text{(B1)}.$$

We will discuss the validity of this assumption in Appendix E.

Further, we solve EQS (4) through (8). Since EQ (4) is a nonlinear function, the side force CF_i contains higher harmonics. Hence

$$CF_i = e_1 \sin \omega t + e_2 \sin 2\,\omega t + e_3 \sin 3\,\omega t + \cdots \qquad \text{(B2)}$$

where e_i is a complex number. Since e_1 is the response to this input frequency, we focus on only e_1. Hence,

$$CF_i \equiv e_1 \sin \omega t \qquad \text{(B3)}.$$

Substituting EQS (4) through (8) and (B1) into (B2), we obtain

$$e_1 = \frac{9}{8}\frac{1}{1+\left(\frac{l^*_{di}}{u}j\omega\right)^2}\left(1-\frac{l^*_{di}}{u}j\omega\right) \qquad \text{(B4)}$$

where

$$l^*_{di} = \frac{\sqrt{3}}{4}\frac{C_i}{ck_{tirei}} \quad \text{(B5)}$$

where l^*_{di} is relaxation length. It should be noted that not only C^*_i but also l^*_{di} include the properties of the tire, steering system, and suspension(Table 2).

Rearranging EQ (B4), we obtain the frequency response function of CF_i to α_i in the form

$$\frac{CF_i}{\alpha_i} = \frac{1}{1+\frac{l^*_{di}}{u}j\omega}C^*_i \quad \text{(B6)}$$

where

$$C^*_i = \frac{3}{4}C_i \quad \text{(B7).}$$

Since

$$\alpha_i = \frac{3}{2}\frac{\mu_i}{C_i}\sin\omega t \quad \text{(B1),}$$

thus

$$0 \le |CF_i| \le \mu^*_i \quad \text{(B8)}$$

where

$$\mu^*_i = \frac{9}{8}\mu_i \quad \text{(B9).}$$

Then we approximate EQ (B6) using the complex cornering stiffness concept[4][5] to simplify the motion equations. Namely, complex cornering stiffness can approximate $A_f(\omega)$ into a quadratic of ω with high accuracy, while the exact form of $A_f(\omega)$ is quartic.

Assume

$$\left|\frac{l^*_{di}}{u}j\omega\right| << 1 \quad \text{(B10)}$$

and

$$\left|1+\frac{l^*_{di}}{u}j\omega\right| \approx 1 \quad \text{(B11).}$$

We can thus make EQ (B6) a Maclaurin's expansion for l^*_{di} and obtain

$$\frac{CF_i}{\alpha_i} = \left(1-\frac{l^*_{di}}{u}j\omega\right)C^*_i \quad \text{(B12).}$$

Using the above functions, we can estimate the slalom diagram with a linear calculation.

APPENDIX C
CALCULATION OF $A_f(\omega)$

Using EQS (13) through (21) and (B10) through (B12), we can obtain vehicle response in the form[4][5]

$$\frac{r(\omega)}{\delta(\omega)} = G^*_r \frac{\left(T^*_{r0}+T^*_r\right)j\omega + 1}{\frac{1}{{\omega^*_n}^2}(j\omega)^2 + 2\frac{\zeta^*}{\omega^*_n}j\omega + 1} \quad \text{(C1)}$$

$$\frac{\beta(\omega)}{\delta(\omega)} = G^*_\beta \frac{\left(T^*_{\beta 0}+T^*_r\right)j\omega + 1}{\frac{1}{{\omega^*_n}^2}(j\omega)^2 + 2\frac{\zeta^*}{\omega^*_n}j\omega + 1} \quad \text{(C2)}$$

$${\omega^*_n}^2 = D^*_\omega {\omega^*_{n0}}^2 \quad \text{(C3)}$$

$$\zeta^*\omega^*_n = D^*_\zeta D^*_\omega \zeta^*_0\ \omega^*_{n0} \quad \text{(C4)}$$

$$D^*_\omega = \frac{1}{1+\frac{1}{I_d}\left\{\frac{g}{u}\left[cic_f C^*_f T^*_f + cic_r C^*_r T^*_r\right] + C^*_f C^*_r \frac{g^2 T^*_f T^*_r}{u^2}\right\}} \quad \text{(C5)}$$

$$D^*_\zeta = 1 + \frac{-\left(1-\frac{l_r}{l_f}\frac{C^*_f\, l\, g}{u^2}\right)C^*_f\frac{u\,T^*_f}{l} + \left(1+\frac{l_f}{l_r}\frac{C^*_r\, l\, g}{u^2}\right)C^*_r\frac{u\,T^*_r}{l}}{cic_f\, C^*_f + cic_r\, C^*_r} \quad \text{(C6)}$$

$${\omega^*_{n0}}^2 = \frac{C^*_f C^*_r}{I_d}\left(1 + A^* u^2\right)\frac{g^2}{u^2} \quad \text{(C7)}$$

$$\zeta^*_0\ \omega^*_{n0} = \frac{1}{2}\left[cic_f\, C^*_f + cic_r C^*_r\right]\frac{g}{u} \quad \text{(C8)}$$

$$T^*_{r0} = \frac{u}{C^*_r\, g} \quad \text{(C9)}$$

$$T^*_{\beta 0} = \frac{I_d}{\frac{1}{1+A^* u^2}\left(1-\frac{u^2}{C^*_r\, g\, l_r}\right)}\frac{u}{C^*_r\, g} \quad \text{(C10)}$$

$$cic_f = \frac{l_f + I_d\ l_r}{l}$$

$$cic_r = \frac{I_d\ l_f + l_r}{l}$$

$$G^*_\beta = \frac{1}{1 + A^* u^2}\left(1 - \frac{u^2}{C^*_r\ g\ l_r}\right)\frac{l_r}{l} \quad \text{(C11)}$$

$$G^*_r = \frac{1}{1 + A^* u^2}\ \frac{u}{l} \quad \text{(C12)}$$

$$A^* = \left(\frac{1}{C^*_f} - \frac{1}{C^*_r}\right)\frac{1}{l\ g} \quad \text{(C13)}$$

$$T^*_i = -\frac{l^*_i}{u} \quad \text{(C14).}$$

Using EQS (C1) through (C13), we can obtain frequency response function of $[u(\beta j\omega + r)/g]$ to CF_f and CF_r as

$$\frac{\left[\dfrac{u\left(\beta j\omega + r\right)}{g}\right]}{CF_f(\omega)} = \frac{\dfrac{u}{g}\left(\dfrac{\beta}{\delta}j\omega + \dfrac{r}{\delta}\right)}{-C^*_f\left(1 - \dfrac{l^*_{df}}{u}j\omega\right)\left(\dfrac{\beta}{\delta} + \dfrac{l_f\dfrac{r}{\delta}}{u} - \dfrac{\delta}{\delta}\right)}$$

$$= \frac{\dfrac{(j\omega)^2}{\omega_s^2} + \dfrac{2\ \zeta_s\ j\omega}{\omega_s} + 1}{\dfrac{(j\omega)^2}{\omega_{skid}^2} + \dfrac{2\ \zeta_{skid}\ j\omega}{\omega_{skid}} + 1} \quad \text{(C15)}$$

and

$$\frac{\left[\dfrac{u\left(\beta j\omega + r\right)}{g}\right]}{CF_r(\omega)} = \frac{\dfrac{u}{g}\left(\dfrac{\beta}{\delta}j\omega + \dfrac{r}{\delta}\right)}{-C^*_r\left(1 - \dfrac{l^*_{dr}}{u}j\omega\right)\left(\dfrac{\beta}{\delta} - \dfrac{l_r\dfrac{r}{\delta}}{u}\right)}$$

$$= \frac{\dfrac{(j\omega)^2}{\omega_s^2} + \dfrac{2\ \zeta_s\ j\omega}{\omega_s} + 1}{T_{skid}\ j\omega + 1} \quad \text{(C16).}$$

Recalling the definition of $A_i(\omega)$,

$$A_i(\omega) = \frac{\mu^*_i}{\left|CF_i(\omega)\right|}\left|A_{max}(\omega)\right| \quad \text{(C17).}$$

Hence

$$A_i(\omega) = \left|\frac{A_{max}(\omega)}{CF_i(\omega)}\right|\mu^*_i \quad \text{(C18).}$$

Performing the linear calculation, we obtain

$$\frac{A_{max}(\omega)}{CF_i(\omega)} = \frac{u\left(\beta\ \omega j + r\right)}{CF_i(\omega)}$$

$$= \frac{\left[\dfrac{u\left(\beta\ \omega j + r\right)}{g}\right]}{CF_i(\omega)}g \quad \text{(C19).}$$

Therefore

$$A_i(\omega) = \left|\frac{\left[\dfrac{u\left(\beta\ \omega j + r\right)}{g}\right]}{CF_i(\omega)}\right|\mu^*_i\ g \quad \text{(C20).}$$

Substituting EQS (C15) and (C16) into EQ (C20), we can obtain EQS (22) and (23) respectively.

APPENDIX D
CALCULATION OF V_{skid}

First, we define the *skid function* as

$$S(\omega) = \frac{CF_f(\omega)}{CF_r(\omega)} \quad \text{(D1)}$$

$$= R\ \frac{A_f(\omega)}{A_r(\omega)}$$

$$=\left|\frac{T_{\text{skid}}\ j\omega+1}{\frac{(j\omega)^2}{\omega_{\text{skid}}{}^2}+\frac{2\ \zeta_{\text{skid}}\ j\omega}{\omega_{\text{skid}}}+1}\right| \quad \text{(D2)},$$

where

$$R=\frac{\mu^*_{\mathrm{r}}}{\mu^*_{\mathrm{f}}} \quad \text{(D3)}.$$

The skid function is shown in Fig. D1. Secondly, we define ω_p as the solution of

$$\left.\frac{\mathrm{d}\,S(\omega)}{\mathrm{d}\omega}\right|_{\omega=\omega_p}=0 \quad \text{(D4)}.$$

At ω_p, the value of the skid function is maximized. Further, at V_{skid}, the maximum value of the skid function must equal R. Thus the equation

$$S(\omega_p)=R \quad \text{(D5)}$$

holds at V_{skid}. Hence we substitute $u=V_{skid}$ into EQ (D5) and then solve this equation for V_{skid} on the asumtion that $T_{skid}=0$, yielding

$$V_{\text{skid}}{}^2=\frac{C^*_{\mathrm{r}}}{C_{\mathrm{r}}}\frac{R^2}{I_{\mathrm{d}}}\left(\Delta_1+\Delta_2\sqrt{1-\frac{1}{R^2}}\right)V_{\mathrm{tt}}{}^2 \quad \text{(D6)}$$

where

$$\Delta_1=\left(cic_{\mathrm{r}}-\frac{l^*_{\mathrm{dr}}}{l}\right)^2-\left[(I_{\mathrm{d}}-1)^2\left(\frac{l_{\mathrm{r}}}{l}\right)^2-2cic_{\mathrm{r}}\frac{l^*_{\mathrm{dr}}}{l}\right]R^2 \quad \text{(D7)}$$

$$\Delta_2=\left(cic_{\mathrm{r}}-\frac{l^*_{\mathrm{dr}}}{l}\right)^2-(I_{\mathrm{d}}-1)^2\left(\frac{l_{\mathrm{r}}}{l}\right)^2R^2 \quad \text{(D8)}$$

$$V_{\mathrm{tt}}{}^2=\frac{C^*_{\mathrm{r}}}{C_{\mathrm{r}}}\ \frac{l}{2}\ C_{\mathrm{r}}\ g \quad \text{(D9)}.$$

Assuming $I_d \doteqdot 1$ and $l^*_{dr}/l \doteqdot 0$, we can make EQ (D6) a Taylor expansion for $I_d=1$ or $l^*_{dr}=0$ and then obtain EQS (24) through (27).

APPENDIX E
ACCURACY OF LINEAR CALCULATION

We assumed

$$\alpha_i=\alpha_{\max i}\ \sin\omega t$$

in EQ (B1). However, there is no assurance that this slip angle promotes A_r, which is the maximum value of the rear side force at each frequency. Considering this assumption, this section discusses the accuracy of this linearized calculation.

Suppose A_r is generated at

$$\alpha_i=\alpha_{\max i}\ (1\text{-}\Delta)\sin\omega t \quad \text{(E1)}$$

where $0 \leqq \Delta \ll 1$. Further, assume $l^*_{dr}=0$, $l_f=l_r=l/2$, and $I_d=1$ for convenience. Replacing EQ (B1) with EQ (E1), we obtain

$$C^*_i(\Delta)=\frac{3}{4}\ C_i\ \left(1+\frac{3}{2}\Delta+\Delta^2\right) \quad \text{(E2)}$$

$$\mu^*_i(\Delta)=\frac{9}{8}\ \mu_i\ \left(1-\frac{1}{3}\Delta-\Delta^2\right) \quad \text{(E3)}$$

using Maclaurin's expansions for Δ.
Thus C^*_r increases and μ^*_r decreases with increasing Δ.
Next, we assume

$$\omega^2=C^*_{\mathrm{r}}\frac{g}{l} \quad \text{(E4)}$$

and

$$u^2=C^*_{\mathrm{r}}\ l\ g \quad \text{(E5)}$$

as typical conditions to promote rear-wheel skid, i.e., this frequency is the undamped natural frequency of the skid function and this velocity is the square root of twice the track-on-track speed. Hence, this frequency is most likely to produce rear-wheel skid and this speed is even higher than V_{skid}. Substituting EQS (E2) through (E5) into (23), we make EQ (23) a Maclaurin's expansion for Δ and obtain

$$\left|A_{\mathrm{r}}(\Delta)\right|=\sqrt{\frac{\mu^*_{\mathrm{r}}}{2}}\sqrt{1-\frac{26}{9}\Delta^2} \quad \text{(E6)}.$$

Thus, $\Delta=0$ maximizes $A_r(\omega)$, and the assumption of EQ (B1) is valid in rear-wheel skid frequencies. This shows the high accuracy of the rear-wheel skid estimation as long as EQ (4) is used for the effective axle characteristics.

Next, we will discuss the accuracy of front-wheel skid estimation. At a front-wheel skid frequency, $A_{max}(\omega)$ is determined by the limit of the front side force, and so the rear maximum side force does not increase. Hence, the rear slip angle is lower than α_{maxr}, and the actual C^*_r will be larger than $(3/4)C_r$, which is used in this linear calculation, as shown in EQ (E2). The estimation values of $A_f(\omega)$ and $A_r(\omega)$ using $(3/4)C_r$ as C^*_r will be lower than the exact values at the front-wheel skid frequency, which does not assure accurate estimation of front-wheel skid.

Therefore, this linear calculation is valid for rear-wheel skid prediction, which includes V_{skid} estimation.

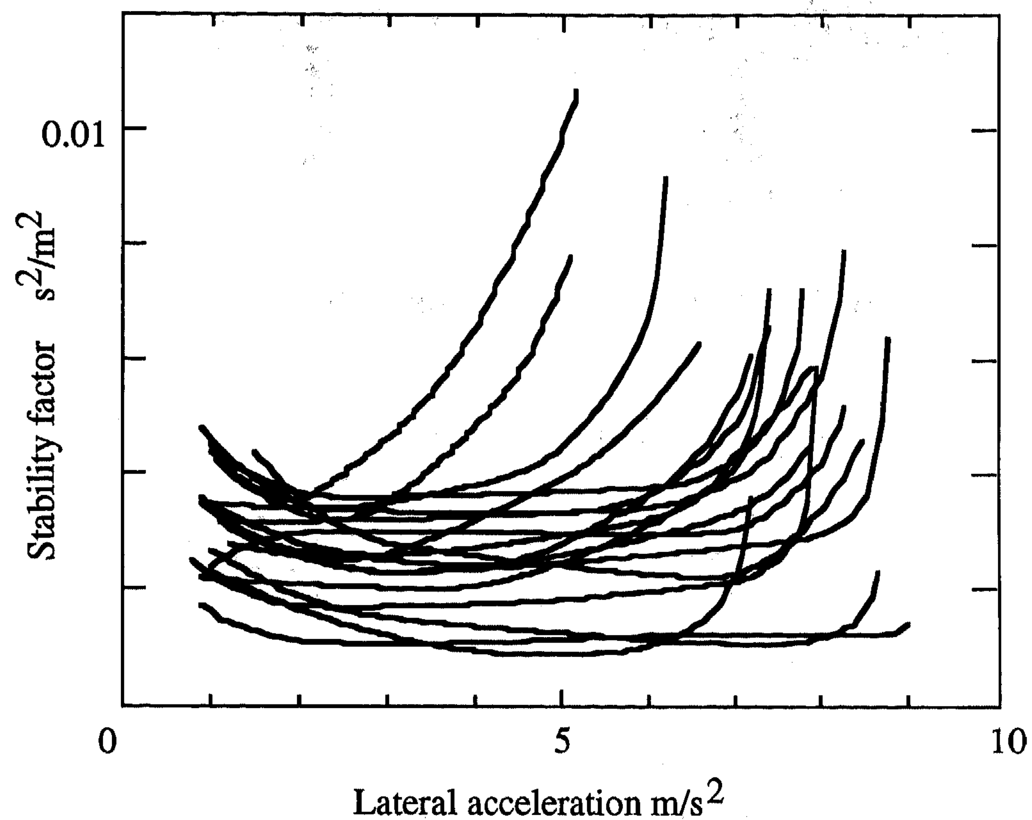
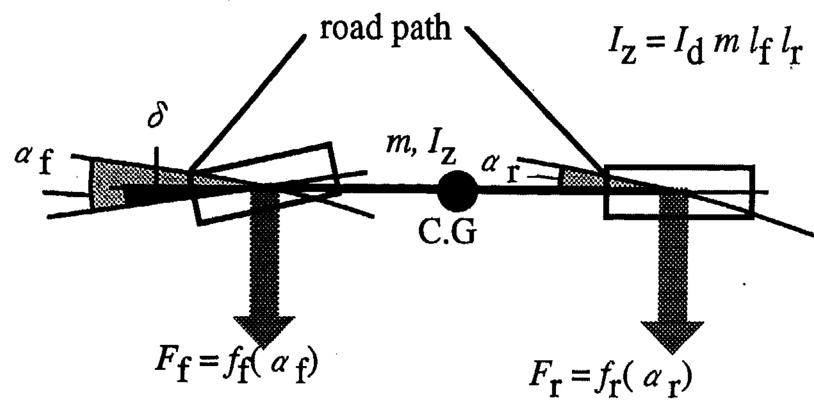
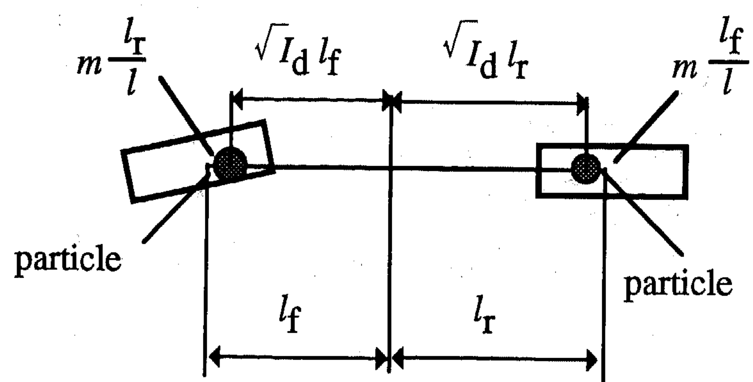


Figure 1 Measuring result of stability factor. These tests were conducted in constant radius (30 R) turning from approximately 1 m/s^2 up to the maximum lateral acceleration. From this figure, we can see that almost all current passenger cars tend to understeer at any lateral acceleration. Therefore these cars tend to front-wheel skid.
Reproduced with permission from Motor Fan Road Test, Motor Fan 1994.2 ~ 1995.12, Sanei-Syobo Co., Tokyo, 1994~1995.



(A) Description in system of rigid body



(B) Description in system of particles

Figure 2 Equivalent Vehicle Model. This paper also describe an equivalent vehicle model in a system of particles to focus on the difference between the two particles in motion.

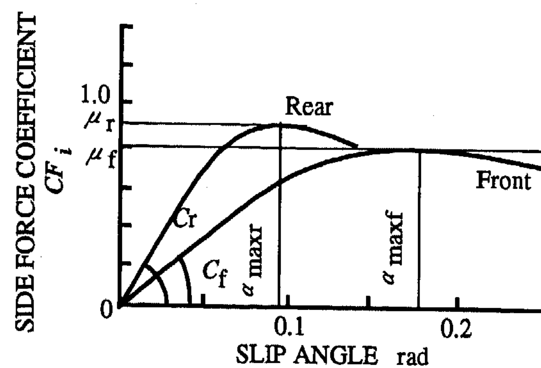
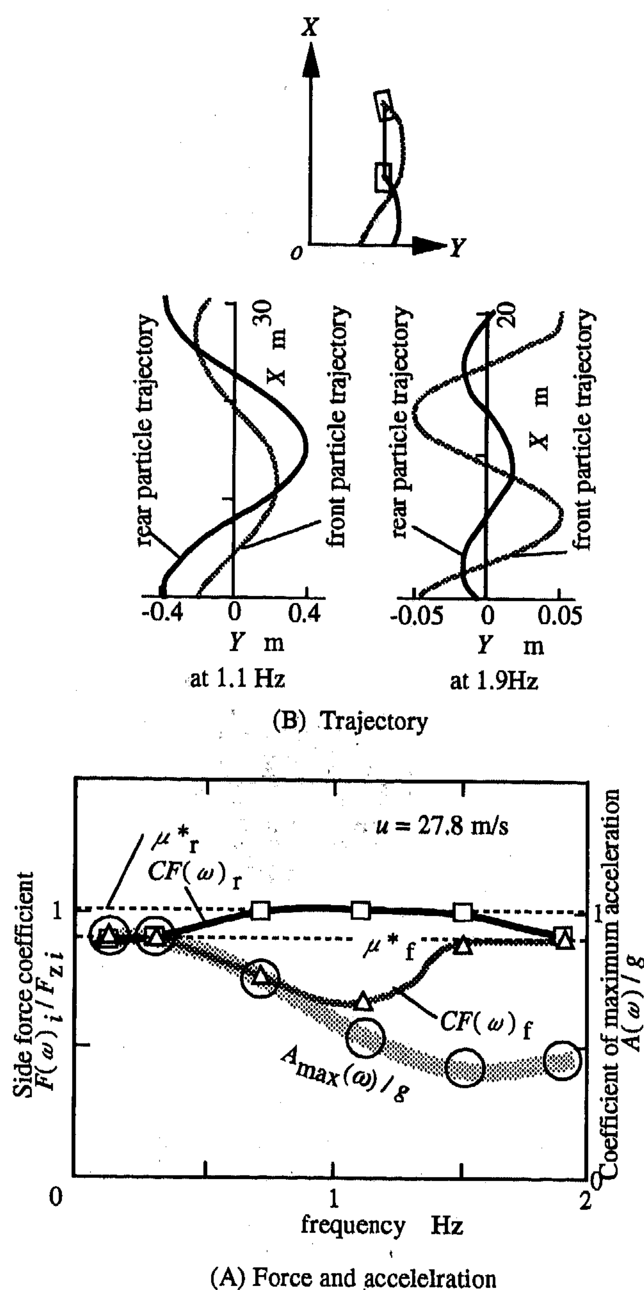


Figure 3: Effective axle characteristics in steady state. This figure shows the tire model in steady-state turning. Focusing on oridinary vehicles, which tend to understeer at any acceleration as shown in Fig. 1, we assume $C_r > C_f$ and $\mu_r > \mu_f$. In addition, we assume a time lag in the transient state based on Ref[3].



(A) Force and accelelration

Figure 4 Simulation result of response to sinusoidal steering input. At 0.1, 0.3 and 1.9 Hz, since the front axle produces its maximum force, the vehicle tends to front-wheel skid. While at 0.7 and 1.1Hz, the rear axle yields its maximum force, so the vehicle tends to rear-wheel skid. The reason is illustrated by the particle trajectories. Namely, rear-wheel skids occur when the rear particle moves in the lateral direction more than the front particle.

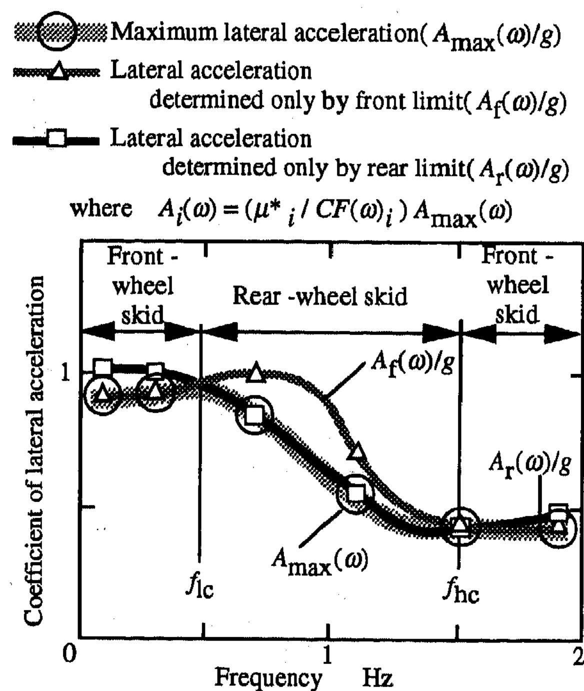


Figure 5 Slalom Diagram. We transform figure 4(A) to simply illustrate skid tendency and maximum lateral acceleration.

Linearized calculation

Coefficient of extended maximum lateral acceleration as determined by front limit $(A_f(\omega)/g)$ only

Coefficient of extended maximum lateral acceleration as determined by rear limit $(A_r(\omega)/g)$ only

Simulation

△ Coefficient of extended maximum lateral acceleration as determined by front limit $(A_f(\omega)/g)$ only

□ Coefficient of extended maximum lateral acceleration as determined by rear limit $(A_r(\omega)/g)$ only

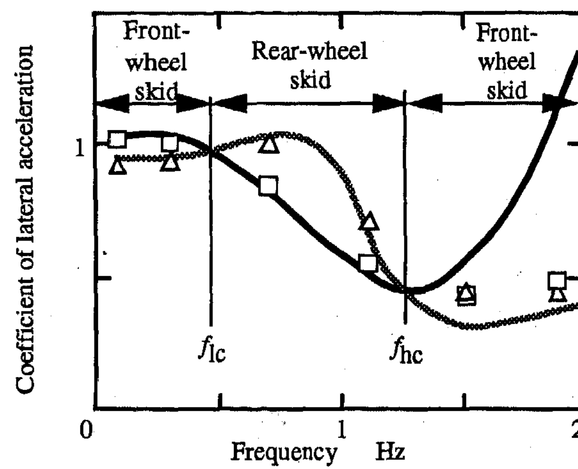


Figure 6 Accuracy of linearized calculation. The numerical simulation takes much time to detect the optimum steer angle for promoting maximum lateral acceleration. Since this linearized calculation does not require much time and agrees with the numerical simulation, we can use the linearized simulation hereafter in this paper.

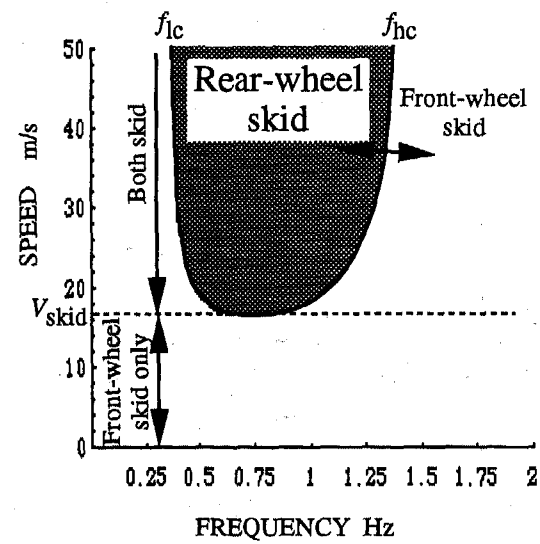
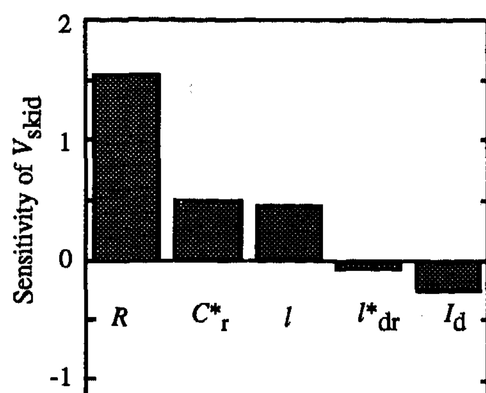
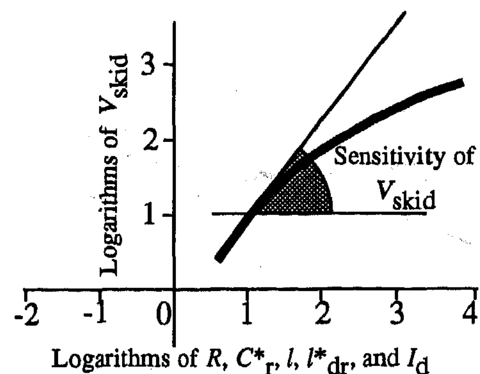


Figure 7 Effect of speed on rear-wheel skid tendency. V_{skid} denotes the critical speed in the rear-wheel skid zone. V_{skid} is adopted as an index of rear skid tendency.

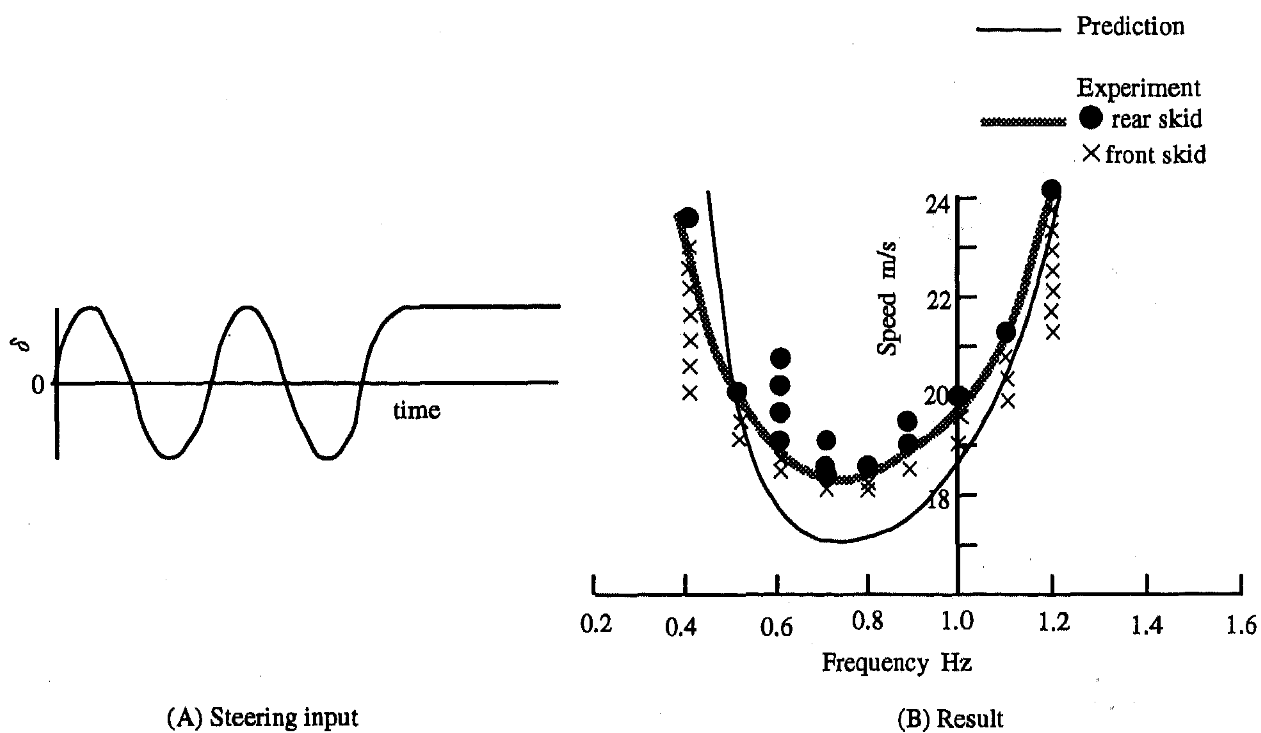




(A) Sensitivity of parameters

(B) Explanation of parameters

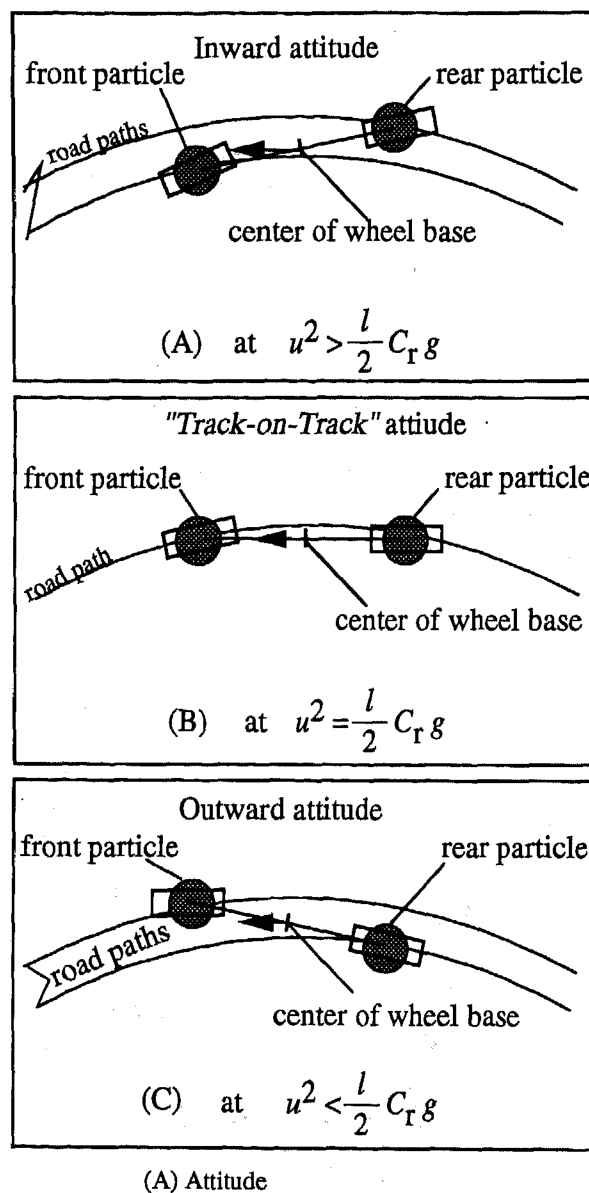
Figure 8. Sensitivity of V_{skid}. This figure shows that an increase of the ratio of friction balance (R) enhances V_{skid}, as is well known. Hence, we deduce that this sensitivity result is correct. Further, an increse of C_r and l and decrease of l^*_{dr} and I_d also enhance V_{skid}.



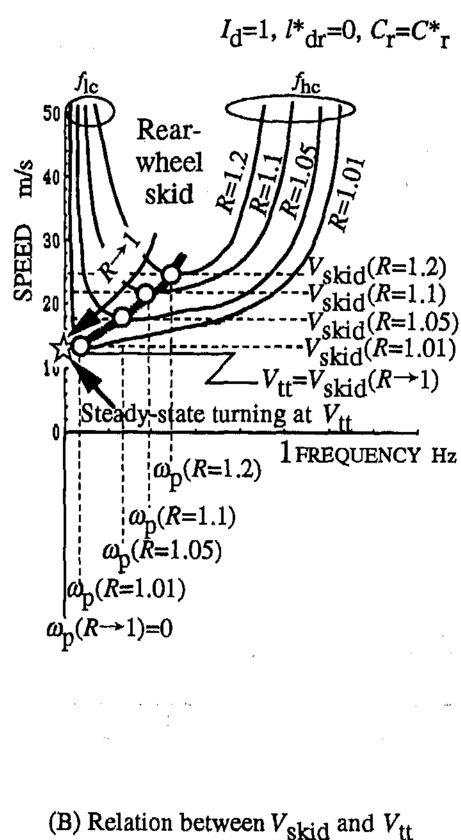
(A) Steering input

(B) Result

Figure 9 Verification experiment. These results generally agree with those expected.



(A) Attitude



(B) Relation between V_{skid} and V_{tt}

Figure 11 Track-on-Track speed. At *track-on-track speed,* the front and rear tracks correspond during steady state turning. This speed is V_{skid} under the assumption of I_d=1, l^*_{dr}=0, C_r=C^*_r and $R\to 1$. As shown in Fig. 11(B), we can examine rear-wheel skid with the steady-state turning at V_{tt}. Over this speed, the rear particle trajectory goes beyond the front as well as vehicle behavior in rear-wheel skid as shown in figure 5. In addition, the particles lie just on the respective axles because I_d=1.

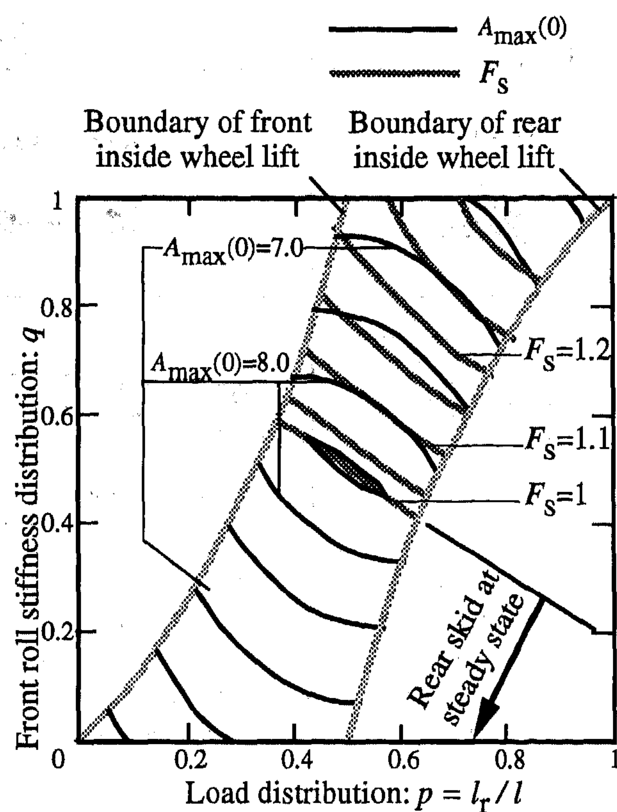


Figure 10 effect of roll moment distribution on friction ratio and maximum lateral acceleration. Increase of F_s can reduce rear skid tendency. Further, an increase of p and q is able to increase F_s. Hence front-end heavy and front heavy distribution of the roll moment can decrease rear skid tendency. This fact is well known empirically. However, this also reduces maximum lateral acceleration during steady-state turning.

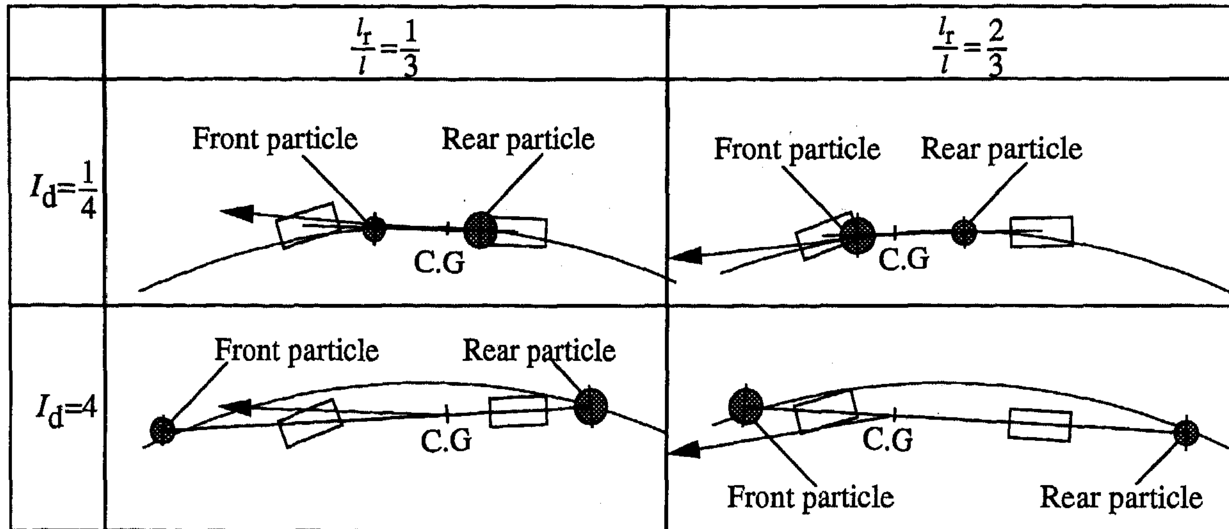


Figure 12 Affect of yawing moment(I_d) on particle trajectories. This figure shows the attitude angles at the speed (V_{par}) where both trajectories of the front and rear particles correspond. V_{par} is the expansion of track-on-track speed (V_{tt}) into genaral cases, where I_d dose not equal 1. Hence we can replace V_{tt} by V_{par} in the discussion of V_{skid}. Suppose a rear-end-heavy vehicle. A larger I_d makes inward attitude at V_{par}. In other words, this implies $V_{par}>V_{tt}$ when $I_d>1$. Therefore a larger I_d increases V_{skid}. Conversely, for front-end-heavy vehicles, a smaller I_d enhances V_{skid}.

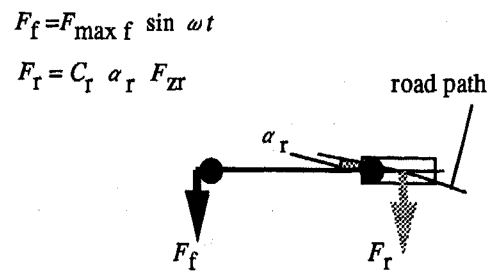
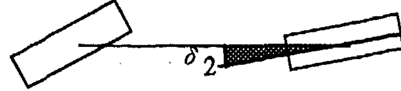


Figure 13: Vehicle model for front force input;. This model equals the equivalent vehicle model provided we choose a steer angle that promotes maximum front side force.

Table 1 Effect of Active Rear Steer (ARS) on V_{skid}. Since ARS affects attitude angle, ARS enhances V_{skid}. In particular, control for β=0 and dβ / dt =0 makes V_{skid} infinity.



Control law	V_{skid} when $I_d=1$	G for $V_{skid}=\infty$
$\delta_2 = G\delta$	$\dfrac{1-\dfrac{G}{R^2}}{\left(1-\dfrac{C_r^*}{C_f^*}G\right)^2(1-G)}\left[R^2-\dfrac{C_r^*}{C_f^*}G+\sqrt{\left(1-\dfrac{1}{R^2}\right)\left(\dfrac{C_r^*}{C_f^*}\dfrac{G}{R}\right)^2-R^2}\right]V_{tt}^2$	$\dfrac{C_1}{C_2}$
$\delta_2 = G\dfrac{r}{u}$	$\left[\left(1+\dfrac{G}{l}\right)^2-\dfrac{1}{R^2}\dfrac{G}{l}\right]F_s V_{tt}^2$	impossible
$\delta_2 = G u(\beta j\omega + r)$	$\dfrac{R^2}{\left(1-C_r^* gG\right)^2}\left[1-\dfrac{1}{R^2}\left(\dfrac{l_r}{l}+\dfrac{l_f}{l}R^2 C_r^* G g\right)\right]V_{tt}^2$	$\dfrac{1}{C_2 g}$
$\delta_2 = G_\delta \delta + G_r r$ or $\delta_2 = G\dfrac{T_n j\omega+1}{T_d j\omega+1}\delta$ for $\dot{\beta}=0$ and $\beta=0$	∞	G for $\dot{\beta}=0$ and $\beta=0$ (Referencies[6][7])

Table 2 Affect of suspension charactristics on complex cornering stiffness

Suspension characteristics	C_r	l_{dr}
Compliance of lateral force	$\dfrac{CC}{1-CC\,F_{zr}\dfrac{d\,\delta_r}{d\,F_r}}$	0
Roll center height	CC	$-CC\dfrac{m\,h}{RS}rch_r$
Suspension lateral stiffness	CC	$\dfrac{CC\,F_{zr}}{k_{sus\,r}}$

Table A1 Categories of relationships between front and rear tire characteristics. Caption. As shown in Fig.1, almost all current passenger cars tend to understeer. at any lateral acceleration. Hence this paper assesses of case 1. Therefore we must assume $\mu_r > \mu_f$ and $C_r > C_f$. (This table was produced by relating the two concepts, the force and moment method [1] and the handling diagram [6], presented by the author.)

CASE	Steering characteristics around center	Steering characteristics around maximum acceleration	Skid at maximum acceleration turning	Stability factor	Effective axle characteristics
1	U.S.	U.S.	Front-wheel skid	Lateral acceleration; maximum acceleration 1; Front-wheel skid; 0; -0.005; 0.005; -1; O.S Stability factor U.S	Side force coefficient F_i / F_{zi}; maximum acceleration 1; Front-wheel skid; 0; 0.1; front; rear; -1; α_i
2	O.S.	U.S.	Front-wheel skid	Lateral acceleration; maximum acceleration 1; Front-wheel skid; -0.005; 0; 0.005; -1; O.S Stability factor U.S	Side force coefficient F_i / F_{zi}; maximum acceleration 1; Front-wheel skid; 0; 0.1; front; rear; α_i
3	U.S.	O.S.	Rear-wheel skid	Lateral acceleration; Rear-wheel skid; 1; maximum acceleration; -0.005; 0; 0.005; -1; O.S Stability factor U.S	Side force coefficient F_i / F_{zi}; maximum acceleration 1; Rear-wheel skid; 0; 0.1; rear; front; α_i
4	O.S.	O.S.	Rear-wheel skid	Lateral acceleration; Rear-wheel skid; 1; maximum acceleration; -0.005; 0; 0.005; -1; O.S Stability factor U.S	Side force coefficient F_i / F_{zi}; maximum acceleration 1; Rear-wheel skid; 0; 0.1; rear; -1; front; α_i

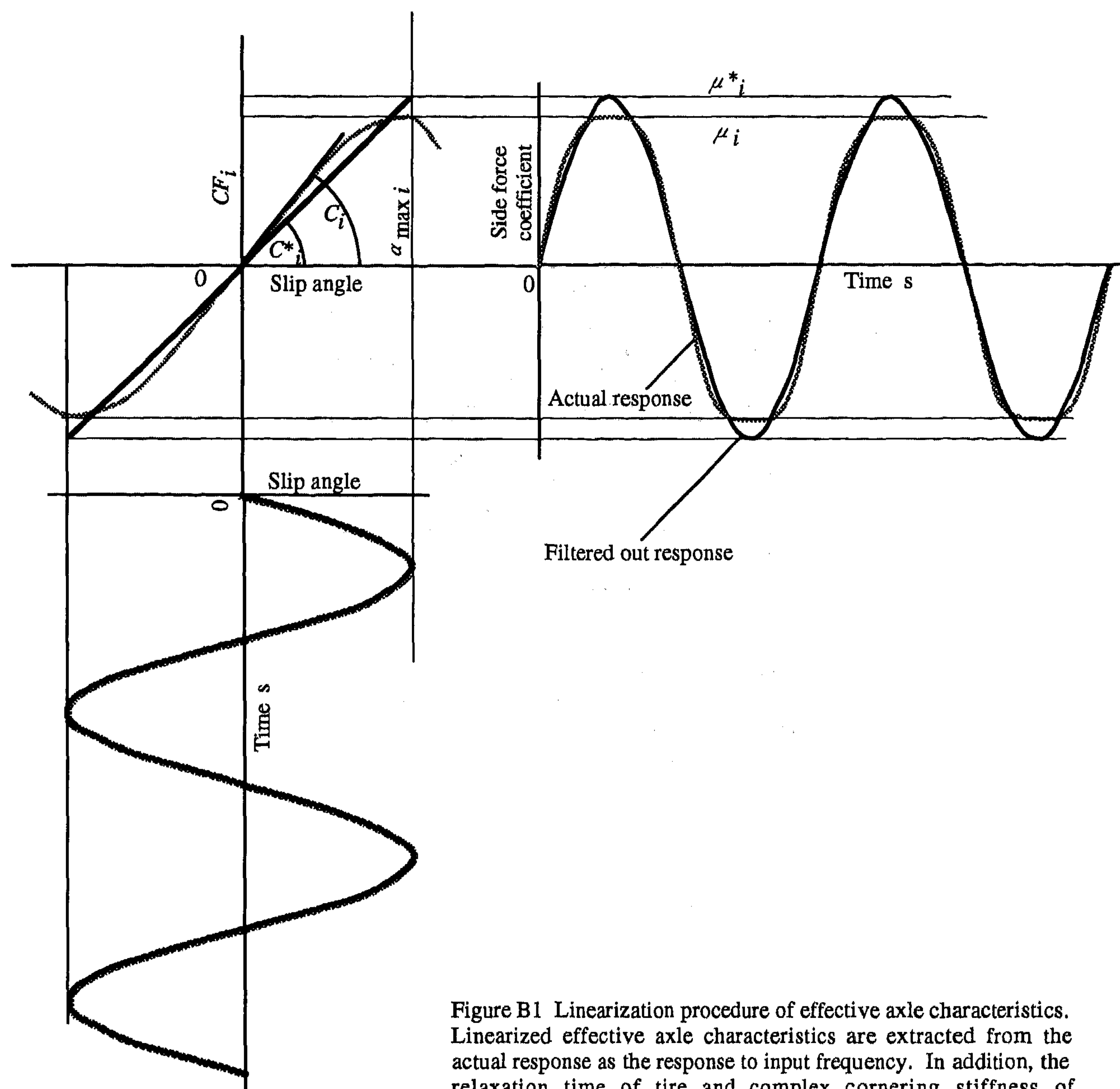


Figure B1 Linearization procedure of effective axle characteristics. Linearized effective axle characteristics are extracted from the actual response as the response to input frequency. In addition, the relaxation time of tire and complex cornering stiffness of suspension have been omitted from this figure for brevity of explanation.

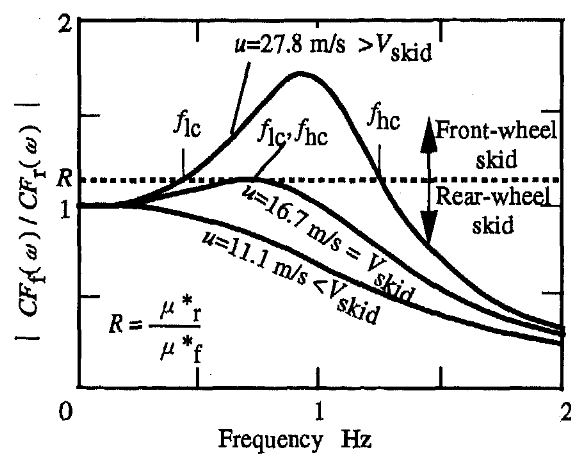


Figure D1 Skid function; Caption. Skid function is defined to establish the V_{skid} formula. At V_{skid}, the peak value of the skid function corresponds to friction balance ratio (R).

970379

Analysis of Energy Consumption for Various Power Assisted Steering Systems

Geoffrey P. Dyer
A.E. Bishop & Associates Pty. Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The energy consumption of conventional open-centre hydraulic power steering was mapped to pinpoint areas of potential for improvement. Using this data, equivalent maps were generated for existing, and potential, competing systems. These maps were used to identify likely future trends in this area.

The results immediately illustrate the gross inefficiency of the conventional system due to high parasitic losses and indicate not only that systems that operate purely "on-demand", such as electric power steering (EPS), are likely to be the optimum solution at least as far as energy consumption is concerned but also that interim solutions such as electro-hydraulic power steering (EHPS) can make significant contributions.

INTRODUCTION

Power assisted steering is penetrating ever further across (and down) the automobile market. Even cars with engine capacities as low as 700 to 800 cc are sold with power steering now, a situation which was unheard of less than 10 years ago! With the rising concern for the level of energy consumption by automobiles and also the consequent level of production of greenhouse gases, it has become necessary/desirable to analyse the energy consumption of the conventional open-centre hydraulic power steering system in order to establish its significance with respect to the total vehicle consumption and identify the most effective areas to target for improvement. With several existing and potential alternatives to this system available, it is also useful to compare these systems by using steering history data to simulate their respective energy maps.

There are a multitude of parameters that affect the energy consumption of open-centre power steering systems. They include:

a) Oil supply flow rate (to steering gear)
b) Losses in the flow control valve
c) Flow losses in the hydraulic lines
d) Flow losses in the power steering valve
e) Friction and viscous losses in the pump
f) Weight on the front wheels of the vehicle
g) Front end geometry of the vehicle
h) Boost characteristic of the power steering valve
i) Pump speed vs. time history
j) Assistance pressure vs. time history

Parameters **a** to **e** are controlled by the system design and can vary greatly from vehicle to vehicle. They have, potentially, a large effect on overall energy consumption. Parameters **f** to **h** are governed by the vehicle specification and their effect can only be gauged by comparison testing of vehicles of disparate weight and configuration. Parameters **i** and **j** are the result of a combination of driver behaviour and demands of road conditions and affect the energy picture through their effect on parameters **a** to **e**. From the recording of the steering history of a vehicle, we can calculate the variations in these parameters and establish their influence on the end result.

In order to generate representative maps of energy consumption, two vehicles were selected for energy mapping: Vehicle A, a large rear wheel drive family sedan, of Australian manufacture and with a mass of 1500 kg and Vehicle B, a small/medium sized front wheel drive family sedan, of European manufacture and of 1100 kg mass. These were chosen so as to establish the possible differences in system requirements caused by their widely conflicting specifications. Overall average fuel consumption for these cars was found to be 16 l/100 km and 10 l/100 km, respectively, when used mainly for suburban commuting with the occasional country trip. In order to obtain a complete picture of steering requirements, data was collected for a wide variety of driving conditions (not standardised drive cycles as in fuel consumption testing) and also for a variety of drivers. Track testing was also carried out to establish the steering performance envelope for extreme manoeuvres such as emergency lane change and dry parking.

Using the steering histories from these two vehicles, maps of energy consumption were generated and analysed for its significance relative to the total energy consumption of the vehicle and the relative contributions of the different components of the system. Also generated were maps of energy consumption for alternative systems and these were compared with the open-centre system.

OVERVIEW OF SYSTEMS

Currently, there are four main categories of system which would be considered contenders for power assistance of automotive steering:

- Conventional open-centre
- Open-centre with constant speed drive for the pump
- Closed centre
- Electric power steering

The fairest means of comparing these systems is to reference the amount of energy supplied by the base power source (in most cases the engine crankshaft) and, therefore, the efficiency of the means used to drive the system must be taken into account.

CONVENTIONAL OPEN CENTRE (OC) - A pump of fixed displacement is driven, mechanically and at a fixed ratio, from the engine crankshaft. Pump flow is, therefore, proportional to engine speed. In order to maintain constant control valve characteristics, a flow control valve (FCV) is used to maintain constant flow through the valve by bypassing the excess flow around the steering gear, usually internally in the pump.

The main contributors to the low efficiency of the conventional system are the parasitic losses associated with the continuous supply of oil to the power steering valve as well as the non constant speed of the pump drive (being proportional to engine speed). Not only does the pump have to work continuously against the back pressure of the system (due to FCV, power steering valve and feed and return lines) but it can also be pumping oil at up to 5 to 6 times the rated flow of the FCV against the same pressure. This latter situation is a consequence of the pump being engine driven at a fixed ratio.

In this case, the efficiency of the mechanical (usually belt) drive to the pump must be applied.

OPEN-CENTRE CONSTANT SPEED DRIVE FOR PUMP (OCCS) - The pump for an open-centre system is driven, mechanically or electrically, at a constant speed. This negates the need for a flow control valve in the pump as there is no excess flow to bypass the steering gear.

This system still has parasitic losses due to the continuous flow of oil through the power steering valve but the system back pressure would be reduced by the absence of a FCV and, also, there would be no excess flow for the pump to supply.

There are three likely scenarios for driving this system:

An Electric Motor (EHPS) - this involves the efficiency of the belt drive to the alternator as well as that of the alternator/motor/battery package.

A Constant Speed Accessory Drive - a CVT driving all accessories at constant speed. The efficiency of both the transmission and the belt drive would apply.

Engine Maintained at (or Near) Constant Speed - a CVT main transmission maintains the engine at relatively constant speeds with varying throttle openings. The pump would then run at (or near) constant speed with only the efficiency of the belt drive to take into account.

CLOSED CENTRE (CC) - An accumulator is used to store oil at high pressure for use "on demand", ie. flow only occurs when the steering is actually displaced. Recharge of the accumulator is achieved by a pump, intermittently driven electrically or mechanically, as required by the stored pressure.

Whilst this is an "on demand" system having no parasitic losses, all flow required by the steering is supplied by dropping full accumulator pressure, via the steering ram, to atmosphere. Hence, energy is lost when the steering pressure requirement does not equate to full accumulator pressure.

The accumulator is charged by a pump which is driven intermittently, based on the demands of the stored pressure. Again, this pump can be driven mechanically or electrically with similar considerations for efficiency to those stated above.

ELECTRIC POWER STEERING (EPS) - The steering is directly powered by electric motor, usually via a gearbox, and controlled electronically. A torque transducer replaces the sensing function of the valve in the hydraulic system.

This system is close to a true "on demand" system, only supplying energy to the steering as and when it is required and only supplying as much as is required. The efficiency of the system is then only affected by the efficiency of the energy conversion mechanisms: the power supply to the motor; the motor and the gearbox.

DATA ACQUISITION

Analysis of the energy consumption required the following in-vehicle data to be recorded on the cars:

- Vehicle speed.
- Pump driving torque.
- Pump speed.
- Differential boost pressure.
- Rack travel.
- Oil temperature.
- Fuel flow.

All of these parameters could be recorded directly, via transducers, or from signals already available on the vehicle. However, because of the difficulty of running a torque transducer on the input-shaft of the pump in the aggressive environment of the engine bay, it was decided to calibrate the pump for input torque versus output pressure (just down stream of the FCV) and pump speed. This was carried out on a test rig, in the laboratory, and an algorithm developed empirically to allow the calculation of pump torque from these two variables. One benefit of recording the time/pressure/pump speed histories is that the energy loss results can be recalculated for varying system back pressures.

Hence, the actual parameters, measured and recorded in the vehicle, were:

- Vehicle speed.
- Engine speed.
- Pump speed.
- Oil temperature.
- Differential pressure.
- Pump output pressure.
- Rack position.
- Fuel flow.

The data acquisition system sampled all the variables at 5000Hz. At each sample point, the software would take the first 50 data points (0.01 s), average them, calculate all the required parameters and store the results. These sample points were collected every 0.09 to 0.2 s, depending on the time to calculate the stored information.

DATA PROCESSING

From the measured inputs, the following parameters directly relating to the performance of the system on the vehicle, were calculated:

- Energy absorbed by the pump.
- Work done on the rack.
- Efficiency of the system.
- Distance traveled by the vehicle.
- Energy absorbed by the pump/km traveled.
- Work done on the rack/km traveled.
- Average vehicle speed.
- Maximum OC power absorbed by the pump during test.

Additional parameters were calculated, relating to the equivalent consumption for alternative systems:

- Energy used if the system was CC.
- Volume of fluid used if the system was CC.
- Volume of fluid used/km if system was CC.
- Energy used if the system was OCCS.
- Maximum rack power during the test.
- Maximum OC power absorbed by the pump during the test.
- Maximum OCCS power during the test.

These calculations did not take into account the efficiencies of the component parts of each system.

At the end of each test, a file was generated to store the summary of the parameters for the whole test.

VEHICLE TESTING

A wide variety of routes were used for the mapping exercise for Vehicles A an B so as to give a reasonably balanced picture for vehicle use in a metropolitan environment. Types of route used included:

- Local suburban (trip distances less than 8 km.).
- Local suburban (trip distances between 8 and 25 km.).
- Local city suburbs to outer suburbs (trip distances approx. 40 km).
- Local suburbs to inner city in rush hour.
- Low speed manoeuvring.
- Freeway.
- Country roads (trip distances up to 500 km).
- Around country towns.
- Twisty country roads.

As well as energy mapping, testing was carried out to define the steering performance requirement envelope for each vehicle. This was carried out on a test track and included emergency avoidance manoeuvres (lane change) and "dry parking" (winding the steering from lock to lock with the service brake applied).

Lastly, Vehicle B was laboratory tested for fuel consumption on a rolling road dynamometer, to Australian standard highway and urban drive cycles, both with and without the power steering pump connected. This was deemed necessary when road testing for changes in fuel consumption proved to be inconclusive.

CALIBRATION OF PUMP INPUT TORQUE

Figure 1 shows the relationship between input torque and outlet pressure for the pump on Vehicle A and is reasonably representative of that for Vehicle B. It shows a basically linear relation between torque and pressure and, thus, is largely independent of pump speed. The only case when this does not hold true is that of the "on- centre" condition, ie. when the steering has little or no torque applied to it. As illustrated in Figure 2, although the on-centre outlet pressure stays reasonably constant, the torque rises linearly with speed. This is attributed to viscous losses in the pump and bearings as well as a slight rise in internal supply pressure to the FCV at higher pump speeds. It is applied as a speed correction factor to the

empirical equation for torque vs pressure and is only significant when output pressures are low.

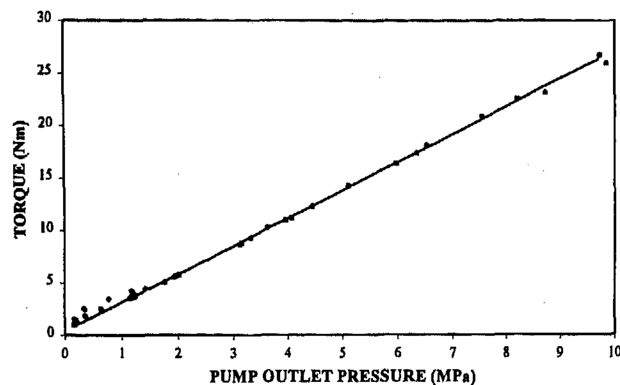


Fig. 1 Pump Drive Torque vs Outlet Pressure.

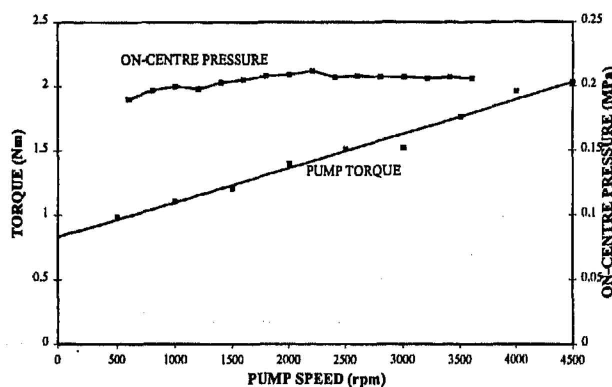


Fig. 2 Torque and Pump Pressure vs Pump Speed: On Centre Pressure

Hence, the equation for the pump input torque was developed as:

T = Pc.Pv + Sc.RPM + C
Where:
T = Pump Input Torque (Nm)
Pv = Pump Pressure (MPa)
RPM = Pump Speed (rpm)
Pc = Pressure Constant (0.02559 for vehicle A)
Sc = Pump Speed Constant (0.0003 for vehicle A)
C = Torque Constant (0.4134 for vehicle A)

The above algorithm was generated with rig testing carried out at 50 - 60°C and, as such, would only hold true for those temperatures. However, changes in temperature (Vehicle A would run with oil at approximately 90°C and Vehicle B 55°C) will have two effects: both viscous losses in pipes and viscous losses in the pump bearings will increase with lower temperatures. Whilst the losses in the piping can be considered significant, they are catered for in the above algorithm for pump torque. The variation in internal losses in the pump were considered to be of less significance and were not considered in this analysis.

DISCUSSION OF RESULTS FOR OC SYSTEM

Figure 3 shows the collective results for the efficiency of the OC system (work done by rack ÷ input energy to pump x 100%) for each test run with respect to average speed. Apart from the obviously low overall result, the interesting trend to notice is that the efficiency increases quite markedly at lower average speeds. Maximum efficiency was approximately 6% and the minimum efficiency less than 0.01%. This phenomenon can be explained by inspection of the steering usage relative to average journey speed.

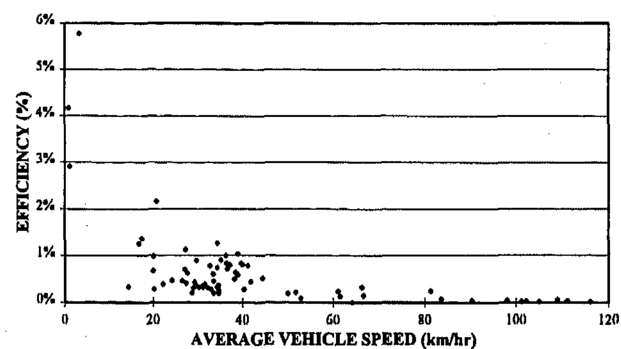


Fig. 3 Efficiency of OC System vs. Average Speed.

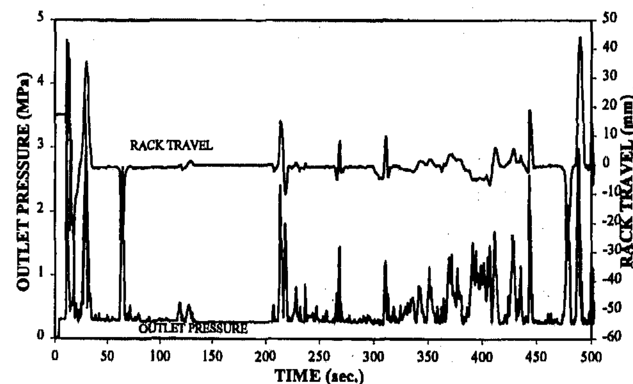


Fig. 4 Rack Travel and Pump Pressure History for a Typical, Suburban Route.

Figure 4 is the steering history for a typical, short, suburban drive in the Sydney area. The distance represented is 5 km and overall average speed was 34.8 km/h. Of immediate interest is how little work the steering is actually required to do. Apart from the gross steering movements at the start and end of the run (exit car park and entry into driveway) and during turns at intersections, the steering spends a very large percentage of the time at (or near) on-centre even though this route includes very little straight road. Also noticeable is the

significant "waiting time", from 134 secs to 206 secs, where the vehicle was stationary at traffic lights. This amounts to 14% of the total journey time and is relatively low when compared to average figures for waiting time in town driving.

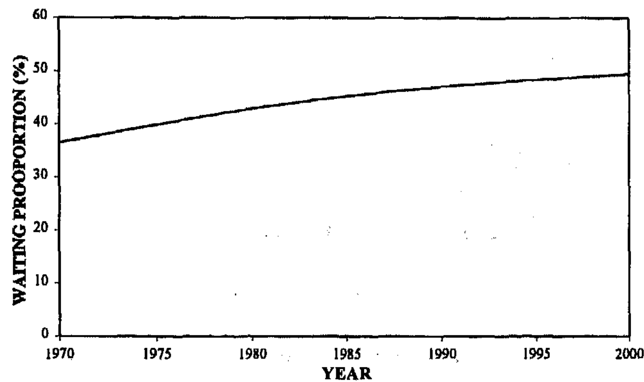


Fig. 5 Development of Proportion of Waiting Times for Urban and City Traffic[1]

Figure 5 represents the development of waiting times, ie. that proportion of time which the vehicle remains stationary, with no steering input and engine at idle, for urban and city traffic from 1970 and projected forward to the year 2000 (reference 1). From this data, waiting times of nearly 50% would be expected in the urban environment The pump on vehicle A, at idle (pump speed 760rpm), absorbs 90W, although this can easily be doubled by resting a hand on the steering wheel during the waiting period. This means that the waiting time is a significant factor in the energy consumption of the steering in town driving and, since this is a large component of the usage of the average vehicle, it is also significant in the overall picture. But this factor would not account for the increase in efficiency at lower speeds. Indeed, it would tend to point toward the opposite effect.

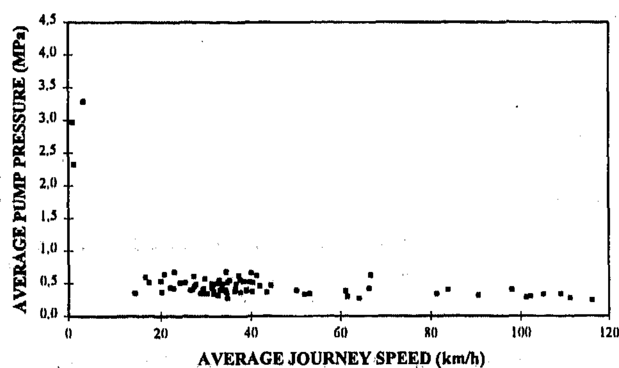


Fig. 6 Average Pump Pressure vs. Average Journey Speed: Vehicle A.

To explain this, we have to look at the trends in steering history for different average journey speeds.

In Figure 6, it will be noted that the average pump outlet pressure falls very quickly with increasing speed from just over 3 MPa at 3 km/h (car park manoeuvring) to of the order of 0.5 MPa at 20 km/h. From here, the pressure drifts down toward the on-centre pressure of 0.2 MPa because, as the speed increases, the steering is required to carry out less work. Hence, the parasitic losses of the on-centre regime become more dominant.

Figure 7, for Vehicle B, illustrates very much the same trends except that the pressures are generally 0.2 - 0.3 MPa higher throughout the range of speeds. This is a direct result of the piping arrangements for this vehicle which had an on-centre pressure of approximately 0.5 MPa as opposed to 0.2 MPa for Vehicle A with similar flow rates. The fact that on-centre losses are dominant means that the back pressure, generated by the piping as well as the power steering valve housing and the power steering valve, are features to focus on with regard to optimising the system for energy consumption.

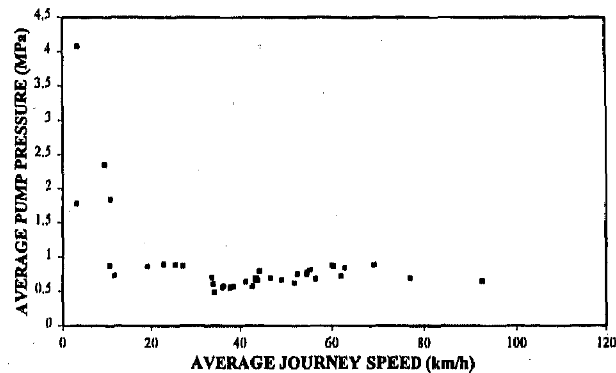


Fig. 7 Average Pump Pressure vs. Average Journey Speed: Vehicle B.

Confirmation of the importance of the on-centre losses can also be found by looking at the percentile bands

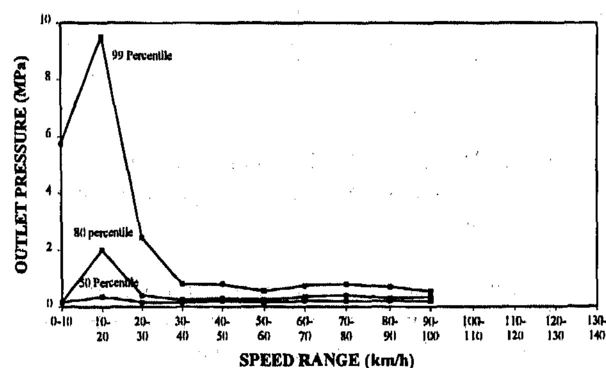


Fig. 8 Percentile Bands of Pump Pressure vs. Speed Range: Inner City in Peak Hour.

of pump pressure with respect to speed. Figures 8, 9 and 10 detail the percentile bands of pump pressure for three widely differing routes and types of driving.

One can immediately identify the different steering demands of these routes: city driving with high waiting times and low speed manoeuvring; free moving suburban driving with many intersections and roundabouts and fast, country driving with long, flowing curves which demand reasonable steering pressure but little in the way of rack travel. Given these differences, it is interesting to note that, for the vast majority of the speed ranges, the pressure is at (or near) the on-centre pressure for 80% of the time on all three routes.

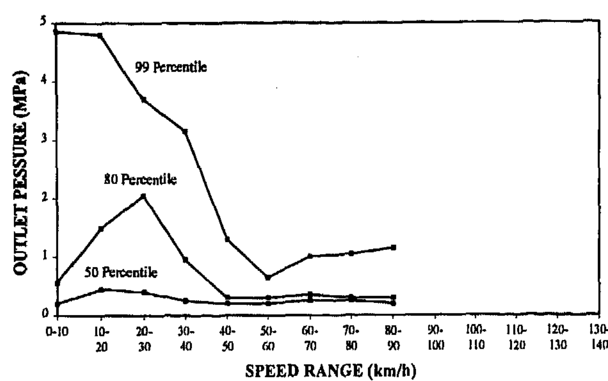


Fig. 9 Percentile Bands of Pump Pressure vs.Speed Range: Suburban Back Streets in Light Traffic.

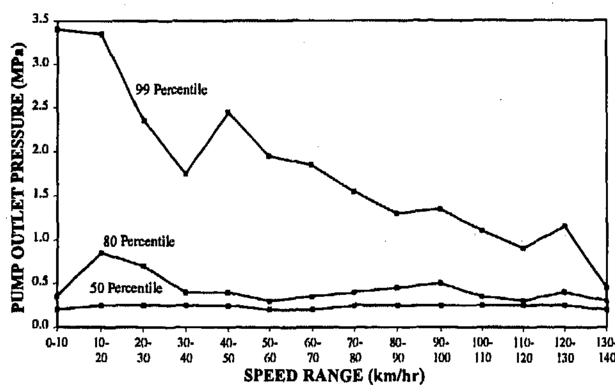


Fig. 10 Percentile Bands of Pump Pressure vs.Speed Range: Fast, Country, Single Carriageway Roads.

If we now look at the percentile bands for rack deviation from centre for the same three routes (Figures 11, 12 and 13), it is evident that, in all cases above approximately 40 km/h, the rack deviates more than 5mm extremely infrequently. This means that, at higher speeds, the amount of work done by the rack is low. At lower speeds, the work rate is dependant on the driving conditions: city driving requiring more rack work at low speeds (10 - 20 km/h); suburban driving requiring more rack work at moderate speeds (10 - 40 km/h) due to the number of intersections and roundabouts to be negotiated.

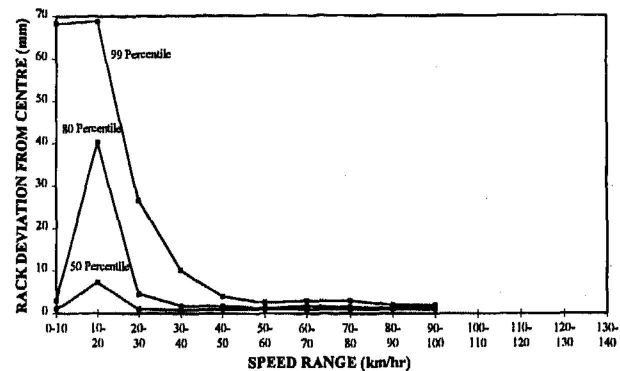


Fig. 11 Percentile Bands for Rack Deviation From Centre: Inner City in Peak Hour.

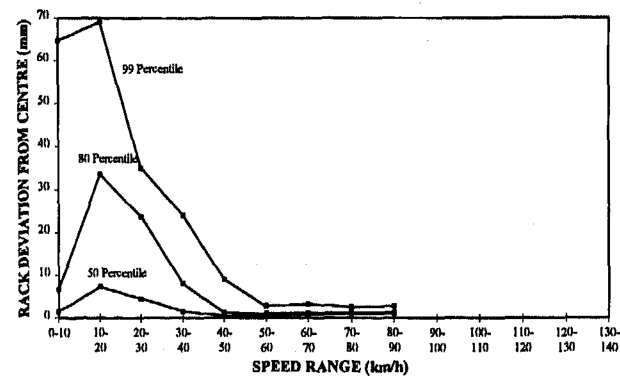


Fig. 12 Percentile Bands for Rack Deviation From Centre: Suburban Back Streets in Light Traffic.

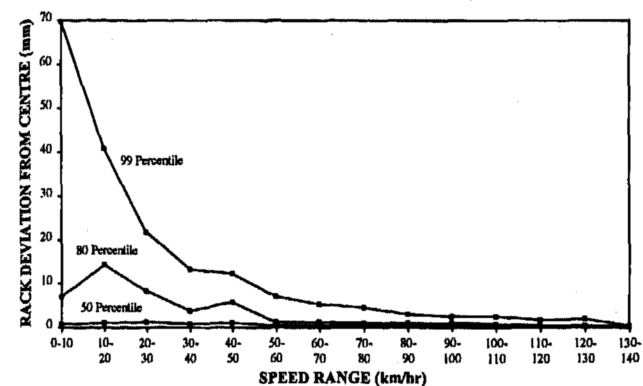


Fig. 13 Percentile Bands of Rack Deviation From Centre: Fast, Country, Single Carriageway Roads

THE OVERALL PICTURE - The average pump speed varies approximately linearly with average vehicle speed (see Figure 14) except at low speeds when gear changes have an effect. On Vehicle A at 100 km/h, the pump speed is approximately 2500 rpm and the input power is 400W when on-centre (see Figure 15). This translates into an energy consumption of 14.4 kJ/km with zero output force at the rack.

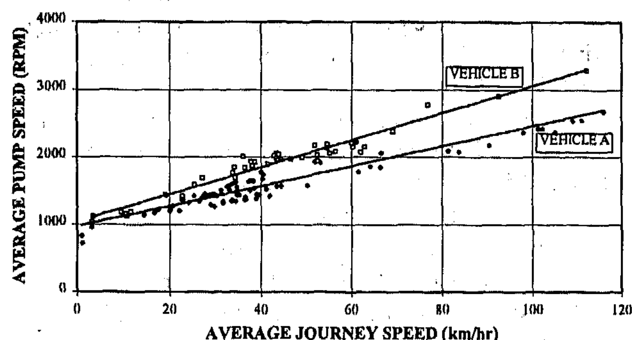


Fig. 14 Pump Speed vs. Average Journey Speed

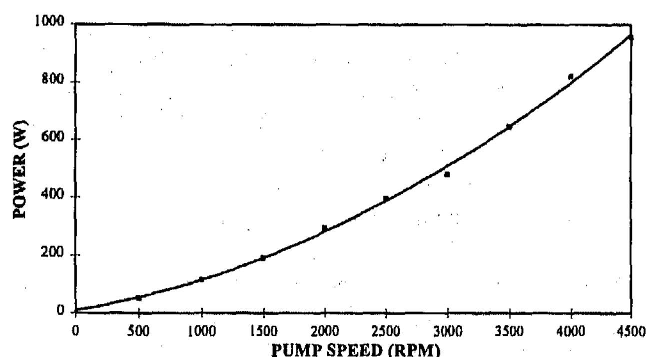


Fig. 15 Input Power vs. Pump Speed: On Centre

Referring back to Figures 6 and 7, the average pump pressure at 120 km/h is very close to that for the vehicle stationary and at idle rising slowly as the speed reduces to 20 km/h. Below 20 km/h, the pressure can rise dramatically due to the amount of gross steering movement which can be made at these speeds. As an approximation, the average pump pressure varies as follows:

Ave. Pump Pressure a (Ave. Vehicle Speed)$^{-1/2}$

Figure 16 shows the energy consumption per kilometre for Vehicle A as a function of average vehicle speed. There is a strong correlation with speed, the energy consumption/km varying approximately with the inverse of the average speed.

Vehicle B displays the same relationship between speed and consumption (see Figure 17) except that the consumption at any given speed is higher than that for vehicle A. This can directly be attributed to the higher on-centre back pressure that this vehicle displays.

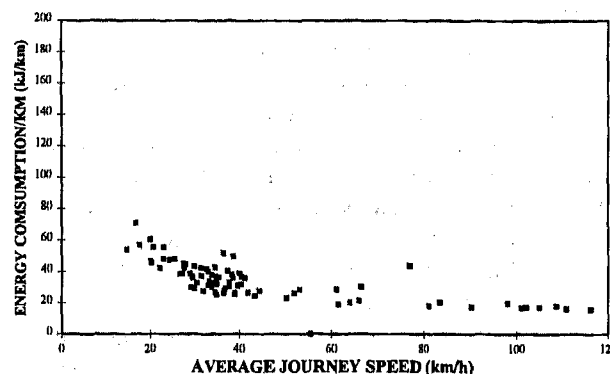


Fig. 16 Energy Consumption / km vs Average Journey Speed: Vehicle A.

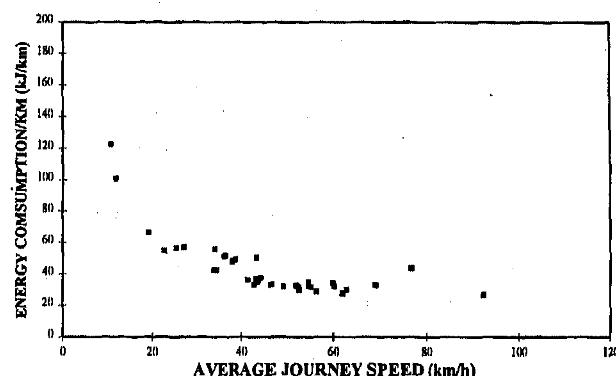
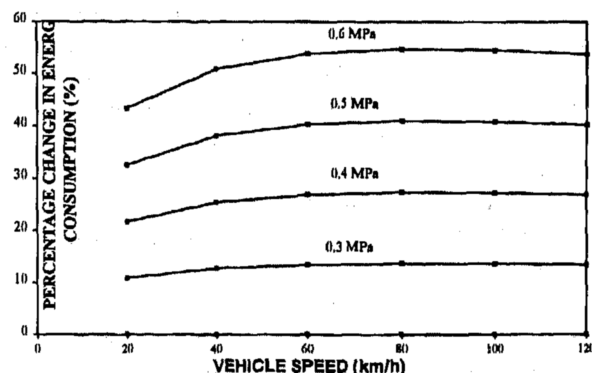


Figure 17 Energy Consumption/km vs Average Journey Speed: Vehicle B.

Having established the importance of the on-centre regime, the effect of the on-centre back pressure should be investigated.



.Fig. 18 Change in On-Centre Energy Consumption vs. Back Pressure.

Figure 18 shows the change in energy consumption, to be expected by changing this parameter, calculated using the empirical formula for pump torque developed earlier.

The level of change is reasonably constant except that it falls away at speeds below 20 km/h, due to the increased demand for rack work increasing the input power requirement whilst the absolute on-centre losses actually decrease at these speeds. Vehicle B runs with a back pressure of between 0.4 and 0.5 MPa. From Figure 18, one would expect an increase in consumption, over that of Vehicle A, of approximately 30% at 100 km/h which is confirmed by experimental data.

How does this affect the fuel consumption of the host vehicle? Overall energy consumption for Vehicle A's steering was 23.28kJ/km and that for Vehicle B was 31.12.kJ/km and overall average speeds were 55.6km/h and 47.1km/h, respectively. The efficiency of the drive to the pump from the crankshaft would be approximately 0.9, so crankshaft energy output would be 25.87 and 34.58 kJ/km, respectively. Modern automobile engines, at the duty levels represented here, are approximately 25% efficient in terms of output at the crankshaft relative to fuel input. Hence, the steering requirement for energy from fuel would be 103.48kJ/km and 138.32kJ/km or 0.32 l/100 km and 0.43l/100 km (using a calorific value of 32.25MJ/l for petrol).

Laboratory testing for fuel consumption on Vehicle B gave raw results (in l/100km) of:

	HWY	URBAN
Pump Connected	7.66	10.38
Pump Disconnected	7.52	10.09
Difference	0.14	0.29

Obviously, on the rolling road, no steering inputs could be made. This was accounted for by using the difference in results for energy consumption/km and on-centre losses at the equivalent average pump speeds to generate a correction factor. The modified differences in consumption then became: 0.23l/100 km (HWY) and 0.46 l/100 km (URBAN) which, given the mix of urban and highway usage represented by the mapping exercise, is in close agreement with the above results for the test vehicles.

Track testing of the two vehicles produced the following results for maximum system performance requirement:

Vehicle:	A	B
Maximum Rack Speed (m/s)	0.15	0.15
Maximum Rack Accel. (m/s^2)	0.90	0.92
Maximum Rack Force (kN)	6.0	7.1
Maximum Rack Power (W)	800	700

It is interesting to note how similar the requirements are, given the widely differing specifications of the two vehicles. The energy maps are also very similar except for the extra back pressure for Vehicle B which is the sole reason that its steering system absorbs more energy than that of Vehicle A.

The OC system, then, is grossly inefficient, having an overall efficiency of the order of 0.03%. This is due to the large parasitic losses associated with the constant supply of oil to the power steering and the pump being driven at varying speeds above that required to supply the rated flow of oil.

The question to ask now is: "How does this fit in to the picture of the automobile as a whole?". Currently, overall driveline efficiencies (fuel in to tractive effort out) are in the region of 8 - 10%, fuel consumption for a medium sized sedan is 8 - 10l/100 km and fuel is relatively inexpensive, especially with respect to the overall costs of running a vehicle. In the life of the vehicle (160,000km), 690l of fuel will be used to power the steering which, currently in Australia, costs approximately A$500 or A$50 - 80 per year. Of growing concern is the 1660kg of CO_2 produced by burning this fuel. The steering currently accounts for approximately 4% of the fuel consumption of the vehicle but efforts are being made to improve the overall efficiency of the vehicle (the US Government PNGV target is 2.94l/100 km for a medium sized sedan) which will increase the significance of the steering contribution. Indeed, with the apparently constant absolute energy requirement, regardless of vehicle size, the application of power steering to smaller vehicles means that the significance of this energy consumption increases quite markedly and this, along with packaging problems, may have already caused manufacturers to look to alternative systems.

Very little can be done to improve the efficiency of the OC system apart from paying attention to the back pressure generated by the system as a whole.

ALTERNATIVE SYSTEMS

OCCS SYSTEM - Mechanical Drive - A transmission to convert crankshaft output to constant speed would have an efficiency of approximately 0.9 and, with the belt drive (also 0.9), would have an overall drive efficiency of 0.81.

Losses would be reduced by reducing on-centre losses (by lowering pump internal back pressure - no FCV) and no extra losses due to the speed of the pump at higher road speeds.

Electric Motor Drive (EHPS) - With the use of an electric motor, the efficiency of the whole electricity supply system comes into play. Assuming the accessory drive efficiency is 0.9, the alternator average efficiency is approximately 0.4 and the electric motor efficiency is approximately 0.6, the total system efficiency is 0.22.

(Alternator and motor efficiencies are estimates based on information from reference (2)) This would seriously disadvantage this system if it is designed to supply the same flow rate as the OC system which is set to account for the highest rack speeds. These speeds are rarely seen in general driving and, since the speed of the electric motor can be controlled electronically, the EHPS system is normally designed around a lower flow rate (typically ¼ of the conventional figure) and the motor accelerated when rack speed demands extra flow. This has compounding benefits: the system would be expected to run at marginally lower back pressures and, for the vast majority of the time, the pump will require only 25% of the power at equivalent steering pressures. This is not now a true constant speed system.

CC SYSTEM - For this system, the amount of energy input required is calculated by multiplying the volume of oil required to move the rack by the accumulator pressure and applying the efficiencies of the component parts of the system. The parts, common to both electrically and mechanically driven systems are: accumulator (0.9)and pump (0.9). Energy is saved in this system because it is a true "on demand" system. In theory, when the rack is stationary and only required to supply a force, the system will not be using any energy due to there being no active flow from the accumulator. This situation doesn't take into account the problem of leakage of fluid from the accumulator, through the valving.

In calculating the CC energy consumption for vehicle A, the mean accumulator pressure was assumed to be 13MPa which would require a rack piston area of $682mm^2$ to achieve the 8.86kN maximum rack load at the standard relief pressure (10MPa and $886mm^2$ piston area). The highest fluid volume requirement occurs during maneouvring in a narrow roadway. The volume required for a five point turn (carried out in 30sec.) was 415ml. This means that the accumulator should be of the order of 500ml capacity and the pump should have a delivery rate of approximately 1.0l/min. Any leakage will have the effect of adding a parasitic loss to the system since the stored energy, in the form of pressure, is continually being vented to atmosphere. Since the leakage rate would be reasonably constant, the additional burden on energy consumption (kJ/km) would be inversely proportional to the vehicle average speed.

Leakage rates will be determined by the design of the control valving and are difficult to estimate. For the purposes of our comparison, here, a leakage rate of 0.2l/min was selected as representing a severe leakage problem and this represents a parasitic loss off 43.3W or 53.5W at the input to the pump.

Mechanical Drive - Here the efficiency of the drive to the pump (possibly integrated into the engine camshaft) would be approximately 0.9 and total efficiency of 0.73 would be applied.

Electric Drive - The efficiency of the electric motor system, as before, would be 0.22 and the total efficiency would be 0.175.

EPS SYSTEM - Whilst this is an "on demand" system, an electric motor draws a current even if it is only supplying a torque at zero rotational speed. Hence, the efficiency of the system is more difficult to calculate because the efficiency of the motor is dependant on its speed.

Motor torque requirement is calculated from the rack force and the mechanical advantage (from maximum rack speed (0.15m/s) and maximum motor speed (assumed 4000 rpm giving a ratio of 26,667 rads/m). Motor torque has a relationship with current which is independant of speed (reference 2) and this is used to calculate the instantaneous current requirement. However, in order to control the motor current, the voltage applied has to be varied due to the back EMF generated being proportional to motor speed. Hence, when calculating the energy requirement, both motor speed and torque have to be considered. For the purposes of this analysis, a conventional 12V brush DC motor of 1000W capacity was selected as the basis of these calculations. This is not necessarily the most suitable equipment but could be presented as a "worst case" scenario.

In addition to the motor characteristics, we have to apply the efficiency of the rest of the electrical system as well as that of the gearbox/steering gear system. When motoring (eg. dry parking) the figures would be: accessories drive (0.9); alternator (0.4); gearbox (at best 0.9) and steering gear (say 0.9). The total for this would be 0.29. Battery efficiency (approximately 0.7) would only enter the equation infrequently, such as when parking.

Higher voltage systems would be of definite advantage in the electrical efficiency area by reducing current, and, hence, resistance losses in the motor windings.

COMPARISONS

From figure 19, it can be seen that the OCCS mechanical system consumes between 15% and 50% of the energy consumed by the OC system. At very low speeds (driving in car parks), this rises to nearly 100%. This is because the OC system is running at the same speed as the OCCS at these vehicle speeds.

Figure 20 demonstrates the same comparison for the EHPS system. It is immediately apparent why these systems are designed with reduced flow. With the conventional flow rate (same as the OC system), the energy consumption is higher than that of the OC system up to 60 - 70km/h. This is due to the lack of efficiency of the electrical energy conversion.

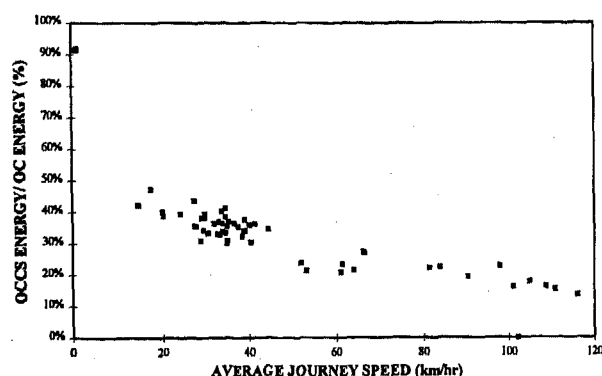


Figure 19 Comparison of Energy Consumption: OC to OCCS Mechanical Drive.

The reduced flow system, on the other hand, provides significant savings of between 50% and 90% over the OC system.

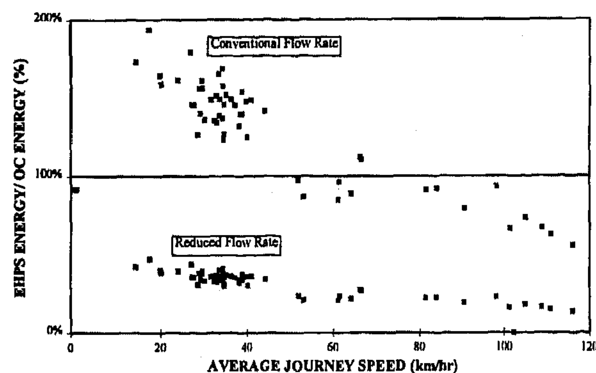


Figure 20 Comparison of Energy Consumption: OC to EHPS

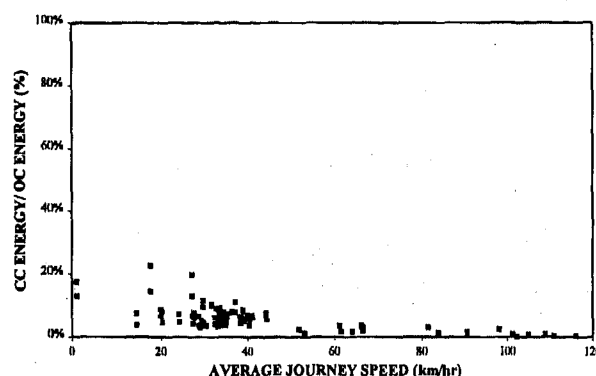


Figure 21 Comparison of Energy Consumption: OC to CC Mechanical Drive.

The energy consumption of the mechanically driven CC system is between 1% and 19% of that of the OC system if one assumes zero leakage. Adding the 0.2l/min leakage rate (see figure 21) raises the upper limit to approximately 23% but has very little effect on the relative consumption at higher average speeds.

With electric drive for the pump, the CC consumption (without leakage) rises to levels similar to that for the EHPS reduced flow system, although the CC system has an advantage at higher average speeds. The rise, again, is due to the inefficiency of the vehicle electric system. With accumulator leakage (figure 22) the maximum is raised to over 60% of that of the OC system.

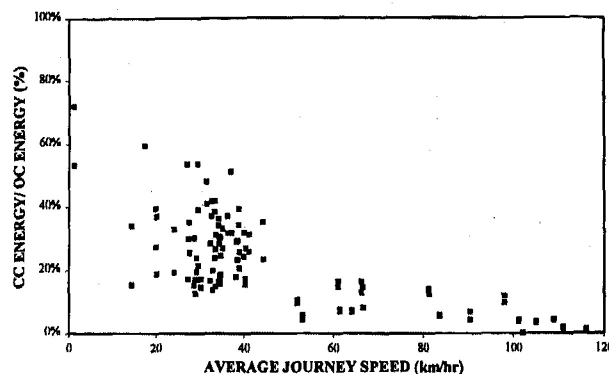


Figure 22 Comparison of Energy Consumption: OC to CC Electrical Drive.

Finally, figure 23 illustrates the energy consumption trends for the EPS system. Generally, consumption is of the same order as that of the CC mechanical with the CC system having an advantage at higher average speeds. This advantage at higher vehicle speeds is due to the fact that rack motion decreases with increasing speed (see figures 11,12 and 13) and the CC energy consumption is directly proportional to rack travel (ignoring leakage) whereas the EPS system can be drawing current even when the rack is stationary.

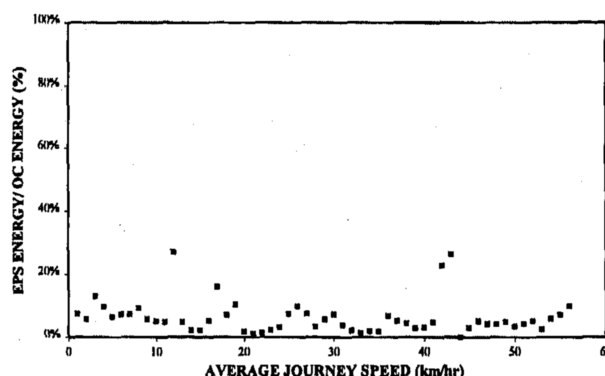


Figure 23 Comparison of Total System Energy OC vs. EPS.

CONCLUSIONS

The most energy efficient systems are the two "on demand" systems: CC mechanical and EPS, both systems exhibiting similar levels of energy consumption of less than 20% of that of the OC system. Even an accumulator leakage rate of 0.2l/min does not severely disable the CC system. However, future trends in powering the automobile would be likely to include an IC engine which is not required to run continuously (eg. hybrid vehicles) and these would probably require electric drive to accessories such as steering. This would tend to devalue the benefits of the CC system as the electrically driven version suffers a consumption penalty of approximately 2.5:1 over its mechanically driven brethren. Therefore, the EPS system would be most applicable in this scenario.

In terms of the current vehicle specifications, significant energy advantages can be gained from the use of the EHPS system, providing the "reduced flow" scenario, described above is utilised.

REFERENCES

1. Reproduced from information supplied by ROBERT BOSCH (AUSTRALIA) PTY. LTD.
2. Estimates of electrical efficiencies derived from information supplied by ROBERT BOSCH (AUSTRALIA) PTY. LTD.

970385

An Energy Efficient Electromagnetic Active Suspension System

Francis B. Hoogterp
U.S. Army TARDEC

Joseph H. Beno and Damon A. Weeks
The University of Texas at Austin

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The technology thrust to develop an effective electromagnetic actuator for application in an active suspension system has precipitated a fresh look at the active control schemes in an effort to reduce the required force levels of the actuator. The resulting "near constant force" control algorithm is described and its ability to greatly reduce vehicle sprung mass motion is documented through simulation and single wheel station laboratory test stand results. The vehicle power and energy requirements associated with this unwanted vehicle vertical are analyzed and comparisons between the corresponding passive and active systems are presented. The success of the active system leads naturally to the conclusion that a passive suspension equipped vehicle will become power limited at a much lower speed than will this active system when traversing severe cross-country terrain.

INTRODUCTION

The possibility of utilizing an electromagnetic (EM) actuator in a high-wheel travel active suspension system has recently gained considerable interest in both the commercial and the military sector [1-2]. The bus developers are interested in the potential benefits of active suspension for ride comfort and vehicle roll control as well as for the accompanying height and attitude control. Height and attitude control can be used to greatly simplify potential solutions to the handicap access requirements the bus manufactures must meet. The interest in an EM actuator based system is predicated on the availability of an energy efficient, and appropriately packaged system, that if exists would eliminate the need for a complex hydraulic system.

A passive suspension system is simply a traditional suspension system composed of springs, dampers (i.e. shock absorbers), and various structural linkages [3]. The springs and dampers of a passive system are appropriately tuned for the expected input frequencies, to attenuate the terrain input (as seen by the tires). The purpose of the shock absorbers (or dampers) is to dissipate the energy that is input into the chassis by the ground through the suspension system. The more energy that is allowed to be input to the chassis, the more energy that the shock absorbers must dissipate, and the more energy that the vehicle's power train must provide. A semiactive suspension system is defined here as a passive system which has adaptive damping characteristics. The system has the traditional springs and dampers, but the vehicle motion is constantly monitored, and the damper characteristics are reduced (or set to zero) when the damping forces would increase chassis motion. It has been demonstrated that semiactive suspension systems can moderately increase the ride quality of a vehicle while simultaneously slightly decreasing the energy dissipated by the suspension system [4-5].

An active suspension system represents a distinct departure from traditional suspension system concepts. Springs and dampers are no longer required (though for practical reasons springs are often still included). The suspension forces that are normally provided by the springs and dampers are now provided by some form of actuator (e.g. hydraulic, pneumatic, or EM) based on the output of vehicle motion based control algorithms. The active suspension control research for rough terrain negotiation has primarily focused on more sophisticated approaches to the attenuation of the ground induced inputs [6-7]. The success of active suspension has been quite impressive in the areas of on-road stability and handling and has been demonstrated repeatedly. The cross-country ride problem, however, has not proven as easy to conquer to date.

The military's interest is active suspension systems also begins with significantly improving vehicle ride and stability, but for the military, cross-country ride is the focus of interest. The Army's thrust to remove hydraulics from its ground combat vehicles precipitated its interest in EM actuator based active suspension systems. The technology thrust to develop an electric drive combat vehicle has only served to heighten this interest since adequate electric power would be available on the vehicle and the addition of a hydraulic system for active suspension would only add to the overall vehicle complexity.

The intended application of such a system would be to obtain maximum high-speed cross-country performance desired by military applications. The major barrier to making such an application practical has been the weight and power requirements of an electromagnetic actuator based system. Since the actuator weight for such a system is driven largely by the actuator force requirements and therefor indirectly by its power demands, minimizing the required active suspension power will also greatly reduce the system weight.

An active suspension control algorithm is presented that requires significantly less power (and correspondingly lower forces) than do other algorithms. This "near constant force"

strategy attempts to prevent the excitation of the vehicle's sprung mass rather than to attenuate the motion once it has begun. The resounding success of this algorithm is demonstrated through a series of simulation and laboratory test stand results. The laboratory test stand results include test data obtained from a single wheel station, rotary EM actuator, supporting a sprung mass of 4500 kg as well as test data from a reduced scale, 4 wheel representation of a high mobility off-road vehicle. Simulation results corresponding to the test data are also presented.

The power and energy requirements of this active suspension configuration are actually shown to be significantly lower than those for the comparable passive suspension configured vehicle. In addition the acceleration levels of the sprung mass are also dramatically reduced resulting in the operational maximum cross-country speed being significantly increased over, not only the passive system, but over other proposed active suspension configurations as well.

NEAR CONSTANT FORCE ACTIVE SUSPENSION CONTROL ALGORITHM

The underlying philosophy for the "near constant force" (or NCF) active suspension system algorithm is quite simple. Since the purpose of a suspension system is to isolate the chassis (or sprung mass) from the motion of the wheel (or unsprung mass) which is trying to follow the terrain, a perfect suspension system (in terms of isolation) would always generate only the static force to the chassis. With a constant vertical force on the chassis, the chassis would then be subjected to zero vertical acceleration and therefor a perfect ride. In practice, of course, this is simply not practical since the vehicle must go up and down hills and the suspension travel is limited to a few inches at best.

The idea persists, however, that the best active suspension control algorithm would be the one that would provide the least variation in suspension forces while minimizing the prospect of exceeding suspension travel. Thus was born the NCF control algorithm. The perfect solution of this problem would require prior knowledge of the terrain, but the development of the NCF algorithm was developed without the assumption of this prior knowledge and yet has been shown in the laboratory to perform at an extremely high level of efficiency.

A model of the original laboratory test stand for which the NCF control scheme was developed, is shown in figure 1. This model represents (as does the test stand) one of the fourteen roadwheel stations and one-fourteenth of the hull mass of a main battle tank (the M1 Abrams). The road-wheel is mounted on a trailing arm and the hull mass is constrained so that it can move only in the vertical direction. The suspension vertical force to the hull is actually applied as a torque between the road arm and the hull at the road arm pivot. Since this model is restricted to vertical motion, it can be viewed the same as a 1/4 car model would be for a four wheeled vehicle. Figure 2 presents the predicted motion of this model's hull mass for the standard torsion bar suspension system on the M1 tank to the terrain input shown. This terrain is extremely rough as indicated by its terrain roughness index of 8.9 cm rms. The gradual hills were removed from the terrain prior to the simulation to comply with the travel

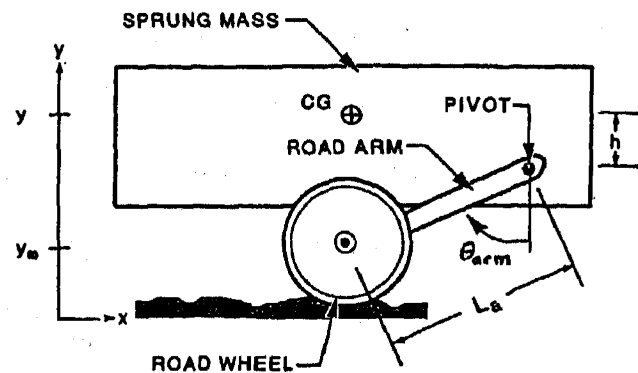


Figure 1 One-dimensional tank suspension model

limits of the laboratory terrain generation hardware. This simulation represented a vehicle velocity of 16 kph and it can be seen from the plots that for this very stiff passive suspension system, the hull tends to follow the terrain.

Any practical active suspension design will include a passive spring to statically support the vehicle and to reduce the overall power requirements of the active system and the EM actuator based system is no different. The demonstration rotary actuator built on a main battle tank scale, incorporated air spring for this purpose. This permitted the spring rate to be kept quite low while still providing the required static force. The first task of the NCF algorithm is to negate the effect of this low rate spring which varies with the arm angle of the trailing arm suspension system. The NCF algorithm was developed based on the premise that it would have available only measurements that could be obtained from the vehicle itself. These could include arm angle and velocity making the cancellation of the spring force a simple matter, and could also include, as needed, chassis (or hull) vertical, pitch, and roll accelerations, velocities, and displacements (or estimations thereof). The resulting actuator

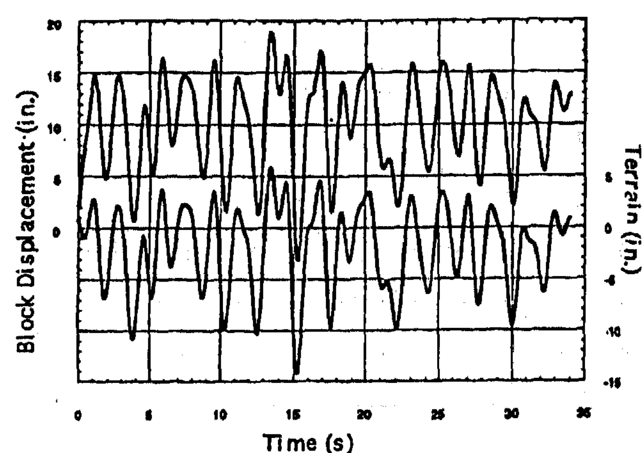


Figure 2 Simulated hull motion at 16 kph for passive M1 torsion bar system

force from the NCF algorithm for a 1/4 car model (i.e. considers chassis vertical motion but not pitch or roll) is given by:

$$F_a = -F_s(\theta_{arm}) - k_1[y - y_{ref}] - k_2 y' - k_3 \theta_{arm} - k_4 \theta'_{arm} \quad (1)$$

where $F_s(\theta_{arm})$ is the force required to offset the air spring's additional force (as indicated, F_s is a function θ_{arm}), y is the vertical displacement of the chassis, θ_{arm} is the angle of the roadarm, y_{ref} is the height of the desired chassis vertical trajectory, and k_1, k_2, k_3, and k_4 are appropriate gains. Of course the displacement of the chassis, y, and of the desired reference trajectory, y_{ref} , are not readily available to the controller and so the gain k_1 is set equal to zero in the NCF algorithm. The term is left in equation (1) to emphasize the fact that it is ideally what is desired to be controlled. However much can be discerned about the nature of the term $[y - y_{ref}]$, from the more easily measured signals y' and θ_{arm} and feedback based on these terms is used instead of on $[y - y_{ref}]$.

Figure 3 shows the hull motion for a 16 kph simulation as did figure 2, but in this case an active suspension system was assumed and the suspension force is determined by the one-dimensional NCF algorithm. Also, as was done for figure 2, the gradual hills had been removed from the terrain. Figure 4 shows the actual results from the laboratory test stand corresponding to the predicted results of figure 3. The comparison is quite comforting and lends credibility to the attempt to greatly decrease hull motion in response to severe cross-country terrains.

The appropriate choice of gains in (1) allows the NCF algorithm to actually respond to hills while effectively filtering out the higher frequency inputs inherent in rough cross-country terrain. Figure 5 presents the simulated hull response to the same terrain used in the previous examples, but where the gradual hills have been left in the terrain. This simulation was performed at a vehicle speed of 64 kph and shows almost a total attenuation of the higher frequency inputs.

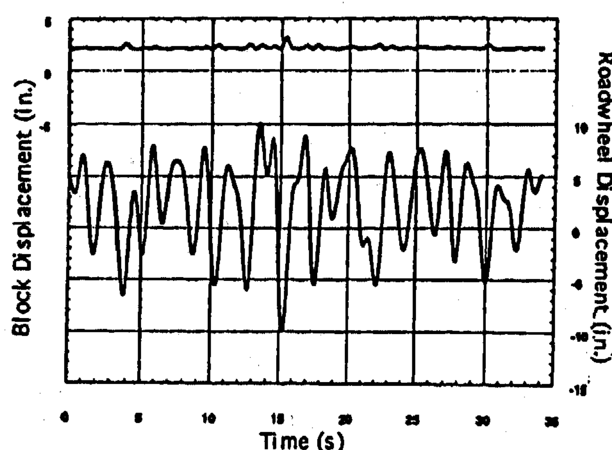


Figure 3 Simulated hull motion at 16 kph for NCF active suspension system

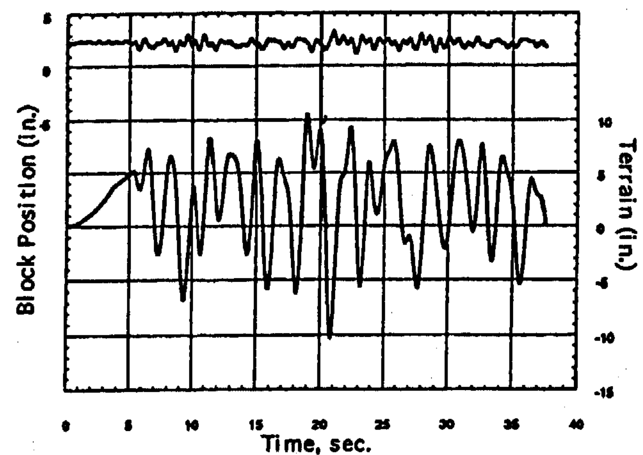


Figure 4 Laboratory test stand hull motion at 16 kph for NCF active suspension system

The version of the NCF algorithm presented in (1) is concerned with minimizing vertical chassis motion only. The extension of the NCF algorithm to a full 3-dimensional vehicle model must also consider the chassis pitch and roll motions in calculations. The NCF actuator force for this full vehicle model can be similarly be written as:

$$F_a = -F_s(\theta_{arm}) - k_1[y - y_{ref}] - k_2 y' - k_3 \theta_{arm} - k_4 \theta'_{arm} - k_5 \theta - k_6 \theta' - k_7 \alpha - k_8 \alpha' \quad (2)$$

where θ and θ' are the chassis pitch angle and pitch rate, α and α' are the roll and roll rate, and the additional k's are appropriate gains as before. This 3-d version of the NCF active algorithm was implemented to control a 30 % scale, 4 wheeled active suspension test rig representing the Army's High Mobility Multi-purpose Wheeled Vehicle (or HMMWV). Preliminary results from this test rig were similar in terms of chassis motion reduction, to those discussed earlier for the single wheel station system.

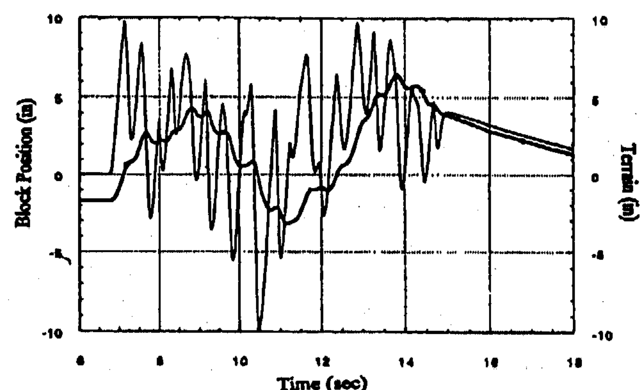


Figure 5 Simulated hull motion at 16 kph for NCF active suspension with hill following employed

POWER COMPARISONS FOR ACTIVE AND PASSIVE SUSPENSION SYSTEMS

It is clear that when the sprung and unsprung masses are moving up and down, that some of the power that was intended to drive the vehicle forward is being expended in the vertical direction. This power is, of course, supplied by the vehicle's engine and is delivered through the inefficiencies of the drive train to provide motion to the vehicle. Power expended in exciting the vehicle vertically is totally wasted, and if it becomes excessive will significantly limit the overall performance of the vehicle. It will be demonstrated that this is in fact the case for the passive suspension system on the M1 when trying to negotiate very rough terrain.

The unsprung mass (roadwheel or tire) is intended to follow the terrain and thus must be continuously accelerated in the vertical direction when traversing rough terrain. The unsprung mass, however, is usually less than 10 % of the sprung mass and thus is a rather minor drain on the system power than is the sprung mass at the same excitation levels. The active suspension system cannot, therefor, effectively reduce the unsprung mass motion of the power required of the engine to produce that motion.

The sprung mass on the other hand has no requirement to follow the high frequency component of the terrain input, and in fact, the suspension systems very purpose is to keep it from doing so. Unfortunately, the passive system, as was demonstrated in figure 2, is not very effective at accomplishing this. In the passive system the power that is input into the unwanted vertical (as well as pitch and roll) motions is ultimately dissipated by the shock absorbers and lost. In order to compare the vertical motion power between the active and passive suspension systems, vertical sprung mass power was computed as:

$$\text{Power} = F\,v = F\,a\,v \qquad (3)$$

where F is force, v is velocity, m is mass, and a is acceleration. The resulting comparisons that are shown below are for a single wheel station as represented in figure 1, where the sprung mass represents $1/14^{th}$ of an M1's sprung weight (the M1 has 14 road-wheels).

Figures 6 and 7 depict the power required to accelerate the sprung masses for the passive and NCF active suspension systems at 4.47 m/s and 17.8 m/s. The simulated results are over the same severe cross-country terrain that was used in the previous examples. The passive system employed the standard M1 torsion bar and rotary damper whereas the active system was modelled with the electromechanical suspension active suspension system described in [1] and utilizing the NCF control algorithms described above. At the lower speed, figure 6 shows that the rms and peak powers associated with both systems is quite small. At 17.8 m/s, however, the difference in power between the two systems is dramatic as demonstrated in figure 7. The passive system requires 540 kW peak and 71 kW rms power just to accelerate the sprung mass of a single wheel station along its vertical trajectory. Projecting this rms power over the 14 road-wheels of the M1 would result in a total sprung mass acceleration rms power of 994 kW, or almost 90 % of the M1 engines rated gross power. It is realized that not all the M1 wheel stations are identical (only 6 of the 14 have dampers) and vertical motion of the sprung mass is somewhat abated over the length of the vehicle, but this power figure does not include contributions from the hull pitch and roll motions either.

It should be noted that the passive M1 could not attain nearly this speed (i.e. 17.8 m/s) across this terrain even if the vehicle and crew could stand the accelerations generated by such a run. The vehicle is in fact power limited long before this speed even though this is approximately the level surface top speed of the M1.

Figure 8 portrays the peak and rms power levels for this same pair of suspension systems as a function of vehicle speed. The terrain is the same cross-country terrain that was used previously and the plots were generated from simulation results using the same single wheel station model. The item of major interest here is the extreme divergence in power requirements at higher vehicle speeds between the passive and the NCF algorithm based active suspension system. Thus high speed travel on such a

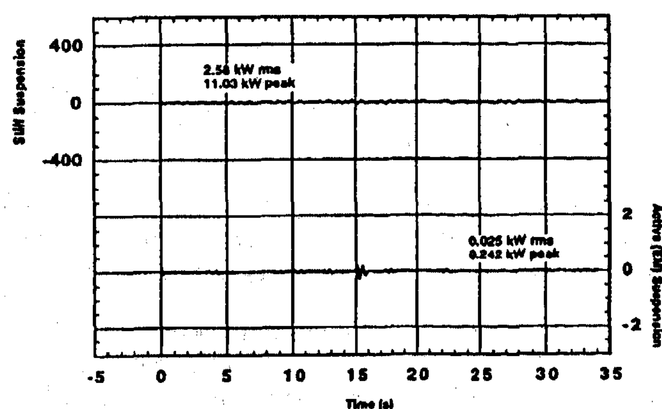


Figure 6 Sprung mass vertical power requirements for active and passive systems at 16 kph

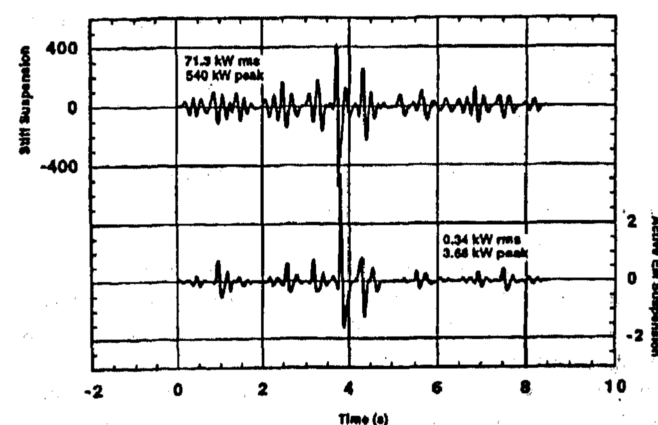


Figure 7 Sprung mass vertical power requirements for active and passive systems at 64 kph

terrain simply cannot be accomplished by this passive suspension system (or probably any other practical passive system). It will require instead, not only an active suspension system, but an active system that significantly reduces unwanted vehicle hull motions.

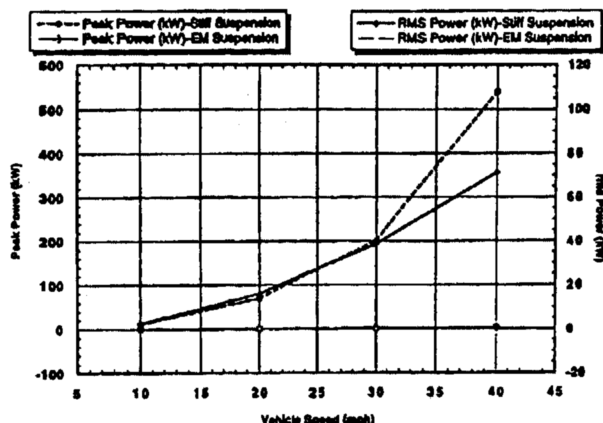


Figure 8 Power consumption between active and passive systems as a function of vehicle speed

ENERGY USAGE

When considering energy consumption of the passive and active suspension systems, the vehicle drive train and active suspension power delivery system efficiencies must be considered. At the 17.8 m/s speed, the passive system required 994 kW rms of power to provide the sprung mass vertical acceleration of the total vehicle. Since the power to accomplish this motion must be provided through the vehicle's drive train, the associated additional power losses can also be attributed to the excessive vehicle motion. If it is assumed that the vehicle automatic transmission and differential each operate at 90 % efficiency, these losses are approximately 233 kW.

These power train losses would represent the total power lost if the energy imparted to the sprung mass were completely recoverable, however this is not the case. Separate simulations show that an individual M1 shock absorber can dissipate as much as 95 to 100 kW of power when the vehicle is driven at 17.8 m/s over this same terrain. It should be noted that there are six shock absorbers on the M1 (on the front two and the rear wheel stations on each side) but the other wheel stations also have some inherent damping associated with them also.

The rms power required for the active system to accelerate the sprung mass was shown to be only 5 kW. This power can be applied by one of two paths to the sprung mass. As in the passive case, the power can be delivered through the vehicle drive train with the partially misdirected goal of vehicle forward motion in mind. Alternatively, the power to accelerate the sprung mass can be delivered directly through the active suspension system actuators. If in this latter case, one assumes that the active suspension power conversion device (for either mechanical to hydraulic or mechanical to electrical) and the suspension force actuator are both 90 % efficient, then the overall efficiency of the power delivery process is 81 % regardless of whether it is delivered through the vehicle drive train or through the active suspension system. Thus the total power required from the engine to provide the sprung mass motion for the NCF active system is only 6 kW even if none of the sprung mass motion energy is recoverable.

The power losses attributable to the unsprung mass vertical motions have not been considered here but they are not avoidable since the wheels must be made to follow the ground in order to be able to provide the necessary hull (or chassis) support. Ground contact is also required, of course, to permit adequate vehicle stability and handling. In either case, the wheel motions of both the active and the passive systems should be quite comparable making the power lost do to unwanted sprung mass motion the major power differential between the two systems.

CONCLUSIONS

The attempt to negotiate severe cross-country terrain at high speeds with conventional suspension equipped vehicles has been shown, to not only impart excessive motion to the vehicle hull or chassis, but to also require such high levels of available engine power that the higher speeds are simply not possible. The "near constant force" (or NCF) active control algorithm has been defined, and when coupled with an EM suspension system has been shown to reduce the sprung mass motion power by a factor of about 200. The NCF approach to active suspension control requires much lower actuator force levels than do previous active control schemes, thus making the prospects of a competitively sized EM actuator quite plausible.

REFERENCES

1. J.H. Beno, F.B. Hoogterp, D.A. Bresie, S.K. Ingram, D.A. Weeks, and W.F. Weldon, "Electromechanical Suspension for Combat Vehicles," SAE Techninal Paper No. 950775, March 1995.

2. F.B. Hoogterp, J.H. Beno, and M.R. Whitmore, "The State of Electromechanical Active Suspension Technology," Proc. of First International Conf. on All Electric Combat Vehicle (AECV), Haifa, Israel, May 1995.

3. D.A. Crolla, D.N.L. Horton, R.H. Pitcher, and J.A. Lines, "Active suspension control for an off-road vehicle," IME Vol. 201, No. D1, pp. 1-10, 1987.

4. L.R. Miller and C.M. Nobles, "The Design and Development of a Semi-active Suspension for a Military Tank," SAE Technical Paper No. 881133, 1988.

5. N.L. Temple and F.B. Hoogterp, "Semiactive Suspension: A Mobility Enhancement for Combat Vehicles," Proc. of ISTVS Conf. on Off-Road Vehicles, London, England, June 1992.

6. D. Hrovat, "Applications of Optimal Control to Advanced Automotive Suspension Design," Trans. of ASME Journal of Dynamic Systems, Measurement, and Control, vol. 115, no. 2B, pp. 328-342, June 1993.

7. P.W. Claar II and J.M. Vogel, "A Review of Active Suspension Control for On and Off-Highway Vehicles," SAE Technical Paper No. 892482, 1989.

970386

Semiactive Suspension: A Mobility Case Study

Nancy L. Saxon and William R. Meldrum, Jr.
U.S. Army Tank-Automotive and Armaments Command

ABSTRACT

The U.S. Army has initiated a program to assess the performance potential of a semiactive advanced suspension system for its combat vehicles. The program utilizes the Bradley Fighting Vehicle as a test bed for a semiactive external in-arm hydropneumatic suspension. The system hardware includes the suspension and associated plumbing, computer controller, and dynamic track tensioner. This paper describes the comparable benefits of the semiactive suspension over the passive suspension through both laboratory testing and simulation results. Transmissibility analysis of the laboratory data document far less hull motion in the semiactive case. Simulated vehicle mobility again shows less vehicle motion and superior ride quality using the semiactive suspension.

INTRODUCTION

The superior mobility of a military vehicle provides the combat crew with a tactical advantage through increased cross country speed. The suspension system plays a pivotal role in evaluating a vehicle's mobility. The primary function of a vehicle's suspension is to cushion the vehicle body and its contents from terrain induced vibrations to permit higher vehicle speeds. Secondary, and no less important, the crew must have a measure of comfort which will allow them to be operationally effective. The human body's tolerance limit to these terrain induced vibrations is reached before the tolerance limit of the vehicle itself. Therefore, the human body tolerance limit is used to evaluate a vehicle's performance and is termed ride quality. The Army has established a quantitative measure of ride quality which relates the terrain, the driver, and the vehicle dynamics. The Army uses ride quality to predict and establish ride characteristics, evaluate suspension systems and analyze weapon stabilization systems. This unit of ride quality is called "Absorbed Power" [1].

Absorbed power was a criteria developed in the late 1960's by the Army through scientific experiments in whole body vibrations using human subjects. These experiments established a human transfer function of energy transmitted to the body. Shortly after these experiments, the Army conducted field tests and established a "driver preferred limit" of 6 watts of absorbed power. In 1980 this limit was incorporated into system performance specifications of Army vehicle solicitations. Figure 1 is a pictorial representation of the ride measurement process. Force and acceleration signals are sent through the human transfer function and integrator respectively and then through an averaging circuit which is then displayed as a single number for a given piece of cross-country terrain.

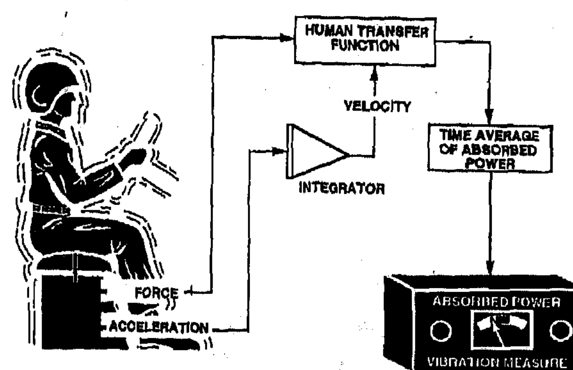


Figure 1. Ride Measurement Process

Not only does the vehicle require good ride quality, it also requires the suspension to provide for such things as good performance on side slopes and accommodate high centers of gravity. It also must utilize the suspension travel so that the vehicle road wheels are on the track and producing tractive effort in all types of terrain situations. In order to accomplish both of these missions, good ride quality and good terrain negotiation,

the selection of the spring and damping rates are a compromise. In the past the standard spring was a torsion bar and the damping was either a linear or rotary shock absorber.

Attempts to develop a suspension which would effectively accomplish both missions stated above lead to research into external in-arm hydropneumatic suspension technology, and more recently semiactive suspension technology [2] [3]. The in-arm suspension is one in which all of the suspension components are exterior to the hull of the vehicle. The suspension components are fitted within the roadarm (there is no separate strut or shock absorber), see Figure 2. The nitrogen gas provides the springing and the damping is provided by friction plates, see Figure 3. The suspension unit is charged with two different nitrogen charges. A higher pressure at the upper end of the nitrogen chamber and a lower pressure near the low pressure piston. As the roadwheel is forced in the jounce direction, the crank forces the low pressure piston down the cylinder increasing the nitrogen pressure giving the initial spring rate. When the lower chamber pressure reaches that of the upper chamber, the high pressure piston also begins to move, thereby increasing the volume of gas to be compressed and softening the spring. The dual piston/nitrogen chambers give the in-arm units their dual spring rates as shown in Figure 4. Some of the benefits of the in-arm system over the torsion bar system are weight and space savings, weight class commonality, and greater load leveling capability. Also, lockout and height control can easily be incorporated into this type of suspension.

Further research lead to the incorporation of the semiactive technology into the external in-arm hydropneumatic suspension units. Incorporation of this technology is considered very low risk and can provide the additional benefits of increased vehicle mobility, platform stability, ride comfort, and faster cross-country speed. Another expected benefit is increased component reliability due to decreased shock loads to the hull.

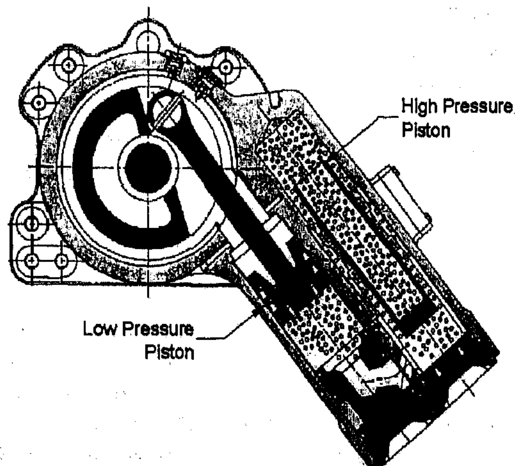


Figure 2. Hydropneumatic External In-Arm Suspension Cross-Section

For the purpose of this paper, a semiactive suspension is defined as one in which the damping forces are computer controlled in real-time. The system uses a damping control strategy to optimize ride and/or gross body motion of the vehicle. Minimizing the vehicle motions results in a more stable, and therefore a more level vehicle, with more even weight distribution over rough terrain. The benefits of semiactive technology has been proven in past tests sponsored by the Army and has shown an increase in a vehicle's cross-country speed by as much as 50 percent on an experimental vehicle [2] [3]. The objective of semiactive control is to minimize vehicle motion without degradation to ride quality and thereby increase cross country speed.

HARDWARE

The basic suspension utilized in this program is an external hydropneumatic in-arm suspension. Previous designs for passive units have included both the heavy and medium weight class. A variation of medium weight class design was incorporated into the Bradley test bed semiactive suspension design and is shown in Figures 2 and 3. Again, the passive technology is fully developed and therefore capable of being utilized as a platform for the semiactive controller.

The Bradley program's main goal is to design and fabricate a medium weight class in-arm vehicle set incorporating semiactive control for field testing. The program also includes the incorporation of an active track tensioning system and suspension lockout.

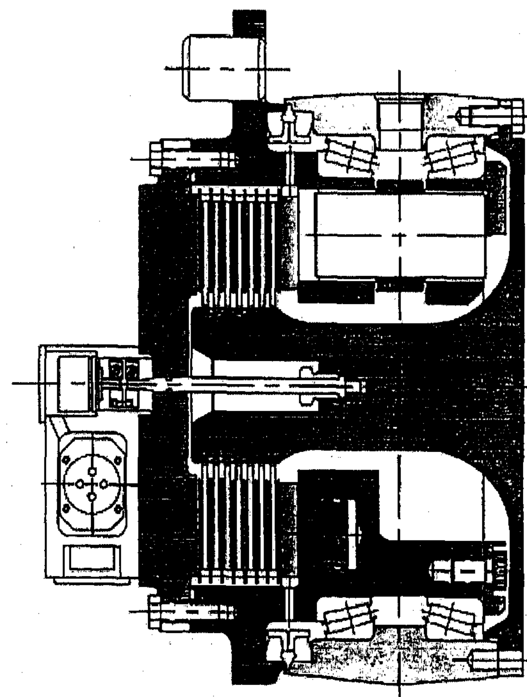


Figure 3. Rotary Friction Damper

Again, the spring curve for the units is shown in Figure 4. As described earlier, this is a dual rate spring providing a softer spring at higher levels of jounce travel. The damping curve is shown in Figure 5. The damper is a "bang-bang" damper and utilizes an "on-off" control strategy with a amplitude controlled dead-band.

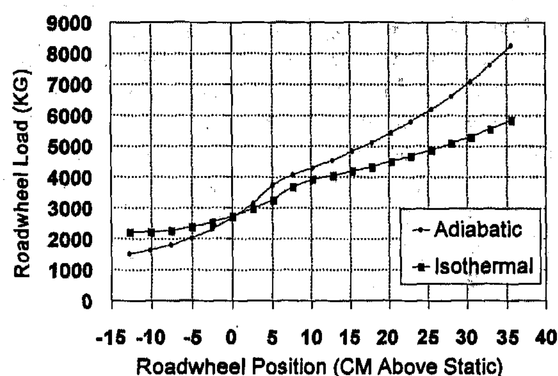


Figure 4. Suspension Unit Spring Curve

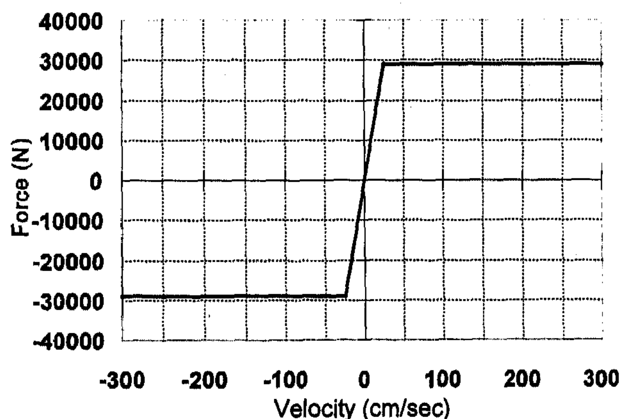


Figure 5. Suspension Unit Damping Curve

LABORATORY TESTING - PHASE I

The Bradley program was set up in a two phase approach. The first phase being a single unit road wheel station prototype to be tested on a suspended mass test rig. The second phase is a full vehicle set fabrication, installation, and field testing.

The laboratory testing was conducted on a single-unit suspended mass test rig, as shown in Figure 6. The simulated hull or suspended mass consisted of a 3208 kg block of concrete attached to a cantilever beam. The suspension unit was bolted to the concrete block and both the simulated hull mass and the suspension were instrumented. An actuator under the road wheel was fed a velocity dependent signal from digitized terrain and sinusoidal inputs.

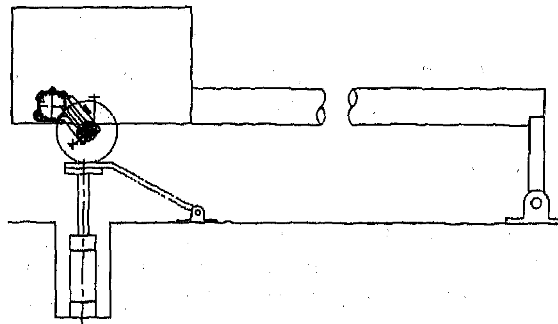


Figure 6. Suspended Mass Laboratory Test Rig

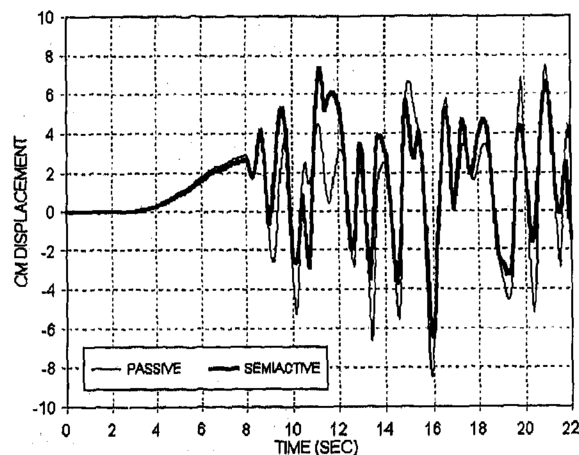


Figure 7. Laboratory Test Results Vertical Displacement 10 cm Bump Course at 32 kph

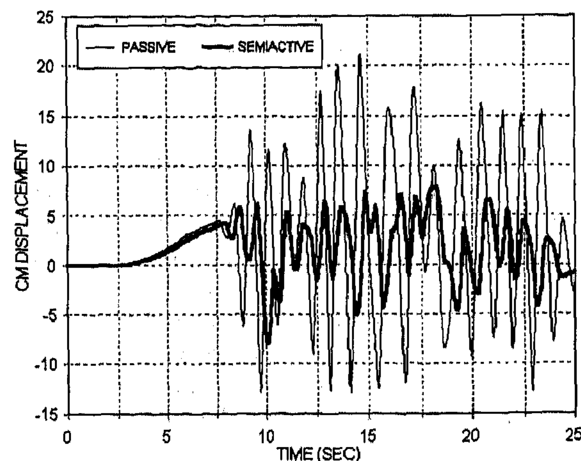


Figure 8. Laboratory Test Results Vertical Displacement 15 cm Bump Course at 32 kph

The objective of the laboratory testing was to insure both the hardware and controller were functioning properly. Minor adjustments in the controller algorithm were made to receive maximum benefit of the semiactive suspension.

Figures 7 and 8 show vertical displacement of the suspended mass for passive versus semiactive suspension configurations over a ten and fifteen centimeter bump course respectively. As shown in the above figures sprung mass amplitude can be markedly decreased with the use of semiactive control.

SIMULATION RESULTS

The simulation code used to predict the performance benefits of the semiactive suspension is a two dimensional dynamics model, Vehdyn II. Vehdyn II is a model developed by the Army Corps of Engineers [4]. It is a module included in the NATO Reference Mobility Model. The model is used to predict gross vehicle motions over rough terrain, to calculate absorbed power and peak acceleration and predict the ride and shock limiting speeds of a vehicle [5]. Figure 9 is a pictorial representation of the Bradley model. Some of the model limitations include the inability to model the roadarms or any nonvertical spring or damper effects, and the track is modeled as a spring element between adjacent wheels. The model is however, quite useful in determining overall vehicle performance without extensive computer time and has been validated with vehicle field tests.

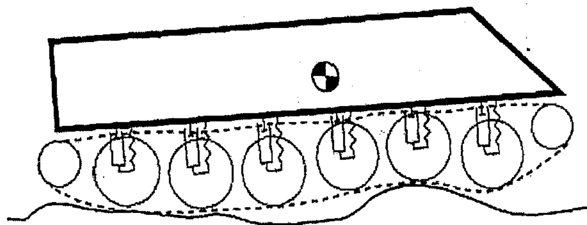


Figure 9. Pictorial Representation of Vehdyn Model

Figure 10 shows a simulated ride curve comparing the semiactive and passive suspension on a Bradley Fighting Vehicle. As shown there is a significant difference in the speed attained before the limiting six-watts of absorbed power is reached. The shock performance over half-round obstacles did not however show significant differences in the speed at which 2.5 g vertical acceleration is reached. Further evaluation and field testing is planned to confirm these findings.

Figures 11 and 12 show the pitch time history at the drivers location of the passive versus the semiactive suspension over a 3.81 and a 8.30 cm RMS roughness terrain. The vehicle speed was selected as the six-watt ride limiting speed for the passive configuration. As can be seen in the plots, the semiactive suspension control will reduce the pitch of the vehicle. More significant reductions can be seen with increasing terrain roughness. Similarly, Figures 13 and 14 show the heave time history over these same courses, and as in the pitch, heave can also be reduced. Table 1 is a consolidation of the RMS, max, and min of the pitch, heave and vertical acceleration signals for various roughness terrain. The maximum benefit for the semiactive suspension can be found in the RMS range up to 6 centimeters RMS. The average decrease in pitch motion over all terrain roughness is 16.2 percent. The vertical acceleration was decreased an average of 51.3 percent. Figure 15 shows a histogram of the vertical acceleration signal for 30.58 kph over a 3.81 cm RMS terrain. As shown in the histogram, the majority of the benefits of the semiactive suspension can be found in the reduction of the frequency of accelerations over 0.25 g's in the positive direction. It is believed that with further optimization of the semiactive algorithm, improvements should also be seen in the negative direction.

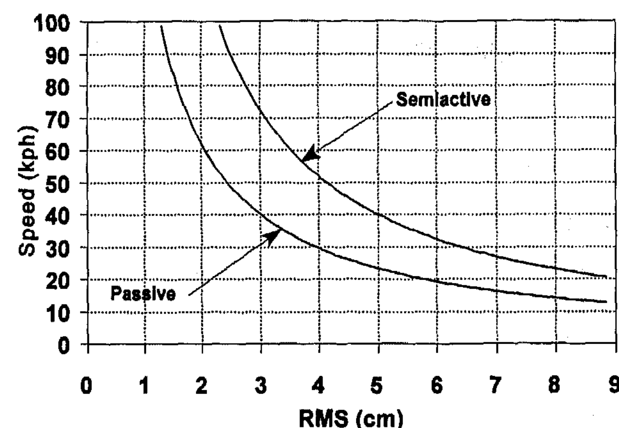


Figure 10. 6 Watt Simulated Ride Curves

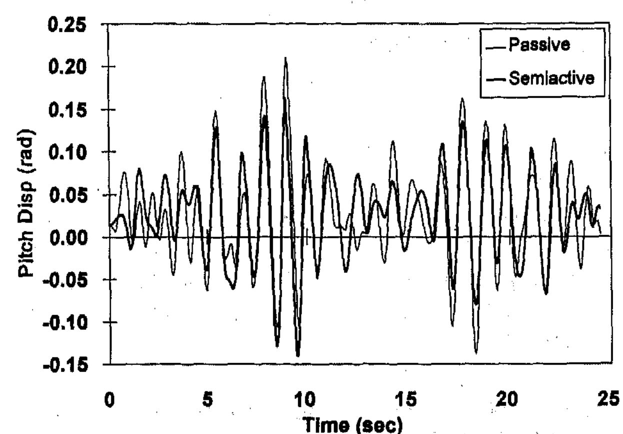


Figure 11. Simulated Pitch Time History 3.81 cm RMS Course @ 30.58 kph

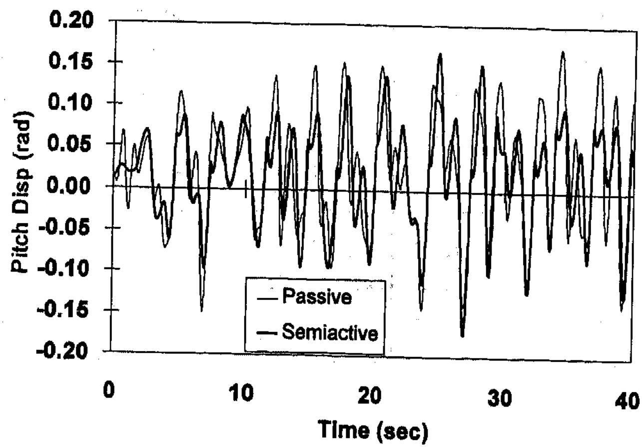


Figure 12. Simulated Pitch Time History 8.3 cm RMS Course @ 11.27 kph

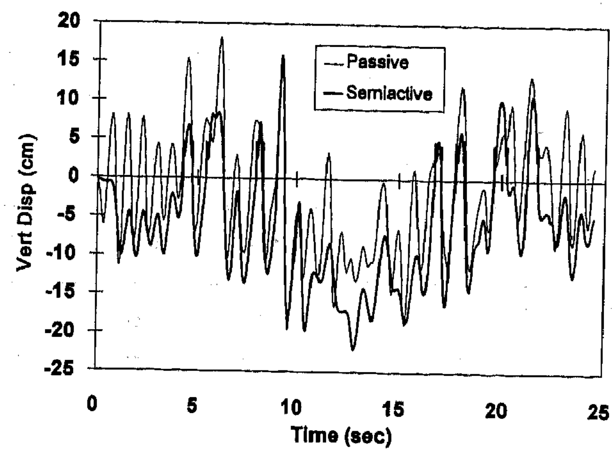


Figure 13. Simulated Heave Time History 3.81 cm RMS Course @ 30.58 kph

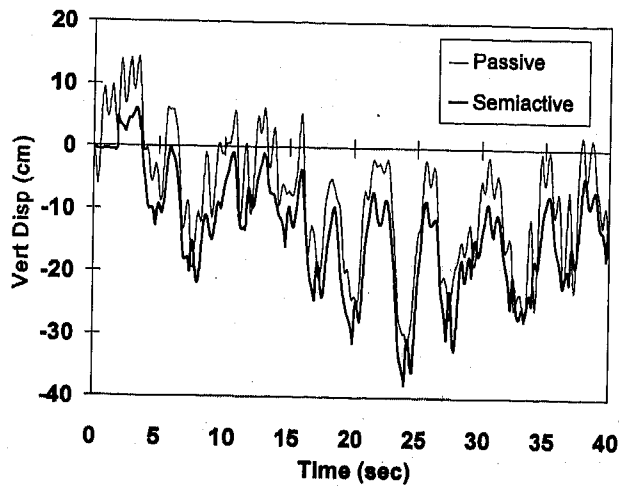


Figure 14. Simulated Heave Time History 8.3 cm RMS Course @11.27 kph

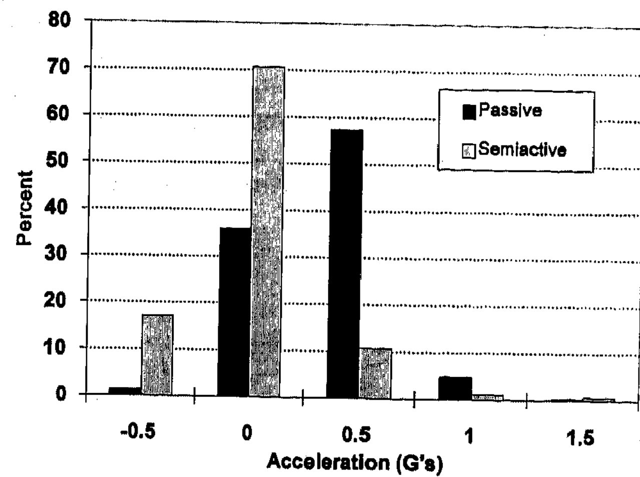


Figure 15. Histogram of Vertical Acceleration Time History of 3.81 cm RMS Course at 30.58 kph

Table 1. Consolidation of Pitch, Heave and Vertical Acceleration Results

Vehicle / Course RMS	Max Pitch (Radians)	Min Pitch (Radians)	RMS Pitch (Radians)	Max Heave (cm)	Min Heave (cm)	RMS Heave (cm)	Max Vert Acc (g's)	Min Vert Acc (g's)	RMS Vert Acc (g's)
Passive / 2.64 cm	0.128	-0.067	0.051	111.79	-6.57	63.44	1.20	-0.74	0.51
Semiactive / 2.64 cm	0.095	-0.120	0.041	105.46	-11.73	59.43	0.64	-0.61	0.18
Passive / 3.81 cm	0.210	-0.138	0.065	18.19	-19.53	7.79	1.83	-0.46	0.49
Semiactive / 3.81 cm	0.164	-0.140	0.057	15.90	-22.02	9.93	1.41	-0.91	0.25
Passive / 4.66 cm	0.154	-0.123	0.068	21.16	-82.27	73.72	1.14	-0.50	0.40

Semiactive / 4.66 cm	0.134	-0.099	0.060	9.04	-87.20	47.29	0.90	-0.78	0.23
Passive / 5.41 cm	0.174	-0.185	0.094	10.08	-197.43	102.04	1.53	-0.39	0.43
Semiactive / 5.41 cm	0.141	-0.165	0.071	4.83	-203.20	105.5	1.79	-0.62	0.27
Passive / 8.30 cm	0.173	-0.149	0.071	14.25	-32.08	12.78	1.33	-0.39	0.55
Semiactive / 8.30 cm	0.167	-0.173	0.062	5.99	-37.92	16.49	1.76	-0.53	0.20

VEHICLE INSTALLATION - PHASE II

Phase II of the project is the fabrication and installation of the vehicle set on the Bradley test bed. One of the advantages of using an external in-arm suspension is the space savings by eliminating the torsion bars from crossing through the hull. Although on this project the space savings will not be utilized since this is a retrofit to an existing vehicle. The existing and new machined hull is shown in Figure 16. The retrofit of the new suspension requires road wheel relocation and the removal and relocation of support rollers, but otherwise the wheel travel or belly clearance is unaffected.

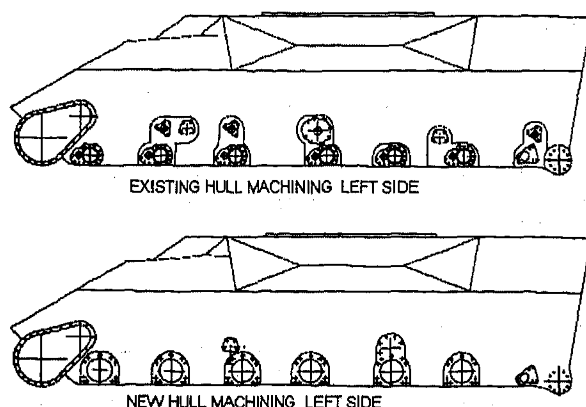


Figure 16. Vehicle Hull Modification

The installation requires plumbing the hydraulic and running the electrical cabling. The hydraulic power comes from a power take-off from the vehicle's engine. The hydraulic reservoir and pump will fit within the engine compartment. The controller resides in a 486 computer built in a rugged case and will mount in the crew compartment. On-board data acquisition is included for all accelerometers and potentiometers. Download links exist to download the data to other platforms.

PLANNED FIELD TESTING

Field testing is planned for the summer of 1997 and will consist of ride, shock, slalom and traverse testing against a passive Bradley Fighting Vehicle at the same weight. Information such as accelerations in the vertical, lateral, and fore-aft directions at different locations within the vehicle will be taken. Comparisons of six-watt ride limiting and shock speeds will be made. Pitch, heave, and roll movements will also be evaluated. Comparisons to the simulated vehicle is also planned.

CONCLUSIONS

Improvements in vehicle mobility can be gained with the use of semiactive suspension technology. This has been proven in both field tests and simulation results. The Bradley Fighting Vehicle has been shown to have superior mobility through the use of a semiactive suspension from both laboratory testing and simulation results. The use of these computer simulation codes make the technical evaluation of advanced suspension systems such as the semiactive suspension presented in this paper both a cost-effective and practical approach to exploring the Army's mobility needs. After initial computer evaluations, hardware should be built and evaluated if the simulation results show promising improvements. Prototype hardware mobility evaluations are still a necessary and prudent approach to long-term vehicle sub-system improvements.

The improvements in ride quality shown in this paper from both the laboratory testing and the computer simulation results are hoped to be duplicated in vehicle testing. Further publications are planned at the conclusion of the vehicle testing phase of this program.

REFERENCES:

1. Lee, R. A., Pradko F., "Analytical Analysis of Human Vibration", SAE Paper No. 680091, Jan. 1968.

2. Hoogterp, Dr. Francis B., Saxon, Nancy L., Schihl, Peter J., "Semiactive Suspension for Military Vehicles", SAE Paper No. 930847, Mar 1993.

3. May, Carl R., Green, Charles E. Mobility Performance Evaluation of the Mobility Technology Test bed (MTTB), Technical Report GL-93-5, Apr 93; U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

4. Creighton, Daniel C. Revised Vehicle Dynamics Module: User's Guide for Computer Program VEHDYN II, Technical Report SL-86-9, May 86; U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

5. Saxon, Nancy L., Bylsma, Wesley W., Simulation in the Role of Suspension Development, Proc. of SCSC of Summer Simulation Conference, La Jolla, July 1995.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Tyrone Bonte from Cadillac Gage Textron Inc. for both the reduction of the laboratory test data and his practical suggestions on the technical content of this paper.

970388

Flow and Energy Pattern in Pyrotechnic Airbag Inflator-Canister System

G. S. Nusholtz
Chrysler Tech Center

D. Wang
Hohai Univ.

E. B. Wylie
University of Michigan

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The highly dynamic process in a pyrotechnic inflator and in a canister-airbag system was simulated by using two compressible gas thermal energy numerical models. First a 2-D model was used to simulate flow through the inflator porous media; then results from the first model were used as input to a second 2-D model to compute pressure, temperature and flow patterns in the airbag. Results show a complete picture of the dynamics of the airbag inflator - canister system during deployment.

INTRODUCTION

During airbag deployment hot gas under high pressure, produced by burning fuel in the center of the inflator, penetrates into the surrounding compound ceramic - metal porous media, exchanging heat with the matrix of the porous media, and exits the discharge holes positioned on the perimeter of the inflator. This process plays a significant role in the amount of thermal energy contained in the gas exiting from the inflator to the airbag, and thus affects the success of the deployment of the airbag. Two numerical models were used to simulate the fundamental transient thermal energy features of the inflator - canister - airbag system during deployment. The results from the two models reveal: (1) the thermal energy distribution among porous media, inflator wall and the gas, (2) the instantaneous pressure and temperature distributions in the ceramic - metal porous media, and (3) the temperature and pressure distributions and flow patterns in the canister and airbag. The aggressive nature of this highly dynamic process makes it technically difficult, if not impossible, to measure these data. Therefore, these numerical results, which are physically reasonable, and which satisfy fundamental principles of fluid mechanics and heat transfer, provide insight to the entire dynamic excitation and response.

In this paper, the fundamentals of the two models are first discussed, followed by an application to a passenger inflator-airbag system with some numerical results.

TWO DIMENSIONAL GAS THERMAL ENERGY MODEL IN THE POROUS MEDIA

BASIC EQUATIONS High temperature and pressure associated with gas generation and flow within the inflator is at a time scale of milliseconds in a spatial domain measured in centimeters. This process involves highly localized thermal non-equiliblium conditions between the gas and porous media. The conditions considered in the development of the equations of conservation of mass, momentum and energy are summarized as follows.

(1) A linear viscous loss term and a nonlinear inertial loss term have been added to the momentum equations.

(2) The heat exchange between porous media and gas follows a linear law, i.e., the heat transferred is directly media and the gas.

(3) Heat conduction of the gas is neglected in the energy equation, since a short transient process (order of 10^{-2} second) is considered in the present case.

(4) The porous media temperature is calculated by the first law of heat energy balance in the porous media only. Heat conduction in the porous media is neglected due to the same reason as stated in (3).

(5) The frame of the porous media is fixed, i.e., no deformation occurs during the transient process.

(6) There is no chemical reaction between the porous media and the gas. [Bachmat and Bear, 1986]

Under the above considerations a model for the 2-D transient compressible gas flow with heat transfer in porous media was derived by the control-volume method. The equations are summarized as follows, beginning with the energy equation, followed by the continuity, and x- and y-momentum equations:

$$L_0 = (u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + \frac{\partial p}{\partial t}) - c^2(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + \frac{\partial \rho}{\partial t}) - (\kappa - 1)[\rho Q + \frac{\rho g}{K}\beta(u^2+v^2) + \rho R_{pm}\beta^2(u^2+v^2)^{\frac{3}{2}}] = 0 \quad (1)$$

$$L_1 = (u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + \frac{\partial \rho}{\partial t}) + \rho(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = 0 \quad (2)$$

$$L_2 = (u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{\partial u}{\partial t}) + \frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{g}{K}\beta u + R_{pm}\beta^2 u\sqrt{u^2+v^2} = 0 \quad (3)$$

$$L_3 = (u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{\partial v}{\partial t}) + \frac{1}{\rho}\frac{\partial p}{\partial y} + \frac{g}{K}\beta v + R_{pm}\beta^2 v\sqrt{u^2+v^2} = 0 \quad (4)$$

The set of equations is completed by including the equality (Wang and Wylie, 1995)

$$L_4 = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} - (\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) = 0 \quad (5)$$

in which p is the pore-volume averaged pressure, u, v are the area-averaged pore velocities(in x-direction and y-direction), K and R_{pm} are the hydraulic conductivity and non-linear resistance coefficient of the porous media, respectively, ρ is density of the gas, c is the isentropic wave speed in the gas, κ is the specific heat ratio, Q is the heat exchange rate, and β is the ratio of pore cross-sectional area to total cross-sectional area.

In addition to the conservation laws a relationship is needed to describe the heat transfer between the gas in the pores and the porous media. The porous media is a heat sink (or source) during the dynamic flow process. The amount of heat exchanged per unit mass and unit time is

$$Q = -\frac{1}{\eta}\frac{K_{ex}a_{sv}}{\rho}(T_g - T_{pm}) \quad (6)$$

in which K_{ex} is the convective heat transfer coefficient between the gas and porous media, heat/(unit area-s-degree), [Wakao and Kaguei, 1982], a_{sv} is the specific surface area of the porous media particles, area/unit volume, [Dullien, 1979, Wakao and Kaguei, 1982], η is the porosity of the porous media, and T_g and T_{pm} are the gas and solid material temperatures, respectively.

At an instant, in a control volume of gas and porous media matrix, δx by δy by unit depth, the time rate of increase of thermal energy in the porous media is just equal to the heat transferred from the gas to the porous media. This neglects heat conduction in the porous media during short transient flows.

$$(1-\eta)\rho_{pm}C_{ppm}\frac{\partial T_{pm}}{\partial t} = K_{ex}a_{sv}(T_g - T_{pm}) \quad (7)$$

The density and specific heat of the porous media are given by ρ_{pm} and C_{ppm}, respectively.

The literature provides relationships to assist in the evaluation of the various coefficients, K, R_{pm}, K_{ex} and a_{sv} in Eqs. (1) to (7) [Ergun, 1952, Sozen and Vafai, 1990].

$$K = \frac{\rho g K_g}{\mu_g} \quad (8)$$

$$R_{pm} = \frac{F\eta}{{K_g}^{1/2}} \quad (9)$$

$$K_g = \frac{\eta^3 d_p^2}{150(1-\eta)^2} \quad (10)$$

$$F = \frac{1.75}{\sqrt{150}\eta^{3/2}} \quad (11)$$

$$K_{ex} = 1.064C_{pg}\rho U(\frac{\mu_g C_{pg}}{k_g})^{-2/3}(\frac{\rho U d_p}{\mu_g})^{-0.41} \quad (12)$$

$$a_{sv} = \frac{6(1-\eta)}{d_p} \quad (13)$$

in which μ_g is the absolute viscosity of the gas, K_g is the permeability of porous media, F is a geometric factor, d_p is the diameter of the equivalent particle of the porous media, U is the total velocity of the gas flow, C_{pg} is the specific heat of the gas, and k_g is the thermal conductivity of the gas. According to Sozen and Vafai, 1990, the heat transfer coefficient in Eq. (12) is valid for Reynolds number, $\frac{\rho U d_p}{\mu_g}$, greater than 350.

Equations (1) to (5), with the substitution of Eq. (6), describe the dynamic gas flow and gas thermal energy response in the porous media matrix. Equation (7) describes the heat exchange between the gas and porous media matrix. They are solved with the characteristics-like method, while Eq. (7), which describes the heat exchange between the gas and porous media matrix, is solved with a finite-difference method. (Wylie and Streeter, 1993), using linear multipliers λ_1, λ_2, λ_3, and λ_4, is carried out in the following manner, $L_0 + \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 L_3 + \lambda_4 L_4 = 0$, which leads to a set of ordinary differential equations.

With a square computational grid $\Delta x = \Delta y$, and selected time step, Δt, integration of the compatibility equations along the path lines leads to the following set of equations.

$$A_{p1}p_P - c_E^2\rho_P = e_E \quad (14)$$
$$A_{p1}p_P + A_1 u_P + A_1 v_P - q_P\rho c^2\Delta t = e_1 \quad (15)$$
$$A_{p1}p_P + B_1 u_P - B_1 v_P + q_P\rho c^2\Delta t = e_2 \quad (16)$$
$$A_{p1}p_P - C_1 u_P - C_1 v_P - q_P\rho c^2\Delta t = e_3 \quad (17)$$
$$A_{p1}p_P - D_1 u_P + D_1 v_P + q_P\rho c^2\Delta t = e_4 \quad (18)$$

in which $q = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)$, and c assumes the local average value. The terms on the right side, e_E, e_1 to e_4, are determined by a bi-interpolation method based on the known values at points A', B', C', and D', Fig. 1. The parameters A_{p1}, A_1, B_1, C_1 and D_1 are a collection of variables. The simultaneous solution of Eqs. (14) to (17) provides variables p_P, u_P, v_P, q_P and ρ_P at each interior point. In order to save computer time and to avoid specification of conditions at singular points, the staggered grid, Fig. 1, is used in the model to discretize the computational domain (Wang and Wylie, 1995).

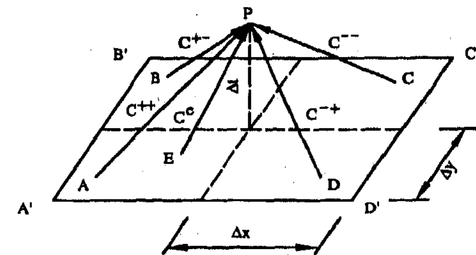


Fig. 1 Path lines in x-y-t space

Boundary Conditions for Gas Flow Computation, and Interface with Porous Media At each boundary, two of the four compatibility equations, Eqs. (15) to (18), are used together with the specified conditions of pressure and/or flow. Additionally, if there is flow out of the domain, Eq. (14) must be used, or if there is inflow, temperature (or a relationship for temperature) must be specified. Three types of boundary conditions which were used in this study are described.

Impermeable Boundary In this condition $q_P = 0$, and $u_P = 0$ at the right and left sides and $v_P = 0$ at the lower and upper limits. A local energy equilibrium condition is used here. This condition is approximated at an instant assuming that the conducted thermal energy flux at the computational node adjacent and interior to the boundary node is equal to the heat flux at the boundary. In numerical form this is

$$A_{P-2} q_{h_{P-2}} = A_P q_{h_P} \tag{19}$$

in which A_{P-2} and A_P are the areas through which heat flows at the node $2\Delta x$ interior to the boundary, and at the node on the boundary, respectively. $q_{h_{P-2}}$ is the unit heat flux through area A_{P-2}, ($q_{h_{P-2}} = -k_g \frac{\partial T_g}{\partial x}$, where k_g is the thermal conductivity of the gas). q_{h_P} is the heat flux at the wall ($q_{h_P} = A_P K_w (T_g - T_w)$, where K_w is the heat transfer coefficient between gas and wall, and T_w is the wall temperature). It is assumed that the thermal energy convected by the gas is small next to a solid wall. This is likely reasonable since the velocity normal to an impermeable wall must be small. With these definitions Eq. (19) becomes

$$A_{P-2} k_g \left(\frac{T_{g_{P-2}} - T_{g_P}}{2\Delta x}\right) = A_P K_w (T_{g_P} - T_w) \tag{20}$$

Specified Pressure and Temperature Boundary $p(t)$, $T_g(t)$ A left boundary position is used as an example of this type of boundary. The equation of state, $p_P = \rho_P R T_{gP}$, and Eqs. (17) and (18) are used to relate the horizontal velocity to pressure and temperature. It is also assumed that $v_P = 0$, and $q_P = 0$ at the left boundary.

Orifice Boundary The orifice equation

$$u_P = C_d \frac{A_h}{2\Delta y} \sqrt{\frac{2(p_P - p_0)}{\rho_P}} \tag{22}$$

is used together with three compatibility equations, $q_P = 0$, and $v_P = 0$ for a right boundary. A_h is the cross sectional area of the orifice and Δy is the grid spacing.

Equation (7) is discretized by an implicit scheme.

$$T_{pm_P}^{n+1} = \Big[\Delta t K_{ex} a_{sv} (T_{gP}^{n+1} + T_{gP}^{n-1}) + \big[(1-\eta)\rho_{pm} C_{ppm} - \Delta t K_{ex} a_{sv}\big] T_{pm_P}^{n-1}\Big] \Big/ \big[(1-\eta)\rho_{pm} C_{ppm} + \Delta t K_{ex} a_{sv}\big] \tag{23}$$

TWO DIMENSIONAL TRANSIENT COMPRESSIBLE GAS THERMAL ENERGY MODEL IN GAS DOMAIN

The time-varying thermal energy model includes the unsteady conservation laws of mass, momentum and energy. The governing equations are derived under certain fundamental assumptions.

(1) The compressible gas is considered non-viscous. The process is highly transient with high pressure, temperature and velocity changes, therefore, the viscous dissipation of momentum and energy is relatively small with respect to other energy components, and may

(2) A heat source or sink may or may not exist inside the the domain. The heat exchange could occur between gas and solid boundary, e.g., between gas and tank wall or between gas and canister wall, where a local energy equilibrium condition would be applied to supply a boundary condition.

(3) Heat conduction of the gas is neglected in the energy equation since the transient process is extremely rapid.

After some manipulation, and introduction of the isentropic wave speed, $c^2 = \kappa p/\rho$, the basic equations, ready for numerical transformation, are summarized as follows (Wylie and Wang, 1996).

$$L_0 = (u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + \frac{\partial p}{\partial t}) - c^2(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + \frac{\partial \rho}{\partial t}) - (\kappa - 1)\rho Q = 0 \tag{24}$$

$$L_1 = (u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + \frac{\partial \rho}{\partial t}) + \rho(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = 0 \tag{25}$$

$$L_2 = (u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{\partial u}{\partial t}) + \frac{1}{\rho}\frac{\partial p}{\partial x} = 0 \tag{26}$$

$$L_3 = (u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{\partial v}{\partial t}) + \frac{1}{\rho}\frac{\partial p}{\partial y} = 0 \tag{27}$$

in which p, u, v are the pressure and velocities(in x-direction and y-direction), respectively. The other symbols have the same meaning as in the first model. Equation (5) completes the set for tranformation in the characteristics - like method. A square grid, Fig. 1, is used and, as in the first model, integration of the transformed equations leads to the numerical model.

Prior to integration of the compatibility equations c was replaced by $\sqrt{\kappa p/\rho}$. This modification was found to be necessary to achieve better accuracy in thermodynamically controlled problems. Integration of the modified compatibility equations along the path lines leads to the following set of equations.

$$\ln\ p_P - \kappa \ln\ \rho_P = e_E \tag{28}$$

$$\ln\ p_P + \frac{\kappa}{c}u_P + \frac{\kappa}{c}v_P - q_P \kappa \Delta t = e_1 \tag{29}$$

$$\ln\ p_P + \frac{\kappa}{c}u_P - \frac{\kappa}{c}v_P + q_P \kappa \Delta t = e_2 \tag{30}$$

$$\ln\ p_P - \frac{\kappa}{c}u_P - \frac{\kappa}{c}v_P - q_P \kappa \Delta t = e_3 \tag{31}$$

$$\ln\ p_P - \frac{\kappa}{c}u_P + \frac{\kappa}{c}v_P + q_P \kappa \Delta t = e_4 \tag{32}$$

in which c assumes the local average value. The terms e_E, e_1, e_2, e_3, and e_4 are determined by the same way as those in the first model.

The treatment of boundary conditions is the same as that in the first model. An additional type of boundary, i.e., specified mass flow rate and temperature boundary $\dot{m}(t)$, $T_P(t)$, is described here. A left boundary position is used as an example of this type of boundary.

$$u_P = \frac{\dot{m}}{\rho_P A} = \frac{\dot{m} R T_{gP}}{p_P A} \tag{33}$$

The equation of state, $p_P = \rho_P R T_{gP}$, and Eqs. (31), (32) and (33) are used to relate the horizontal velocity to pressure and temperature. A is the area covered by each input point, and $\dot{m}$ is mass flow rate.

APPLICATION

The two models were applied to a typical inflator canster airbag system. The transient compressible gas thermal energy model through the porous media was first applied to the inflator with the measured pressure-time history and an estimated temperature-time history as the input data. The transient compressible gas thermal energy model in gaseous domain was then applied to the canister - airbag system using the mass flow rate and temperature time histories obtained from the computations carried out first in the inflator porous media. The operating environmental temperature in his modeling was 21.6 °C. The modeling processes and the esults are described separately in following subsections.

MODELING OF THERMAL - DYNAMIC PROCESS N INFLATOR The computational domain, which represents one quarter of a section through the inflator metal creen, is shown in Fig. 2. The spatial step was Δx = .0335 cm. The actual discharge holes were represented by hree slots, 30.48 cm long and 0.02205 cm wide. The total rea of the slots was equal to the total area of the discharge oles. The thermal conductivity and specific heat of the gas vere 0.0238 J/m-s-°K, 715.4 J/kg-°K; the specific heat and he density of the porous media were 474.4 J/kg-°K and 659.3 kg/m^3; the thermal conductivity, specific heat, and he density of the inflator wall were 220.9 J/m-s-°K, 897.3 /kg-°K, and 2720.0 kg/m^3; the porosity, η, of the porous nedia was 0.7287, and $\beta = \eta$. The hydraulic conductivy and non-linear resistance coefficient of the porous media vere estimated with the relations stated earlier, K = 0.32 n/s, R_{ms} = 1427.2 1/m.

The input boundary conditions were pressure and temerature at the fuel interface and are shown in Figs. 3a nd 3b. The boundary condition at the inflator wall was a o-flow condition, and a local heat energy equlibrium condiion was used here, except for the gas-flow holes. An orifice ormulae, with the 'choke' condition employed, was used at ach discharge hole.

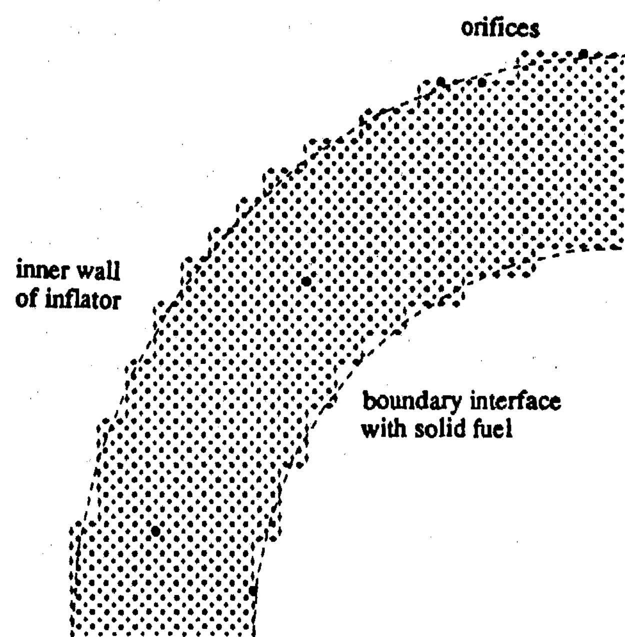


Fig. 2 Computational mesh in the porous metal screen

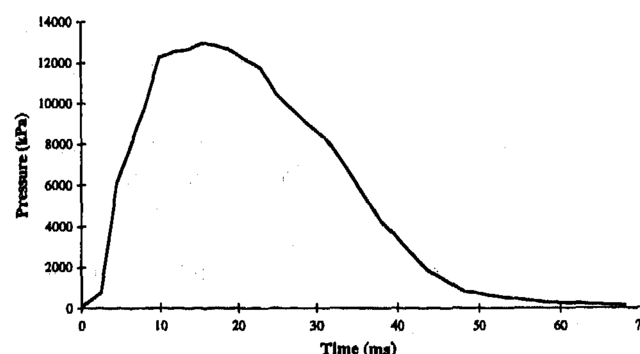


Fig. 3a Input pressure for 2-D porous model

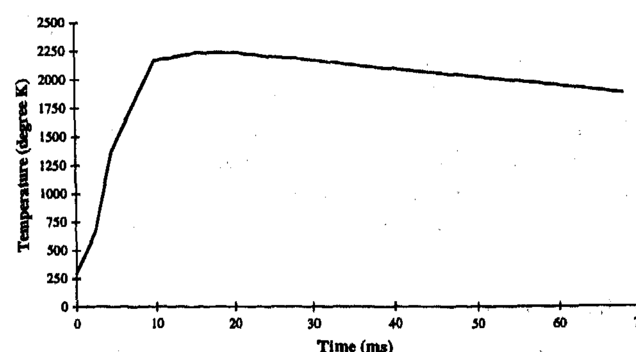


Fig. 3b Input temperature for 2-D porous model

Extensive computations were carried out to calibrate the heat transfer coefficients, K_{ex}, K_w, by using the data available. The values of K_{ex} and K_w were found as $K_{ex} = 7.653 \cdot (10)^6$ J/m^2-s-°K, $K_w = 1.526$ J/m^2-s-°K. These values were used for the final computation.

Graphic displays of the pressure contours, temperature contours, and flow pattern at 13 ms are shown in Figs. 4-6. The flow pattern is represented by the mass velocity, through the discharging holes, conservation of total mass (at the input and exiting through the discharge holes), the time histories of pressure at the input and the discharge holes, the time histories of the temperature of the gas and metal screen at the input and at the discharging holes, and lastly the time histories of total input thermal energy, the energy stored in metal screen, the energy going out with gas and the energy stored in the gas contained in metal screen, are shown in Fig. 7 to 11.

MODELING OF THERMAL - DYNAMIC PROCESS IN CANISTER - AIRBAG The semiaxes dimensions of the ellipse-shaped computational domain were 53.95 and 35.81 cm yielding a volume of the domain equal to the volume of the inflated airbag, 185 liters. The spatial step was Δx = 0.3173 cm. An actual vent hole of 5.08 cm in diameter was represented by two slots, 30.48 cm long and 0.3222 cm wide each. The total area of the slots was equal to the total area of the vent hole. The thermal conductivity and specific heat of the gas were 0.0238 J/m-s-°K, 715.4 J/kg-°K, and the heat exchange coefficient K_w at the canister and inflator walls was 0.171 J/m^2-s-°K.

The input boundary conditions were mass flow rate and temperature, which were obtained from the modeling of the thermal - dynamic process in inflator. They are shown in Fig. 7 by a dashed line and in Fig. 10 by a short dashed line indicated by 'at exit hole (gas)'. The boundary condition at the inflator wall is a no-flow condition, and a local heat energy equlibrium condition was used here, except for the gas-flow holes. An orifice formulae, with the 'choke' condition employed, was used at vent hole.

Two cases are presented, the first has no heat exchange between gas and solid walls, and the second includes heat exchange. The results for the case of no-heat exchange between gas and solid walls, $K_w = 0.0$, are shown in Figs. 12 to 16. Figure 15 provides a time history of the mass balance by showing the accumulated mass inflow at each instant, the instantaneous stored mass, the accumulated mass outflow, and the sum of the latter two. Figure 16 provides a similar time history of the energy at each instant. The results for the case with heat exchange between gas and solid canister and inflator walls show similar behavior as those in the case with no heat exchange. Only the pressure and temperature contours are shown in Figs. 17 and 18. Comparisons between the two cases are presented in Figs. 19 and 20.

CONCLUSIONS

The results from the two models show a complete picture of the dynamics of the compressible flow and thermal energy distribution of the airbag inflator-canister system during the deployment. Overall mass and energy balances are achieved in the whole system, which adds credibility and some substantiation to the numerical results.

Several important implications may be drawn from the results.

(1) The gas density change during this highly transient event plays a significant role in the overall response. Most of the gas passes directly to the discharge holes of the inflator, however, as much as holes merely packs up the void space during the early stages of the fuel burn. This is while the pressure is high and the matrix temperature is still low. Following the peak pressure, and as the temperature of the metal matrix rises the density of the gas drops and the gas exits the inflator.

(2) Two thirds of the thermal energy released from the burning fuel is absorbed by the metal screen during the first 60 - 80 ms period. This amount is the major source of thermal energy responsible for the temperature response on the outer wall of the inflator in the 200 seconds after the deployment.

(3) The pressure drop in much of the region of the metal screen is very small, about 1.0% of peak pressure, and the biggest pressure drop occurs near the discharge holes of the inflator. Even in that vicinity the drop is only about 5% of peak pressure. This suggests that the pressure measured through the tube connecting to the inflator wall may be a reasonable representation of the actual pressure produced by the burning fuel.

(4) There is a large temperature gradient across the porous metal, with the exit gas temperature less than one third of the temperature produced by the burning fuel ($\sim$ 537 °C at 22 ms). The gradient is the largest near the fuel-porous metal interface. This reflects the transient nature of the process and indicates most of the thermal energy absorbed is stored in the inner part of the porous metal screen.

(5) The pressure and temperature are almost uniformly distributed in the airbag at any instant. A big drop in pressure and temperature occurs in a very small region next to the vent hole due to the high discharge velocity.

(6) The maximum pressure is about 69 kPa in the bag and 52 kPa at the vent hole in the non-heat-exchange case, and about 67 kPa in the bag and 51 kPa at the vent hole in the heat exchange case. The pressures are higher than might be expected inasmuch as gas leakage through the bag wall fabric and the canister was not included, nor was any energy consumed in expanding the airbag. The maximum temperature is about 374 °C in the bag and 339 °C at the vent hole in the no-heat-exchange case, and about 357 °C in the bag and 328 °C at the vent hole in heat-exchange case. The amount of heat absorbed by the canister and inflator wall from the hot gas during the deployment is less than 3 kJ, which may raise the temperature of the canister wall about 3 °C.

(7) The obvious temperature gradient along the canister wall, shown in the temperature contour plot, in the case

of having heat exchange, compared with the temperature contour in the case of no heat exchange, indicates the effect of the heat transfer between gas and canister and infaltor wall.

REFERENCES

[1] Bachmat, Y., and Bear, J., "Macroscopic modelling of transport in porous media. I: The continium approach," *Transport in Porous Media* vol. 1 1986, pp. 213-240.

[2] Bachmat, Y., and Bear, J., "Macroscopic modelling of transport in porous media. II: Applications to mass, momentum and energy transport," *Transport in Porous Media* vol. 1 1986, pp. 241-269.

[3] Dullien, F. A. L., "Porous media fluid transport and pore structure," Academic Press, New York, 1979.

[4] Du Plesiss, J. P., and Masliyah, J. H., "Mathematical modelling of flow through consolidated isotropic porous media," *Transport in Porous Media* vol. 3 1988, pp. 145-161.

[5] Du Plesiss, J. P., and Masliyah, J. H., "Flow through isotropic granular porous media," *Transport in Porous Media* vol. 3 1991, pp. 207-221.

[6] Ergun. S., "Fluid flow through packed columns," *Chemical Eng. Progress*, Vol. 48, 1952, pp. 89-94.

[7] Sozen, M., and Vafai, K., "Analysis of the non-thermal equilibrium condensing flow of a gas through a packed bed," *Int. J. of Heat and Mass Transfer* vol. 33 No. 6, 1990, pp. 1247-1261.

[8] Sozen, M., and Vafai, K., "Longitudinal heat dispersion in porous beds with real-gas flow," *Journal of Thermophysics and Heat Transfer* vol. 7 No. 1, 1993, pp. 153-157.

[9] Wakao, N., and Kaguei, S., "Heat and mass transfer in packed beds," Gordon and Breach Science Publishers, Inc. New York, 1982.

[10] Wang, D., and Wylie, E. B., "Transient 2-D compressible flow with heat transfer," FED - Vol. 208, Forum on Fluid Transients, ASME, IMECE, San Francisco, Nov. 1995, pp. 91-99.

[11] Wylie, E. B., and Streeter, V. L., "Fluid Transients in Systems," Prentice Hall, New Jersey, 1993, pp. 410 - 413.

[12] Wylie, B., and Wang, D., "Energy balance in transient compressible flow," Pressure Surges and Fluid Transients, BHR Group, Harrogate, UK, April, 1996, pp. 283-288.

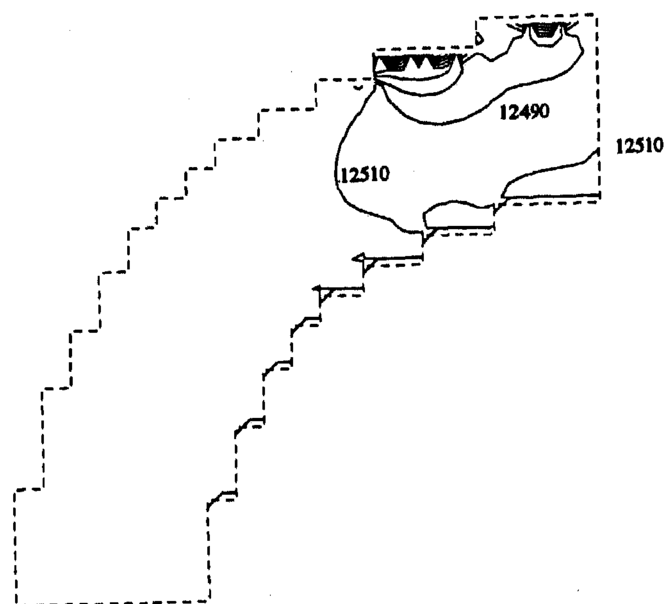


Fig. 4 Pressure contours in ceramic-metal screen at t=13 ms, $\Delta p = 20$ kPa

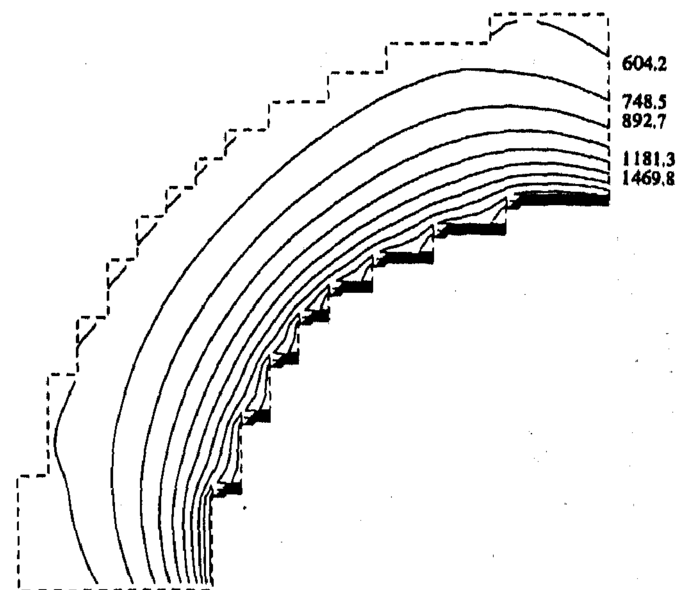


Fig. 5 Temperature contours in metal screen at t=13 ms, $\Delta T = 144.3°$C

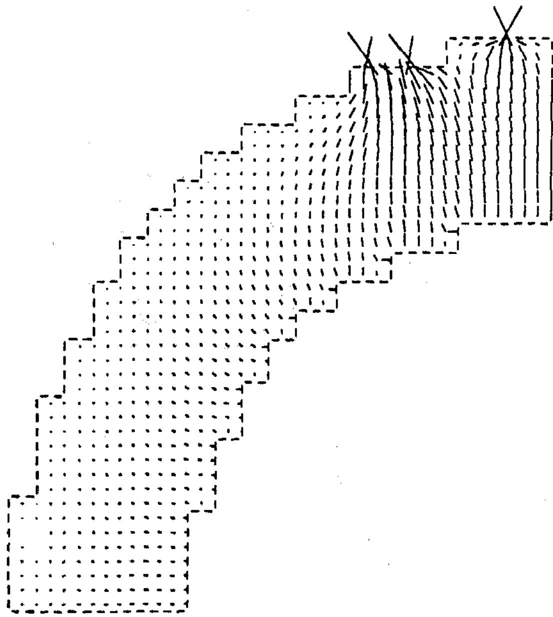


Fig. 6 Mass velocity vectors in metal screen at t=13 ms

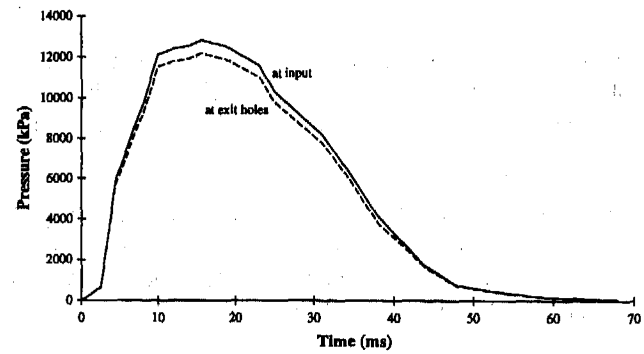


Fig. 9 Pressure time histories at input and at exit holes

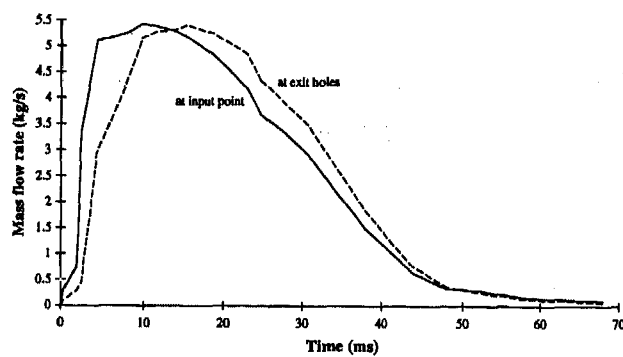


Fig. 7 Computed mass flow rate at input and out of exit holes

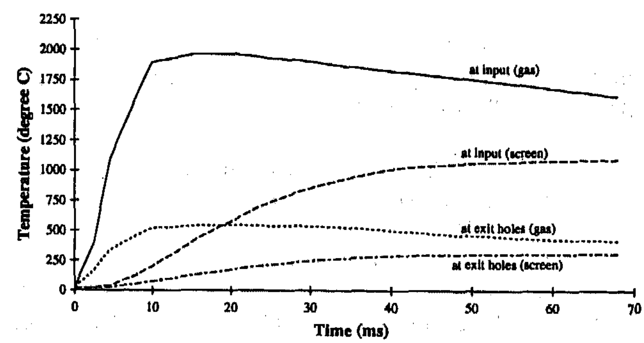


Fig. 10 Temperature time histories

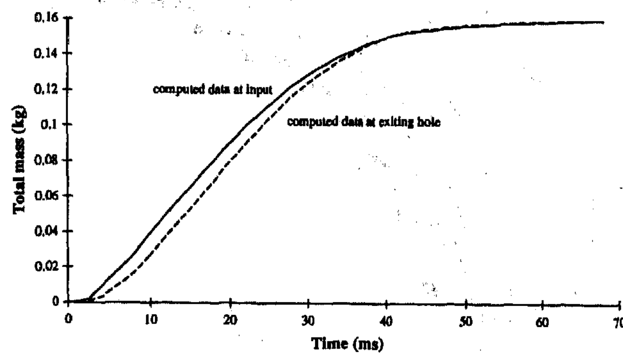


Fig. 8 Computed total mass at input and out of exit holes

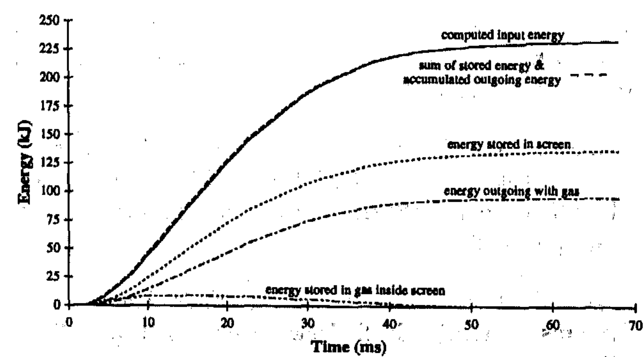


Fig. 11 Energy time histories

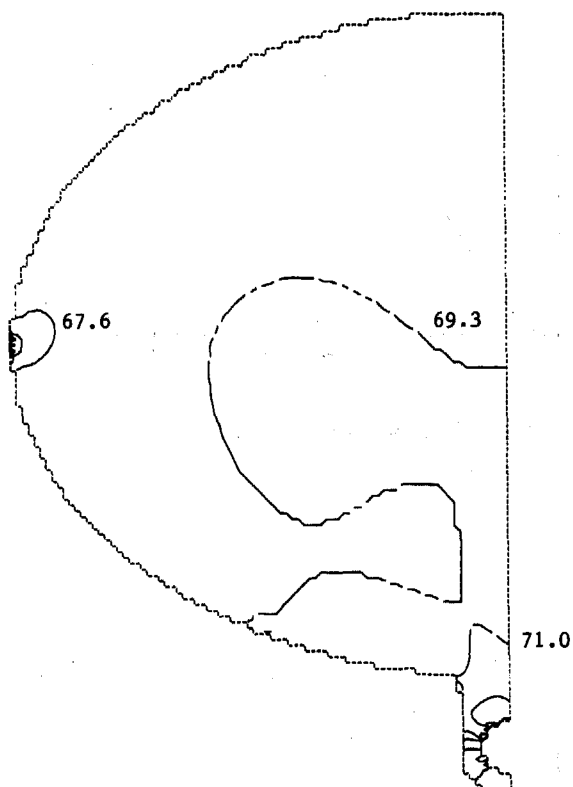


Fig. 12 Pressure contours at t=43 ms, $\Delta p = 0.7$ kPa (no heat exchange)

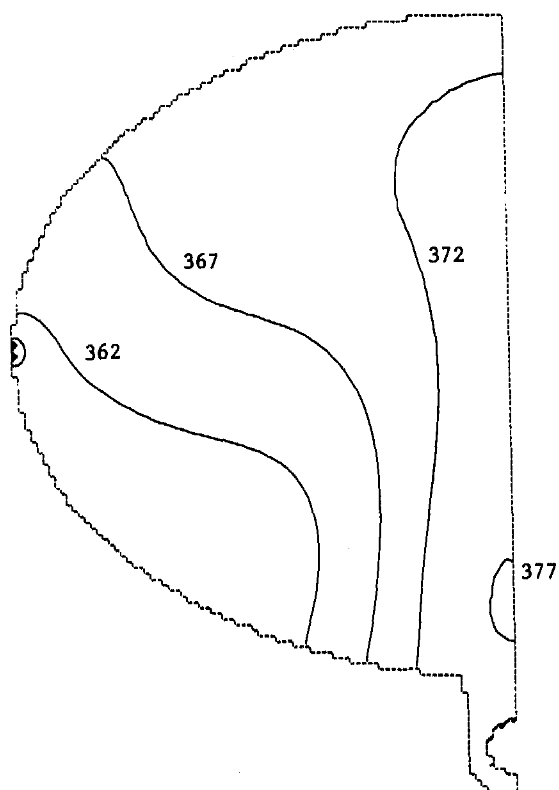


Fig. 13 Temperature contours at t=43 ms, $\Delta T = 5$ °C (no heat exchange)

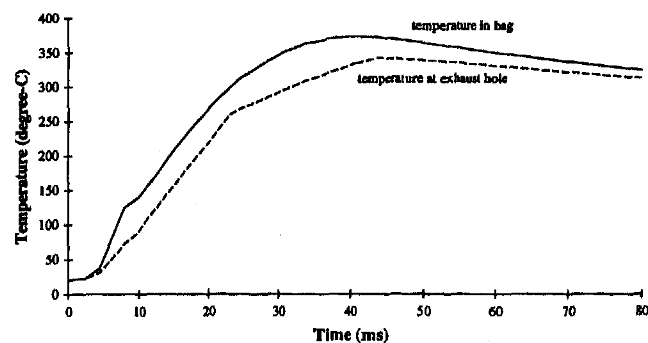


Fig. 14 Computed temperature time history at node A and at vent hole (no heat exchange)

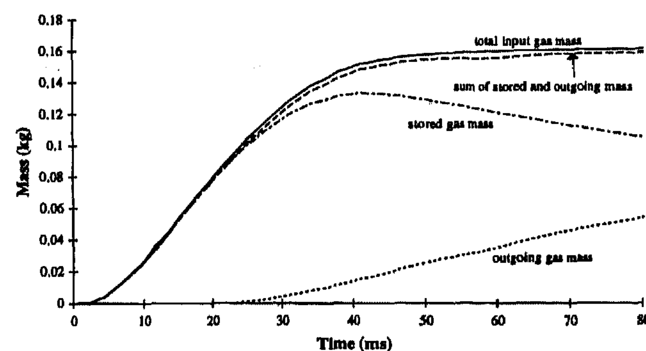


Fig. 15 Temporal mass variations (no heat exchange)

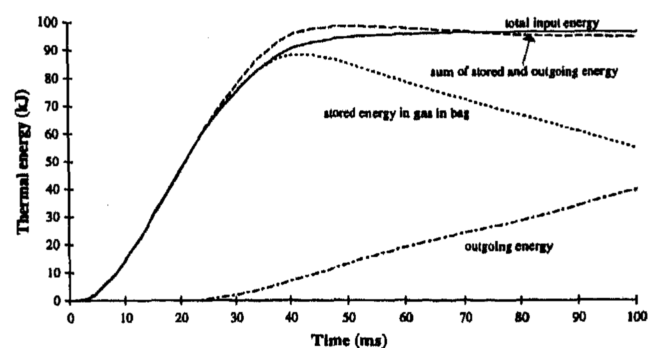


Fig. 16 Temporal energy variations (no heat exchange)

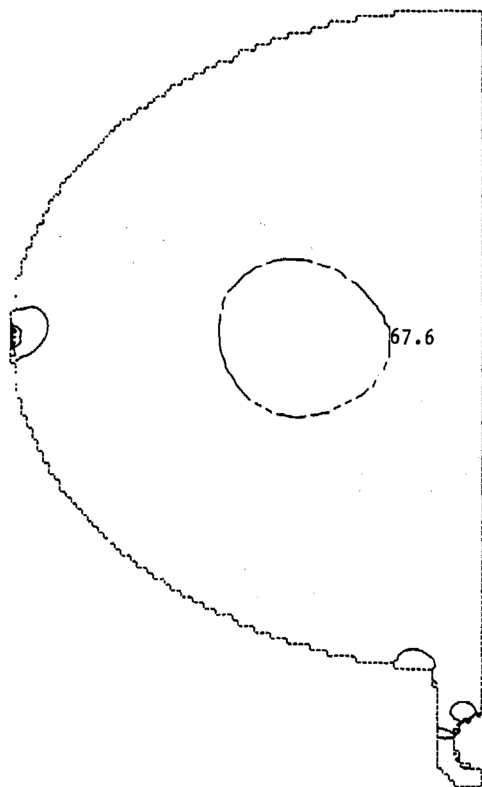


Fig. 17 Pressure contours at t=44 ms, Δp = 0.7 kPa (with heat exchange)

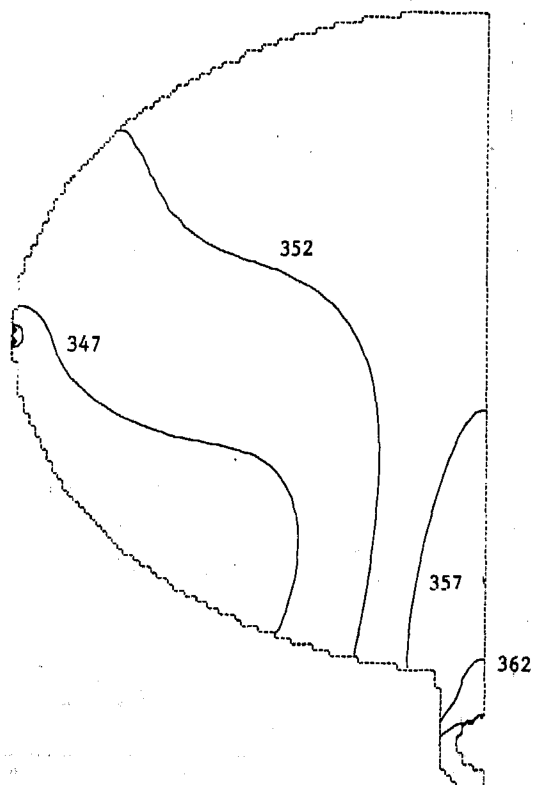


Fig. 18 Temperature contours at t=44 ms, ΔT = 5 °C (with heat exchange)

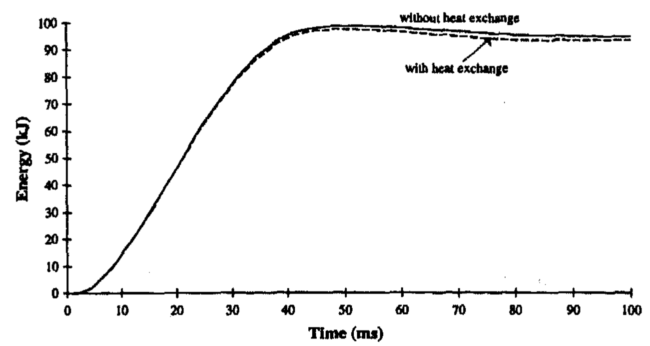


Fig. 19 Comparison of temporal energy variations in two cases

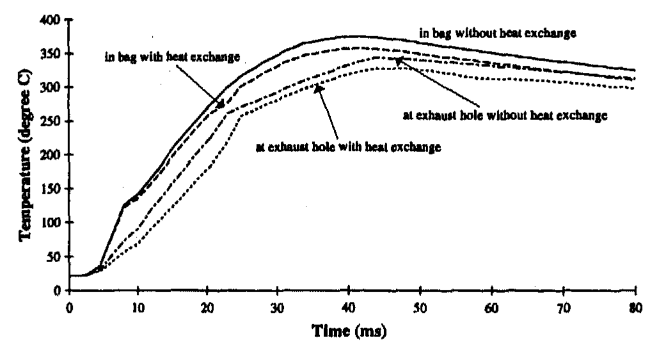


Fig. 20 Comparison of temporal temperature variations in two cases

970390

Cavitation During Head Impact

Guy S. Nusholtz
Chrysler Corp.

Lee G. Glascoe and E. Benjamin Wylie
University of Michigan

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The effects of stress in brain material was investigated with experimental and computational idealizations of the head. A water-filled cylinder impacted by a free traveling mass serves to give insight into what could happen to the brain during impact. Under an impact of sufficient velocity, cavitation can occur on the cylinder boundary opposite impact. Limited internal vaporization of the fluid may also occur during severe impact events. Cavitation occurred in these experiments at accelerations greater than 150 g's. Head forms of different sizing will experience an acceleration magnitude inversely proportional to the size difference to produce a similar pressure/cavitation response.

INTRODUCTION

The effect of stress in the brain as a mechanism of injury is rarely considered in head impacts, even though there have been some experimental studies indicating fluid vaporization (cavitation) through pressure limiting effects. These effects have been observed in impact tests using repressurized human cadavers and live anesthetized primates, where negative gage pressures of less than one atmosphere have not been observed unless skull fracture occurred [1, 2, 3]. Despite the possible fluid vaporization and no observed injury, the authors hypothesized that fluid vaporization limited the gage pressure to a negative one atmosphere. The theory was that the cavitation occurred either in the cerebral spinal fluid or between the dura and the skull, and was not significant enough to cause injury as geometric attenuation effectively reduces the force of cavity collapse.

To study this hypothesis, an experimental and numerical analysis of a simple model of the human head under impact was used [4, 5]. A water filled 14-cm diameter cylinder was struck by a 10-kg free-flying mass. Rigid-body acceleration-time-histories and the pressure at the fluid-cylinder interface were monitored during impact. Comparisons between the experimental results and the results of a two-dimensional computational model were made.

As a result of this work, the following view of head impact was developed: when the head receives a blow and a positive pressure develops under the point of impact, a small cavity, opposite the impact, can form between the skull and the dura. Vaporization limits the maximum pressure drop in the head to vapor pressure. Since the brain is nearly incompressible, the small amount of volume needed to create the cavity is accomplished by either pushing a very small amount of material down the foramen magnum or by skull deformation.

This leads to the following possibility for a negative gage pressure-limiting impact phenomena that could occur at the skull-dura interface. During severe impact to the human head, a greater tension would develop between the cerebral spinal fluid and the dura, than between the dura and the skull. The dura would then move with the brain cerebral spinal fluid. As the dura moves away from the skull, a small cavity (of thickness less than 10^{-5}m) could form representing a slight stretching between the skull and the dura of the periosteum. Although it may be possible that enough surface tension could develop between the skull and dura for a cavity to form on either side of the dura, this does not seem as likely as the above suggested mechanism. Therefore, it is this motion of the dura as coupled with the small fluid flow through the foramen magnum that can have a significant effect on the pressure response of the brain during impact. If the impact is severe enough, it may be possible to produce fluid vaporization and the associated violent cavity collapse. This violent cavity collapse, which could be a mechanism of injury, is strongly modified by the dura motion minimizing the amount of damage that could occur.

*This paper was previously published in part at the 1996 AGARD Conference.

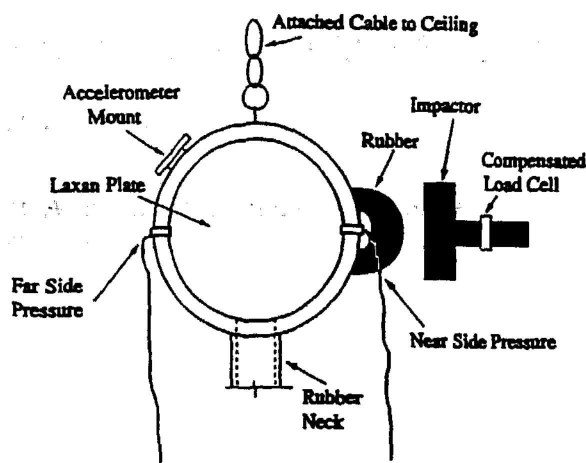


Figure 1: Experimental setup.

Recently, a follow up study indicated that under complex loading conditions, cavitation could occur in the brain material – not just at the boundary [6]. It was found that if the head underwent pre-impact acceleration immediately before the head strike, a vaporous region could form in the brain material and internal cavitation was likely. As a result, it was postulated that, in a vehicle environment, head impact could generate internal cavitation. Seeking further understanding, the same simplistic physical model of the head is used in this current study. In addition to cavities forming at the skull-dura boundary cavities in the brain material are addressed, with implications of potential cellular tissue damage. No attempt is made to model the anatomical structures of the head, the skull-dura boundary, the complex geometry of the head, head impact response, skull deformation, head angular motion, or brain shear. Instead, a simple physical model is used to generate test data. These data are modeled by a numerical simulator to gain insight into cavitation and into the general effect of pre-stressing, such as neck loading, on pressures in the brain material.

PHYSICAL MODEL

The physical model used is briefly outlined herein; it has been discussed in more detail in earlier papers [3, 4, 5, 6]. The simplified mechanical head form consisted of two parts: an internally polished aluminum cylinder and a rubber neck attached to the base of the cylinder, Figure 1. The cylinder was closed at each end with flat plates of 1.3 cm thick plexiglass, and for most of the tests, the plexiglass was reinforced with aluminum channel. To complete the head form the cylinder was connected by a rigid joint to a flexible viscoelastic rubber neck structure; fluid flow between the cylinder and the neck was through a two-centimeter diameter hole at the head neck interface. A nine-accelerometer array, used to measure three-dimensional motion (six degrees of freedom), was rigidly affixed to the aluminum cylinder and pressure transducers monitored the interface of the fluid and the cylinder.

The fluid tap water used to fill the system was boiled for several hours and allowed to return to room temperature. Before filling the cylinder with the fluid the air in the cylinder was evacuated and a partial vacuum was maintained. The vacuum was used to draw the fluid through a valve into the neck, then through the hole in the base of the cylinder. A vacuum was maintained until the cylinder was filled and no visible cavities or vacant spaces could be observed. Before each test series the mechanical head form was impacted several times and inspected to ensure that no air or cavities were visible in the cylinder.

The impact was delivered through the center of mass of the cylinder by a 10 kg free-traveling mass fitted with an inertial-compensated load cell assembly. The impacting surface consisted of a flat rigid 5-cm diameter disk [1, 2, 3]. To bring the contact force time-history closer to that of a head impact, a neoprene rubber hemisphere was fixed to the cylinder so that loading impact to the cylinder came through the neoprene hemisphere. The impact mass velocity was measured by timing the pulses from a magnetic probe which sensed the motion of targets in the mass.

COMPUTATIONAL MODEL

Two finite difference models were used to simulate the impact response of the fluid-filled cylinder, a continuum model and a discontinuous or discrete model. For both models it was assumed that boiling the tap water removed the free gas. Therefore, even though tap water had a significant number of nucleation sites to allow for fluid vaporization in the impacted cylinder and plastic neck, there was no free gas in the water. It was also assumed that even though there was no visible air in the physical model, it was impossible to remove all of the air at the boundaries of the water and the cylinder walls. The air at these boundaries was modeled as a continuous thin layer less than 10^{-6}m thick. Finally, it was assumed that the flow velocities through the neck were proportional to the force acting on the fluid in the neck.

CONTINUUM MODEL. The continuum model is based upon the two-dimensional unsteady equations of motion and continuity in a compressible fluid continuum. The procedure is based upon Henke's method [7] that was applied to fluids by Wylie [8, 9]. The linearized version of the more general compressible flow equations that describe pressure waves in a two-dimensional continuum provide the most simple case to develop the method. A combination of equations for the conservation of mass and momentum in the longitudinal and lateral directions are linearized into a set of four characteristic-like equations. The numerical procedure is developed by integrating these equations along paths in space and time, Figure 2. In this linearized model with constant isentropic wavespeed, $\Delta x = \Delta y$ in the computational grid.

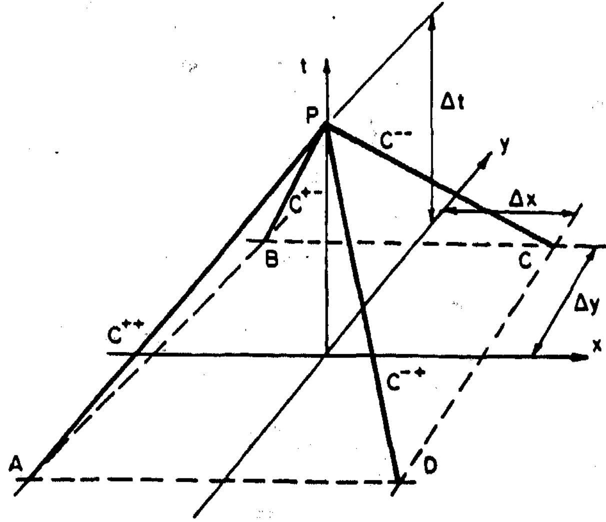


Figure 2: Staggered grid for continuum model.

The following set of equations result:

$$\frac{p_P}{\rho c} + u_P + v_P - q_P \Delta x = e_1 \quad (1)$$

$$\frac{p_P}{\rho c} + u_P - v_P + q_P \Delta x = e_2 \quad (2)$$

$$\frac{p_P}{\rho c} - u_P - v_P - q_P \Delta x = e_3 \quad (3)$$

$$\frac{p_P}{\rho c} - u_P + v_P + q_P \Delta x = e_4 \quad (4)$$

in which

$$e_1 = \frac{p_A}{\rho c} + u_A + v_A + q_A \Delta x \quad (5)$$

$$e_2 = \frac{p_B}{\rho c} + u_B - v_B - q_B \Delta x \quad (6)$$

$$e_3 = \frac{p_C}{\rho c} - u_C - v_C + q_C \Delta x \quad (7)$$

$$e_4 = \frac{p_D}{\rho c} - u_D + v_D - q_D \Delta x \quad (8)$$

where for point P, p_P is the absolute pressure, u_P is the x-direction velocity, v_P is the y-direction velocity, and $q_P = (\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x})/2$. Equations 1 through 4 are applied along lines C^{++}, C^{+-}, C^{-+}, and C^{--}, respectively, Figure 2. This two-dimensional spatial grid in the $x - y$ plane with time represented in the third dimension was used to simulate the plexiglass cylinder. Handling of the vaporization-free gas at the water-cylinder (or skull-dura) interface requires special consideration.

At the cylinder's perimeter the pressure drop is limited to vapor pressure as pressures below this are not physically possible. Under these conditions, in the absence of free gas, the fluid is assumed to vaporize, the pressure is fixed to vapor pressure, and a cavity is permitted to develop. The cavity grows while the liquid moves away form the boundary, and collapses as the liquid moves back to the boundary. In the model a small amount of air is assumed to exist at the wall and isothermal expansion and compression of the gas layer is assumed. The ideal gas law is used to relate the cavity volume to pressure. The gas law may be written as

$$V_{air} = \frac{C_4}{p_p - p_v} \quad (9)$$

where V_{air} is the volume of free gas at any boundary node; p_p is the absolute pressure at that node; p_v is the vapor pressure of water; and C_4 is the product of the gas constant, absolute temperature, and the mass of free gas at the node. The volume of air per unit depth at each boundary node is equal to the air thickness multiplied by two distance intervals.

A boundary condition is used at the neck node to model the effect of the rubber neck on the dynamic response of the cylinder. With a relatively low wave speed in the rubber neck, it was assumed that the neck was long enough so that the far end condition did not influence the dynamic response at the cylinder-neck connection. This meant that the neck acted like an infinitely long tube. The additional equation necessary to interact with the C^{-+} and C^{--} equations was provided by the following one-dimensional compatibility equation for the neck

$$p_P = p_i - \rho\, c_i\, (v_P - v_i) \quad (10)$$

In equation 10 $p_i = 1$ atmosphere (absolute), $v_i = 0$, and c_i is the wave speed in the rubber neck. The variable q_P was set to zero, enabling a solution for the variables p_P, u_P, and v_P at each computational time step. Figure 3 illustrates a square grid of latticework elements. The compatibility equations used in the method of characteristics, without friction produce the following equations for point P in Figure 3.

DISCONTINUOUS MODEL. Because the first method of modeling requires a continuum it can be used to estimate cavitation only at the boundaries but not inside the fluid. This is a result of the discontinuity caused by the cavities formed inside the fluid media. To address this problem a second method is used. A description of this method and its equivalence to the first method are found in [9, 10]. A brief description of this approach as it relates to the cylinder follows.

The numerical procedure to analyze two-dimensional transients is essentially the same as in one-dimensional problems. At internal nodes, or line element intersection points, a common pressure is assumed at each instant and the mass conservation law must be satisfied. A scaling relationship is needed between the acoustic wavespeed, a, in the line element and the acoustic velocity, c, in the two-dimensional domain [9].

$$a = \sqrt{2}c \quad (11)$$

$$p_P = p_A - a\rho(u_2 - u_1) \quad (12)$$

$$p_P = p_B - a\rho(v_2 - v_1) \quad (13)$$

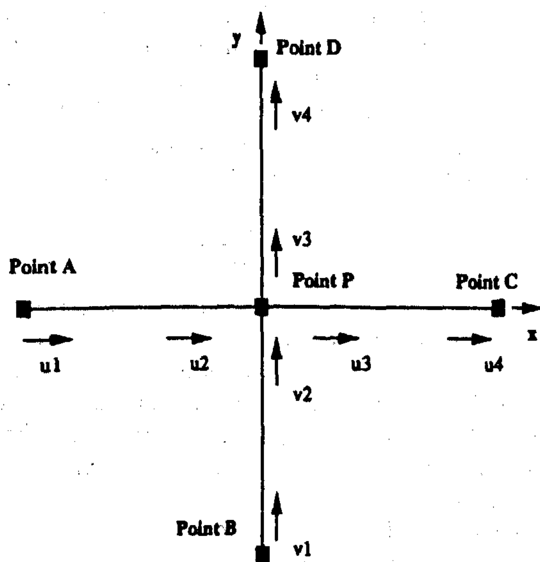


Figure 3: Latticework elements and velocities for discrete model.

$$p_P = p_C + a\rho(u_3 - u_4) \tag{14}$$

$$p_P = p_D + a\rho(v_3 - v_4) \tag{15}$$

where g is gravity, p_P is the pressure at point P, u_1 is the horizontal velocity just to the right of point A, v_2 is the vertical velocity just below point P, etc. Conservation of mass at any instant provides the fifth equation to solve for the five variables $(p_P, u_2, u_3, v_2, v_3)$.

With the second model, when low pressures develop in a region of the domain, one of two procedures is used depending on whether computation is at a boundary or an interior point. At an interior point, the minimum allowable pressure is vapor pressure, and the liquid vaporizes to form a vapor cavity which grows and collapses in response to the fluid dynamics. When the transient would have the pressure drop below vapor pressure, the pressure is set to vapor pressure, and the four velocities at the node, Figure 3, are computed. A time-dependent cavity volume per unit depth, V_V is then calculated from continuity

$$\frac{dV_V}{dt} = (v_{out} - v_{in})\Delta x \tag{16}$$

in which v_{in} is the net velocity to the left or below the volume, and v_{out} is the net velocity to the right or above the cavity. The size of the cavity at each point grows and collapses in response to the dynamic behavior. At the boundary, the procedure is the same as at the boundary in the first method, that is, isothermal expansion and compression of the gas layer is assumed.

RESULTS FROM EXPERIMENTAL AND CONTINUUM MODELS

The parameters used in the model were those of water under standard conditions. The density of water, ρ, was 1000 kg/m^3, the Bulk Modulus was 2.05 GPa; the wave speed, c, was 1435 m/s; the step distance interval, Δx, was 0.00318 m; and the time step, Δt, was $2.22(10^{-6})$ sec. Vapor pressure was set at 4550 Pa (absolute). The acceleration recorded from the physical model was input to the computer model to generate pressure-time histories at various points in the cylinder. As mentioned before, determined efforts were made to remove all of the air from the physical model, however, it is likely that a microscopic layer existed at the fluid/cylinder boundary. Therefore, on the perimeter an initial thickness of air of 10^{-6}m was used to model this effect.

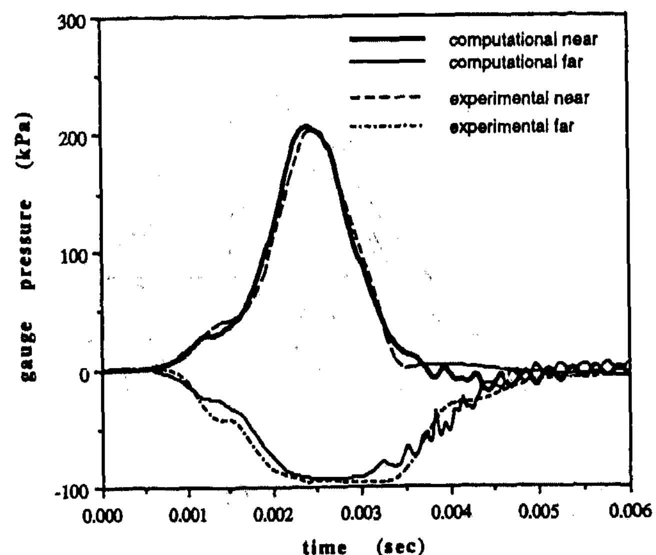


Figure 4: Low acceleration pressure response.

The initial conditions for the model were set such that there was no velocity in the cylinder and the pressure everywhere was atmospheric. The right and left boundaries were then accelerated according to a digitized acceleration from a specific experimental run. Two runs are evaluated here, a low acceleration (5 m/s impact velocity) and a high acceleration (8 m/s impact velocity). Both are analyzed in the model with vaporization at the boundary. The points presented for comparison are the near side and the far side pressure-time histories of the extreme left and extreme right boundaries.

LOW ACCELERATION (5 M/S IMPACT VELOCITY). An impact test was considered appropriate for this study if: 1) the impact velocity was within 0.1 m/s of 5 m/s; 2) no visible air could be seen in the cylinder before and after impact; and 3) the peak angular accelerations during impact were less than 100 rad/s^2. In the set of impact tests that met the above requirements the angular and linear accelerations and velocities were similar and reasonably close to that of a rigid body. Peak linear accelerations were within 5 percent of 190 g's and there was less than 0.01 rad of rotation. Therefore, for this set of tests it was assumed that: the motion of the cylinder was one-dimensional, the acceleration of each point on the alu-

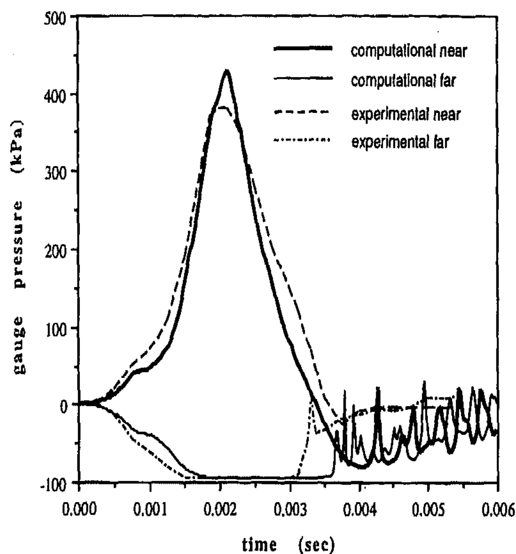


Figure 5: High acceleration pressure response.

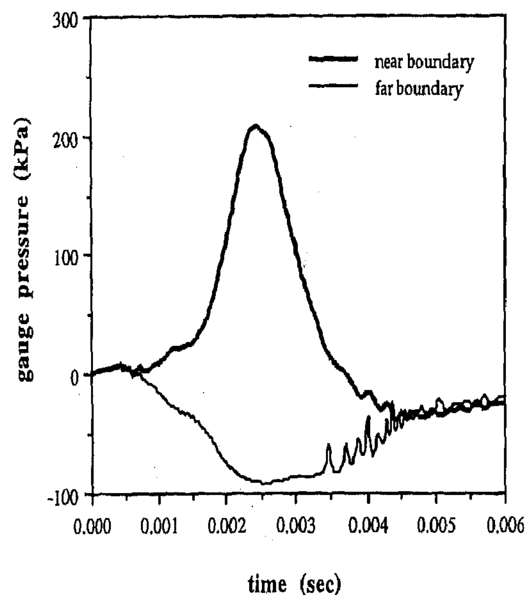


Figure 6: Low acceleration pressure response with 20-g neck acceleration.

minum cylinder was equal to the tangential acceleration obtained from the nine accelerometer array, and that each impact represented a repeatable event. In simulating this set of tests rotation did not need to be considered in the mathematical model. The near side and far side pressures from these tests also represented repeatable phenomena. As can be seen from Figure 4 the computational results compare very well with the experimental results. Both the experimental and computational results show the effect of the vaporization/free gas boundary though the experimental results show a slightly sharper drop to vapor pressure.

HIGH ACCELERATION (8 M/S IMPACT VELOCITY). The requirements for the 8 m/s tests differ from the 5 m/s tests only in impact velocity, that is, the impact velocity had to be within 0.1 m/s of 8 m/s. The motion of these tests was also one-dimensional without rotation. Although these tests were repeatable in terms of acceleration and near side pressures, the far side pressures were not. The differences occur at a time near the end of impact as the negative pressure starts to rise. Additionally, the magnitude of the pressure overshoot varies from test to test.

As is the case for the 5 m/s run, the minimum pressure on the far side for the 8 m/s run is limited to vapor pressure, Figure 5. In the model, low pressure lasted longer implying a larger vapor cavity, and the subsequent pressure rise is due to the vapor cavity's violent collapse. The near side pressure peak of the computational model is somewhat higher (by about 15 percent) than the experimental results. The computational results also show a more narrow peak. For the most part, however, the experimental results and the computational results for the near side compare reasonably well. The far side experimental data show a smaller vapor cavity growth than the computational results as the pressure is at vapor pressure for about 0.4 ms less in duration. However, the experimental runs gave some variations as to the exact time of vapor cavity collapse (6). Some of this variation may be experimental in nature as the sampling rate may be low enough (10,000 Hz) that the exact magnitude of the instantaneous pressure spike after cavity collapse cannot be obtained.

DISCRETE MODEL RESULTS

If it is assumed that the numerical model reliably reflects the response of the physical model at the measured points, then it is reasonable to use the model to estimate the response at points where measurements cannot be made. In addition, it is also reasonable to evaluate impact conditions in the numerical model that were not tested in the physical model. In this study the physical model was struck by a free traveling mass and the pressure response at the cylinder fluid boundary was measured. All cavitation occurred at the fluid cylinder boundary. To build on these results we used the computer model to estimate the cavitation response inside the fluid as a result of an impact similar to that of the test, and also a hypothesized neck loading.

The impacts of the physical model and simulation of these impacts by the computational models represent a stationary-test subject struck by a moving mass. However, in most automobile crashes the occupant is moving and strikes a moving object. The motion normally sub-

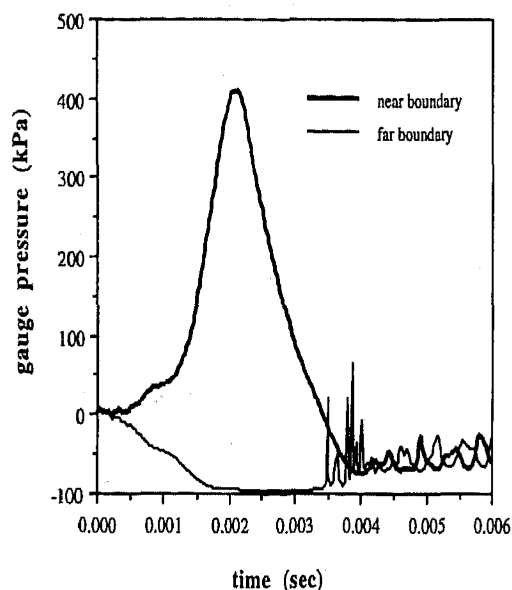


Figure 7: High acceleration pressure response with 20-g neck acceleration.

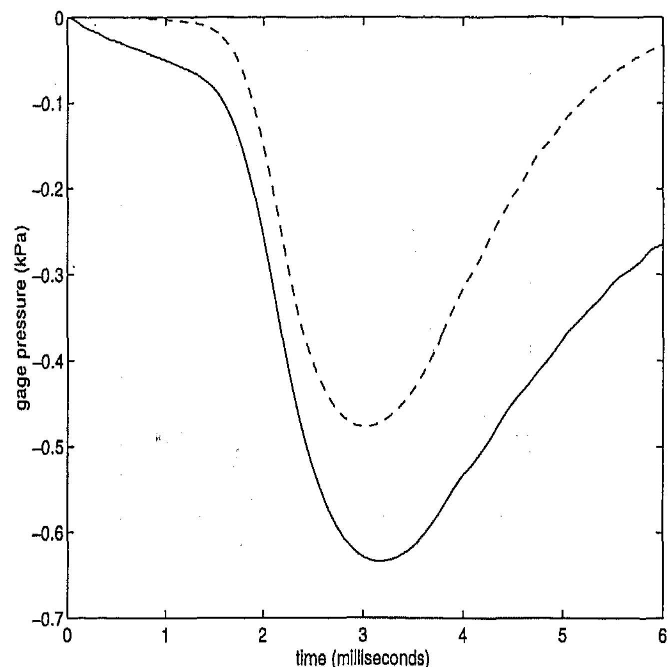


Figure 8: Fluid displacement with time for the continuum model (no internal vaporization, dashed line) and for the discrete model (with internal vaporization and neck loading, solid line) at the high acceleration.

jects the occupant to a force through the neck. This force can produce an acceleration perpendicular to the impact load and is typically of longer duration and lower magnitude. Therefore, the following results represent impacts similar to those above with the addition of a constant acceleration from the neck, in this case the acceleration is about 20 g's. Figures 6 and 7 show the near and far side boundary pressure histories of the low and high acceleration impacts, respectively, with a 20-g upward acceleration at the neck. When comparing Figures 6 and 7 with Figures 4 and 5, respectively, it can be seen that the additional stress incurred by the 20 g's results in a more violent cavity collapse. Greater internal cavitation is predicted by the model under the 20-g neck acceleration. The second model is able to predict an internal vaporous zone with small cavities (on the order of 10^{-6} m in diameter) at a number of internal nodes. Figure 12 shows the internal response of the fluid when the 20-g neck acceleration is added to the acceleration from the impact event. The figure shows the extent of the vaporous zone (the white region) in the cylinder at a time of 3.1 ms. The computational model demonstrates that the higher the acceleration from impact the greater the chance of cavitation in the fluid material. However, as illustrated by Figure 12, even at the low velocity the potential for internal cavitation exists when additional loading is included. By providing a constant vertical acceleration, in addition to the time-varying horizontal acceleration, the vector addition results in a larger resultant acceleration on the container. During peak acceleration the change is small. However, during low horizontal acceleration periods the prestressed vertical loading changes the magnitude and direction of the resultant significantly. This change in the resultant acceleration biases the vaporous zone towards the top of the cylinder. This is particularly evident for this case since the onset of vaporization is well before the time of peak horizontal acceleration, and the largest cavity size is beyond the time of peak horizontal acceleration. The numerical model also shows that the vertical acceleration on the head produces a higher net mass outflow at the neck, Figure 8 (less than 1 mm of fluid displacement), thus contributing to lower overall internal pressure and corresponding vaporization. Therefore, the amount of vaporization associated with this acceleration profile is greater than the amount of vaporization if the same resultant acceleration was entirely in the horizontal direction.

DIMENSIONAL SCALING

It is possible to investigate the effect of cylinder size on cavitation response. In the computational model, the size of the cylinder was reduced by scaling all relevant dimensions (diameter, thickness, neck size, etc.) by a scaling factor. The pressures scale to the sizing ratio, i.e., a cylinder 1/5 the size of the original cylinder undergoes a pressure response 1/5 the size. This can be observed in Figure 9 when vapor pressure limitations are not included in the model. When vaporization is included in the model and when the larger cylinder experiences more vaporization than the smaller cylinder, the pressure observed on the impact side

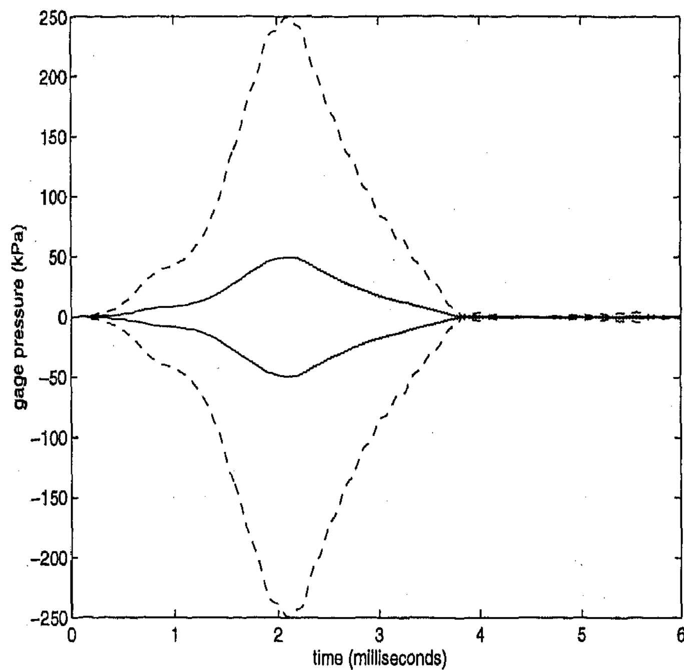


Figure 9: Pressure-time history for high acceleration when cylinder is reduced by a factor of five (solid line) and compared to standard size response (dashed line) when the pressure response is not limited to vapor pressure.

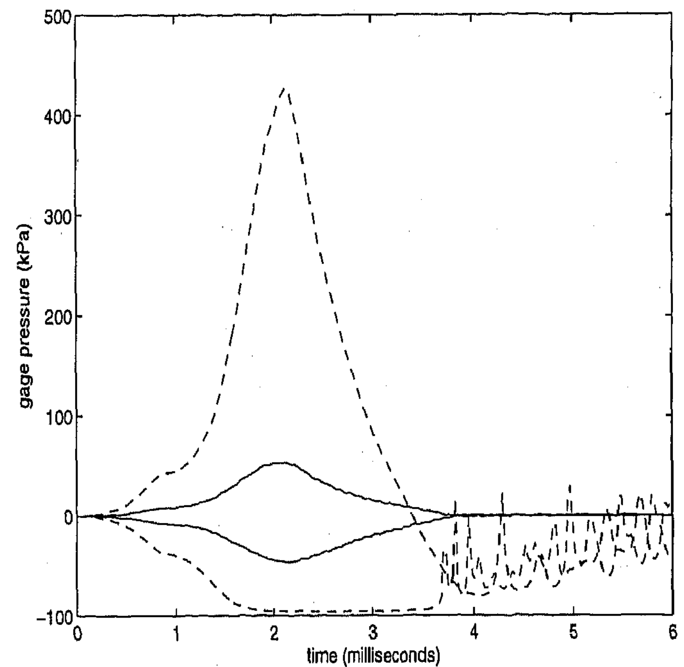


Figure 10: Pressure-time history for high acceleration when cylinder is reduced by a factor of five (solid line) and compared to standard size response (dashed line) when the pressure response is realistically limited to vapor pressure.

of the cylinder is much greater than the scaling ratio, Figure 10. This is because when the distal pressure response is limited to vapor pressure (as it must be to model reality) the impact pressure response must be greater than if there were no vaporization in order to conserve fluid momentum. The effectiveness of scaling directly by the size ratio is demonstrated when a cylinder 1/5 the size of the original is accelerated by a magnitude 5 times greater than the original acceleration, Figure 11. The pressure response for this case is nearly identical to that of the original cylinder undergoing the original acceleration. These results have relevance for comparing the impact response on primates to that of the impact response on humans. In an impact environment, pressure/cavitation response would remain approximately the same if the impact velocity is increased by the dimensional scaling factor. These results confirm what might be expected, that is, during an impact sufficient to produce cavitation for primate studies, greater acceleration of the impacted head form would occur.

DISCUSSION

Complex loading conditions are common in an automobile crash. In most automobile crashes the occupant is moving with the vehicle before it crashes. As the vehicle slows down the occupant keeps moving. Normally the head and torso will move differentially and a load will develop through the neck, resulting in angular as well as linear motion. If the head makes contact with a hard object during

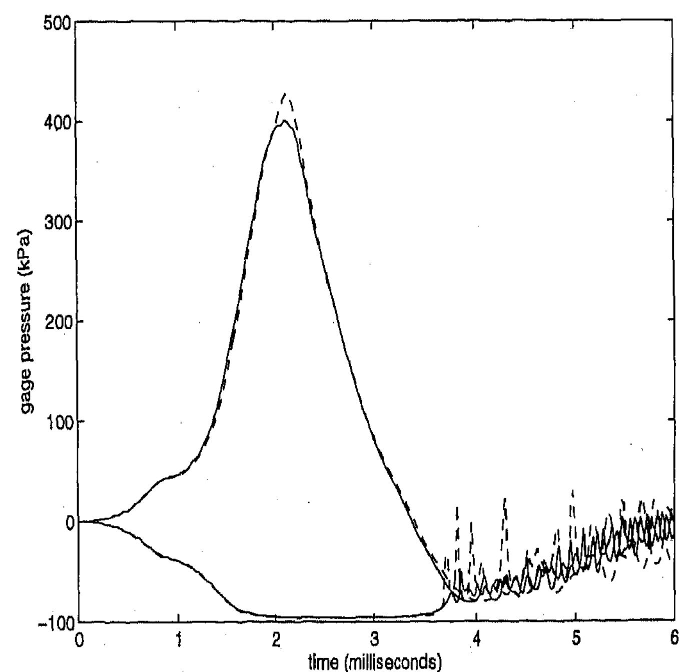


Figure 11: Pressure-time history for high acceleration when the original cylinder size is reduced by a factor of five and the original acceleration is increased by a factor of five (solid line). Also plotted is the pressure response under original 1:1 conditions (dashed line)

impact, the neck force produces an acceleration perpendicular to the impact load, creating a complex loading, that can affect the cavitation response. The influence is similar to what has been observed in low pressure transients in pipes, i.e., gravity may have a profound effect on cavitation during water-hammer events [9]. By changing the angle of the pipe with respect to the earth the cavitation response can be enhanced or eliminated. Therefore, the pre-stressing of the fluid by gravity before a water-hammer event affects the cavitation response.

A discussion of the pressure distribution in a closed container during acceleration may help in visualizing internal cavitation. For a constant horizontal acceleration the pressure pattern is:

$$p = p_0 + \rho k g x \tag{17}$$

where x is measured from the front of the container, p_0 is the reference pressure at the front, k is the number of g's of acceleration, and p is the absolute pressure. Low pressures causing vaporization are always initiated at the boundary. If the low pressure pulse is visualized to be initiated at the far side boundary ($x = D$), the magnitude of the pressure pulse causing vaporization is:

$$\Delta p = -p_v + p_0 + \rho k g D \tag{18}$$

where p_v is absolute vapor pressure and D is the container diameter. The pressure at the wave front is

$$p - \Delta p = p_v - \rho k g(D - x) \tag{19}$$

That is, for any x the pressure is trying to reach a value less than vapor pressure, a necessary condition for internal cavitation. It is the combination of the sloped initial gradient and the limitation in the pressure excursion to not less than vapor pressure that creates internal cavitation. Invariably, the largest cavities develop at the boundary. However, distributed internal vaporization can occur and will be more severe for high acceleration.

Vertical accelerations on the head may be more severe since the additional acceleration may create a high pressure in the vicinity of the neck, which would force material down the foramen magnum thus increasing the likelihood of internal cavitation. Even a small net outflow reduces the overall pressure, which amplifies the opportunity for cavitation. A time-varying acceleration, such as experienced in an accident, complicates a transient pressure pattern, but the above discussion provides the fluid dynamic basis for internal cavitation.

Both the physical and computational model used in this study are over-simplified versions of the human head. However, the same mechanism that influences pressure and cavitation in the cylinder under impact may influence pressure cavitation in the human head under impact. When the head or the cylinder is abruptly accelerated a positive pressure develops under the point of impact and negative pressure develops opposite the impact. When there is enough acceleration then a small cavity from fluid vaporization forms opposite the impact. At low accelerations the model does not compute significant cavitation away from the boundary. The dominant vapor cavity is always located at the skull/fluid boundary. However, if there are large accelerations, if a pre-existing stress is added to the load, or if there are complex loading conditions, then cavitation is possible inside of the fluid.

Therefore, it is reasonable to assume that when testing human surrogates an impact to a stationary head is unlikely to produce cavitation in the brain material. However, it is possible for cavitation to occur opposite impact at the skull-dura boundary, in the brain material near the skull-dura boundary, as well as elsewhere in the brain if there is pre-existing stress or a complex load path. Such conditions would exist in a vehicle crash as the occupant moves with respect to the vehicle before head impact. To attempt to understand what might happen in a vehicle crash, the model was pre-stressed according to estimated neck loadings. In addition, it is possible to generate pre-existing stress from other sources such as previous accelerations, angular acceleration, or angular velocity.

The numerical model predicts a cavitation zone on the boundary opposite impact with more cavitation occurring in the upper far side quadrant than in the lower far side quadrant. There is a variable void fraction throughout the zone at any instant. A more sizable cavity develops at the boundary. The cavitation zone grows spatially during the early stages of the low pressure, then reduces the size and disappears just before the collapse of the boundary cavity. Small amplitude pressure spikes are associated with the condensation of the internal vaporous zone. The largest pressure spike occurs at the instant of boundary cavity collapse. This type of behavior is entirely in keeping with low-pressure transients in liquid pipelines [9]. In this setting, high frequency pressure oscillations are common during the collapse of a vaporous zone, followed by larger amplitude pulses, normally associated with isolated cavity collapse at boundaries, or at high points in the pipeline.

Cavitation has been proposed as a mechanism of counter-coup lesions, which are found in the area of the head opposite impact [11, 12]. This hypothesis states that during impact an area of pressure low enough to cause a cavity exists at the counter-coup site. However, the exact mechanism by which cavitation may cause injury is unknown. Cavitation erosion is a well-known damage mechanism in fluid mechanisms [13]. However, the conditions of occurrence are quite different from those that occur herein and in head impacts in that it normally appears under high velocity flow near an imperfect surface. Small vapor bubbles form, grow and collapse all within 1 to 3 ms. The violent collapse occurs at a time scale of micro seconds, during which time high intensity pressure spikes may do localized physical damage to a solid surface if the phenomena is repeated over a sufficiently large number of cycles. Pressure intensities of the order of a gigapascal may be possible, and temperatures as high as 5,000°C [13]. Because most of the published data addresses damage in metals, cavity creation does not play a significant role; instead, the damage done by cavitation is the result of violent cavity collapse. However, there may be additional mechanisms of damage in brain tissue. When a vaporous region is formed in the

brain, large local strains in the tissues that have contact with each micro-void develop. The micro-voids are generally small enough so that geometric attenuation limit the magnitude of the effect. Nonetheless, if this occurs near a sensitive cellar structure, then it is possible that significant disruption of cellular neurofilament, microtubular structure, or axioplasmic transport might occur. Therefore, the size of the cavity and the forces generated during expansion or collapse do not have to be as large as in metals to cause damage in brain tissue. They could be as much as several orders of magnitude smaller.

Therefore, cavitation damage in brain tissue may be more common and have a greater variety of modes than previously thought. Injury attributed to other mechanism may be the result of cavitation. Cavitation damage/cell-disruption mechanism could be a cause of some of the diffuse axonal injuries (DAI) seen in head injuries from vehicle accidents. It is possible that the cellular disruption causes local damage that does not cause immediate cell death, but instead disrupts the cell activity to the degree that visible injury occurs later [14, 15, 16, 17]. The mechanism of numerous micro-voids forming a large vaporous region, by its nature, would produce damage that could be non-local and defuse throughout the brain. Cavitation requires the high accelerations generally driven by head strikes which are considered the sole or dominant cause of head injuries in the automobile environment (DAI inclusive) [18].

Although the complexity of head injury with apparent multiple injury modes created by many possible injury mechanisms appears to preclude one mechanism, cavitation is consistent with many of the observed biomechanics. For example, the 150-g acceleration for the initiation of vaporous regions to form in the head as an injury mechanism is consistent with the results previously observed [19]. It is also consistent with the logic behind HIC, both acceleration magnitude and duration are important. Short duration accelerations need higher accelerations to produce equivalent cavitation strain or accompanied violent collapse in the brain.

In addition, the smaller the brain cavity the greater the acceleration needed to produce cavitation and the associated damage. This is consistent with observations that it requires greater acceleration to cause damage in the smaller brain sizes [20, 21]. The acceleration necessary to cause damage in a primate is greater than that in a human. This acceleration could be either linear or angular or a combination of both. Angular acceleration adds significant complication for predicting where cavitation will occur in both primates and humans. For example, if the head is spun about its center of mass without linear acceleration (difficult to do if the head is attached to the neck) then cavitation will occur internally without any surface events. If both the linear and angular accelerations are complex such as occurs in a head impact in the automobile environment, then cavitation could occur at multiple locations in the brain.

For both large and small brains it is important to have a pressure gradient (caused by linear or angular acceleration) of sufficient duration to cause cavitation damage. This damage could be produced by impact or inertial load. However, the accelerations needed to produce cavitation are unlikely to occur by inertial loads in an automobile environment. Instead, impacts of the head would be the only way to produce cavitation type damage. This is consistent with previous results [18] in which no injury to the brain material occurred without impact in the automobile environment. Coupled with the acceleration response, impacts produce additional complexity in that the pressure gradients are also influenced by the skull deformation.

IMPLICATION FOR BIOLOGIC SOFT TISSUE MODELING

The results presented here have implication for modeling of biologic material. As a result of cavitation, it is reasonable to expect that soft tissue will have a material model that is a function of the state of stress, both the yield surface and elastic response. For example, shear response is considered important in brain injury mechanisms. As discussed above, when the head acceleration exceeds 150 g's then some form of voids in the brain can occur. How these voids affect the response of the brain in terms of shear, local brain material properties, volumetric and shape changes, energy management and damage properties is unknown. When the head acceleration exceeds 200 g's cavitation response can be violent and there is the chance that the temperature of the collapsing cavity can exceed 5,000°C [22, 23, 24]. In addition, brain material also has weak Van der Walls bonding forces and may behave like organic polymers with both tension/compression and shear response a function of hydrostatic forces. It may turn out that a brain material model, or any soft tissue model, may have to be developed that takes into account the state of stress including the effects of cavitation and micro voids. Furthermore, it has been reported [25, 26] that the Casimir effect can be significant in biological materials. This may also affect how voids form in brain tissue and influence their effect on response and injury.

In the experiments and in the computational model no deformation occurred. Deformation in the human skull can cause both magnitude and phase differences in the pressures as well as contaminate the acceleration response [2]. However, it would seem reasonable to assume that an impact which increases the skull volume, front to back or back to front, would cause greater cavity formation and conversely, an impact that decreased volume would decrease cavity formation.

CONCLUSIONS

This paper presents a limited experimental and computational study of the impact response of the head with regard to cavitation. Because of the complex nature of the cavitation phenomenon during an impact event more work is necessary before the results can be generalized. However, the following conclusions can be made for non-skull fracture cases:

1. Computational models of the head, addressing severe head impacts, should include some form of fluid vaporization at the skull-brain boundary and in the brain material.
2. The potential for cavitation during impact is influenced by pre-existing stress.
3. Vertical loading on the head by the neck before impact can affect the potential for cavitation in the brain material. This is particularly true if the preloading causes a higher pressure at the base of the head, leading to an outflow of material through the foramen magnum.
4. Cavitation in the form of micro-voids inside the brain material may cause local cellular damage that does not produce immediate cell death. It may be one of the causes of diffuse axonal injuries (DAI).
5. The pressure response due to impact can be scaled in proportion with the physical size of the head form unless there is cavitation occurring. Head forms of a different sizing will experience an acceleration inversely proportional in magnitude to the size difference in order to produce a similar pressure/cavitation response.
6. Cavitation may be more common than previously thought because there may be more than one damage mechanism.

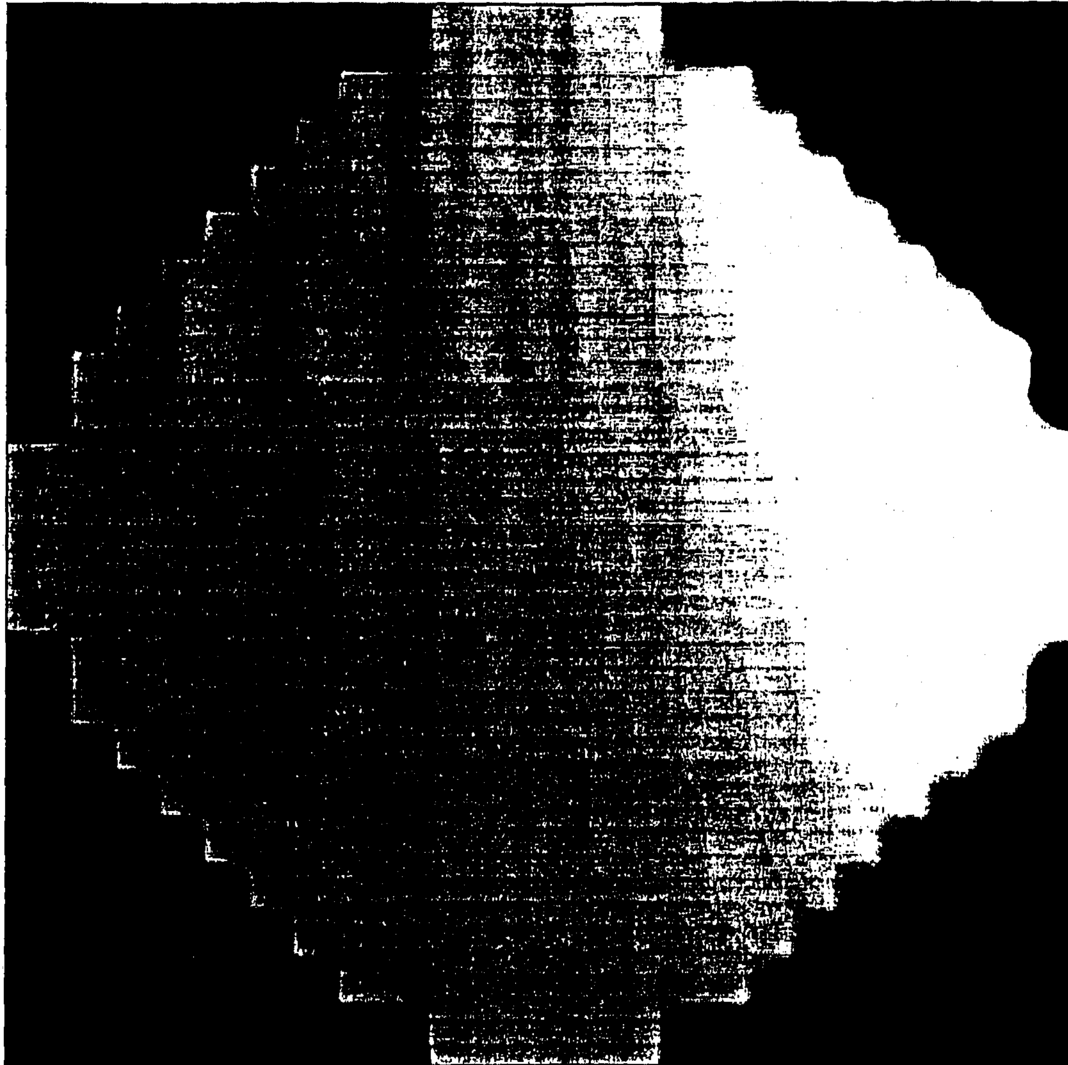


Figure 12: Vaporous region of head model for low acceleration, t = 3.1 ms. Note the vaporous region is in white while the rest of the head model domain is in gray.

References

[1] NUSHOLTZ, G. S., LUX, P., KAIKER, P. S., and JANICKI, M. A. (1984). Head impact response-skull deformation and angular acceleration. Proc. 28th Stapp Crash Conf. 841657, pps. 41-74.

[2] NUSHOLTZ, G. S., KAIKER, P. S., and LEHMAN, R. J. (1986). Critical limitations on significant factors in head injury research. Proc. 30th Stapp Car Crash Conf. 861890, pps. 237-68.

[3] NUSHOLTZ, G. S. and WARD, C. C. (1987). Comparison of epidural pressure in live anesthetized and postmortem primates. Aviat. Space Environ. Med. 88:9-17.

[4] NUSHOLTZ, G. S., KAIKER, P. S., WYLIE, E. B., and GLASCOE, L. G. (1994). The effects of the skull/dura interface and foramen magnum on the pressure response during head impact. 14th ESV Conference, Munich, paper 94 S1 0 26.

[5] NUSHOLTZ, G. S., WYLIE, E. B., and GLASCOE, L. G. (1995). Cavitation/Boundary Effects in a Simple Head Impact Model. Aviat. Space Environ. Med., 66(7):661-667.

[6] NUSHOLTZ, G. S. and WYLIE, E. B., and GLASCOE, L. G. (1995). Internal Cavitation in Simple Head Impact Model. Journal of Neurotrauma, 12(4):707-714.

[7] HENKE, R. (1980). Numerical procedure for predicting the torsional response of solid media. Ph.D. dissertation, The University of Michigan

[8] WYLIE, E. B. (1984). Linearized two-dimensional fluid transients. J. of Fluids Engineering, ASME 106, 227-232.

[9] WYLIE, E. B. and STREETER, V. L. (1993). Fluid transients in systems. Prentice Hall, New Jersey.

[10] WYLIE, E. B. and STREETER, V. L. (1980). Multidimensional fluid transients by latticework. J. of Fluids Engineering, ASME 102:203-210.

[11] BENEDICT, J. V., HARRIS, E. H., and VAN-ROSENBERG, D. V. (1970). An analytical investigation of the cavitation hypothesis of brain damage. Biomechanics and Human Factors Conf. ASME 70-BHF-3.

[12] LUBACK, P. and GOLDSMITH, W. (1988). Experimental cavitation studies in a model head-neck system. J. Biomech. 13:1041-1052.

[13] KNAPP, R. T., DAILY, J. W., and HAMMETT, F. G. (1970). Cavitation. New York; McGraw Hill.

[14] POVLISHOCK, J.T., BECKER, D.P., CHENG, C.L.Y., and VAUGHAN, G.W. (1983). Axonal change in minor head trauma. J. Neuropathol. Exp. Neurol. 42:225-242.

[15] CHENG, C.L.Y., and POVLISHOCK, J.T. (1988). The effect of traumatic brain injury on the visual system: A morphologic characterization of reactive axonal change. J. Neurotrauma 5:47-60.

[16] ERB, D.E., and POVLISHOCK, J.T. (1988). Axonal damage in severe traumatic brain injury: An experimental study in cat. Acta Neuropathol. 76:347-358.

[17] TOMEI, G., SPAGNOLI, D., DUCATI, A., et al. (1990). Morphology and neurophysiology of focal axonal injury experimentally induced in the guinea pig optic nerve. Acta Neuropathol. 80:506-513.

[18] McLEAN, J.A. (1996). Brain Injury Without Head Impact Traumatic Brain Injury Bioscience and Mechanics edited by Bandak, F.A., Eppinger, R.H., and Ommaya, A.K.

[19] MERTZ, H.J., PRASAD, P. and NUSHOLTZ, G.S. (1996). Head Injury Risk Assessment for Forehead Impacts SAE 960099.

[20] GENNARELLI, T.A. (1994). Animate models of human head injury. J. Neurotrauma, 11:357-368.

[21] GENNARELLI, T.A., THIBAULT, L.E., ADAMS, J.H., GRAHAM, D.I., THOMPSON, C.J., and MARCINCIN, R.P. (1982). Diffuse axonal injury and traumatic coma in the primate. Ann. Neurol. 12:564-574.

[22] KNIGHT, P. (1996). Sound, Light and the Vacuum, Nature Vol. 381(6):27.

[23] EBERIEIN, C. (1996). Phys. Rev. Lett 76 3842.

[24] POOL, R. (1994). Science Vol. 286:1804.

[25] SPRUCH, L., Long-range Casimir Interactions, Science, Vol. 272.

[26] NIR, S. (1976). Prog. Surf. Sci. Vol. 8.

970392

Heart Injuries Among Restrained Occupants in Frontal Crashes

Jeffrey S. Augenstein, Elana Perdeck, Jami Williamson, James Stratton, Mary Murtha, and Kathryn Sapnas
University of Miami School of Medicine-William Lehman Injury Research Center

Kennerly Digges and A. C. Malliaris
George Washington Univ.

Louis Lombardo
U.S. DOT, National Highway Traffic Safety Administration

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The William Lehman Injury Research Center has conducted multi-disciplinary investigations of one hundred seventy-eight crashes involving adult occupants protected by safety belts and air bags. In all cases, serious injuries were suspected. Nine cases involved serious heart injuries. These cases are not representative of crashes in general. However, when used in conjunction with National Accident Sampling System; Crashworthiness Data System (NASS/CDS) they provide insight into the most severe injuries suffered by restrained occupants in frontal crashes.

Heart injuries are rare, but when they occur they are usually life threatening. NASS/CDS shows that heart injuries comprise about 0.2% of the injuries in frontal tow-away crashes. In the NHTSA file of Special Crash Investigations (SCI) of air bag cases, heart injuries are reported in 1% of the occupants over 15 years of age. Twenty-five percent of the fatally injured occupants had heart injuries, and 83% of those with heart injury died. In the Lehman Center cases, heart injuries are present in 5.1% of the cases. Forty percent of the fatally injured had heart injury, and 78% of the victims with heart injury died. There is a compelling urgency for the early diagnosis and treatment of heart injuries. Studies of the limited data available provide the basis for premises to better define triage criteria for restrained occupants. This paper suggests two additional triage criteria, based on observations from multi-disciplinary studies. These include: (1) passengers in 2-point belts and crashes of 25 mph or higher, with the lap belt unfastened or with the seat full forward; (2) drivers in crash conditions which delay the air bag deployment or permit the driver to be close to the air bag at deployment.

NASS/CDS DATA ON HEART INJURIES

Heart injuries are rare events. The NASS/CDS samples tow-away crashes on US highways. The risk of any injury to an occupant is about 45%. Less than 0.25% of all injuries in frontal crashes are heart injuries.

For the purpose of this study, only frontal crashes are considered. The NASS/CDS 1988-1995 indicates a yearly average of 2,088,961 passenger car occupants in frontal tow-away crashes (excluding crashes with subsequent rollover or ejection). Forty-five percent of these occupants were injured. On the average, each injured occupant had 2.52 injuries recorded. Serious (AIS 3+) injuries accounted for 3% of all injuries. For those occupants protected by 2-point belts, the risk of serious injury was twice as high as for those protected by 3-point belts or 3-point belts with air bags. The total number of AIS 3+ heart injuries is estimated at 5,100 per year. About 38% of these occur in the 7.5% of the population protected by 2-point belts, while 28% occur in the 49% of the population protected by 3 point belts.

Table 1 shows the fraction of heart injuries as a percentage of all injuries, and as a percentage of serious (AIS 3+) injuries for each type of restraint system. The heart injury fraction among serious injuries is much larger than the fraction of all injuries. This relationship demonstrates that when heart injuries occur, they are nearly always serious injuries.

TABLE 1. Heart Injury Percentages in Frontal Crashes

Restraint	Heart Injury % of Injuries	Heart Injury % of AIS 3+
None	0.15%	4.0%
3-point Belt	0.15%	6.5%
2-point Belt	1.20%	27.4%
Belts + Air Bag	0.03%	0.9%
Air Bag Only	0.30%	5.1%

The data in Table 1 also indicates that the air bags plus belt system have the lowest heart injury fractions of all protection systems.

Among occupants protected by 2-point belts, heart injuries comprise 27.4% of their AIS 3+ injuries. The small number of AIS 3+ injuries result in an error band of 16% for the 27.4% heart injury fraction. Still, the risk of heart injury for 2-point belt protected occupants is much higher than for other protection systems. The increased heart injury fraction for this population is further accentuated by the 2.5 times higher frequency of all AIS 3+ injuries among the 2-point belt protected population compared to the 3-point belt protected population. Each year there are 157,000 occupants

protected by 2-point belts in frontal crashes. About 65,000 thousand of these occupants are injured, and suffer about 1,900 heart injuries.

HEART INJURY MECHANISMS

The William Lehman Injury Research Center supports a multi-disciplinary crash investigation team at the Ryder Trauma Center, one of the most comprehensive Level I Trauma Center in the United States. The criteria for admission to the study are as follows: (1) the subject must have been injured inside a post 1980 passenger car that was involved in a frontal collision; (2) the subject must have been protected by a safety belt, an air bag, or both; (3) at the crash scene, the subject must have met trauma criteria (see Table 2) for suspicion of injuries of a severity which justified transporting to the Ryder Trauma Center, or must have suffered subsequent complications which required emergency interventions; (4) the subject must have agreed to have all records included in the study. All admissions to the Ryder Trauma Center were evaluated for possible admission to the study. All of those that were found to meet study criteria were asked to participate.

TABLE 2. Adult Trauma Criteria

- Systolic BP $\leq$ 90 (Shock)
- Respiratory rate < 10 per minute or > 29 per minute
- Glasgow Coma Scale $\leq$ 12
- Penetrating injury to head, neck, chest, abdomen or groin
- Paralysis
- Amputation proximal to wrist or ankle
- Ejection from motor vehicle
- Paramedic Judgment: High Index of Suspicion of Injury
- Second or third degree burns $\geq$ 15% Total Body Surface Area

Detailed crash and injury investigations are documented for each seriously and fatally injured motor vehicle occupant who was protected by a belt system and/or an air bag. The crash investigation is performed by an experienced crash investigator, and the injuries are assessed and described by the Center's medical staff.

Previous results of multi-disciplinary analysis have been reported in earlier papers [Augenstein, J., [and others]., 1994, 1995 and 1996] . Initial improvements in triage criteria were documented in a NHTSA report [Lombardo, 1993].

In general, heart injuries are serious, and often require immediate and urgent treatment. Yet, in some cases, the heart injury may be difficult to detect. Clues from the crash may assist in detecting these serious and time critical injuries. One of the objectives of the Lehman Center research is to identify clues which may be used to assist in diagnosing heart injuries.

There are several heart injury patterns which have been observed among the injured cases at the William Lehman Injury Research Center. These injuries are depicted in Figure 1 and summarized as follows:

1- Contusion of the surface of the heart. This injury others listed can degrade the heart's function, resulting in lowering the blood supply to the brain and to critical organs. Minor contusions may have no clinical impact. Major contusions may drastically reduce the hearts pumping ability (that is a reduction in cardiac output). Bleeding from the contused outer surface of the heart may produce fluid in the pericardial sac. This is called pericardial tamponade. The result can be severe deterioration of cardiac output.

2- Rupture of the ventricle. This injury is generally associated with compression of the heart while filled with blood. The right ventricle operates at a lower pressure than the left ventricle and can withstand less over pressure. Rupture is most likely to occur in the right ventricle. This injury causes loss of heart function, and bleeding into the pericardium. Immediate treatment is required.

3- Laceration or rupture of the atrium. This injury causes loss of heart function, and bleeding into the pericardium. Immediate treatment is required.

4- Laceration of aorta. This injury frequently occurs at the end of the aortic arch. It may be caused by displacement of the heart relative to the aortic attachment. Immediate treatment is typically required.

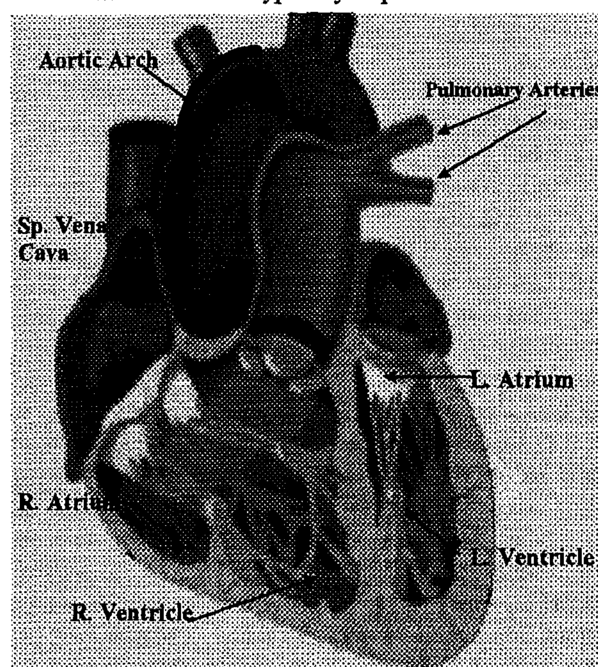


Figure 1. Diagram of Heart

The injuries in cases at the William Lehman Research Center are more severe than observed in NASS/CDS. The heart injuries comprise a larger portion of the injury population. Seventy-eight percent of the occupants with heart injuries died. The injury rate for each restraint system is summarized in Table 3. In this sample, the relative risk of heart injuries for air bag protected occupants is slightly higher than the average rate of 5.1%.

Table 3 indicates that among occupants with severe injuries, heart injuries remain a small percentage. However, 78% of occupants with heart injury died. An estimated twenty-five percent of those who died from heart injuries may have been saved if the time for diagnosing and treating the

injury could be reduced. Observations made at the crash scene may be vital to the rescue and treatment of occupants with heart injuries.

TABLE 3. Number of Cases and Percentage of Heart Injuries in Lehman Center Data

Restraint	Number	Heart Injury
3-point Belt	74	2.7%
2-point Belt	36	5.6%*
Belts + Air Bag & Air Bag Only	68	7.4%
All	178	5.1%

CASES OF BELTED PASSENGERS WITH HEART INJURIES

Only two cases of passengers with automatic 2-point belts have entered the Lehman data. Both cases (100%) had fatal heart injuries. These cases involve healthy young adult men in frontal crashes. Both were front seat passengers, and both may have been asleep at the time of the crash. The drivers in both cases were braking and bracing heavily prior to the crash, and received only minor injuries. These cases are summarized as follows.

CASE 95-029A1



Figure 2. Case Vehicle 95-029A1 Hyundai S-coup, 1994

This crash occurred at 4:33 AM and involved a 1994 Hyundai S-coup (Figure 2). The driver lost control of the vehicle, mounted a curb, impacted a line of small shrubbery, and then a 17 inch diameter palm tree. The impact direction was 12 o'clock. Damage started 5 inches to the left of centerline and continued for 17 inches. The maximum crush was 30 inches, and the delta-V was calculated at 32 mph. Rescue was delayed by the time to call for help in the early hours of the morning and by a lengthy extraction. The victim was air lifted to the Trauma Center, and arrived approximately 45 minutes after the crash. The injuries were as follows:

Laceration, left atrium	AIS-5*
Fracture, anterior right rib cage (2-3)	AIS-3
Abrasions, contusions and lacerations to face, trunk, abdomen, and knees	AIS-1

* *This injury may have been survivable*

The passenger died of his heart injury before it could be diagnosed at the trauma center. More rapid crash identification, location and recovery may have prevented this fatality.

CASE 95-034AP

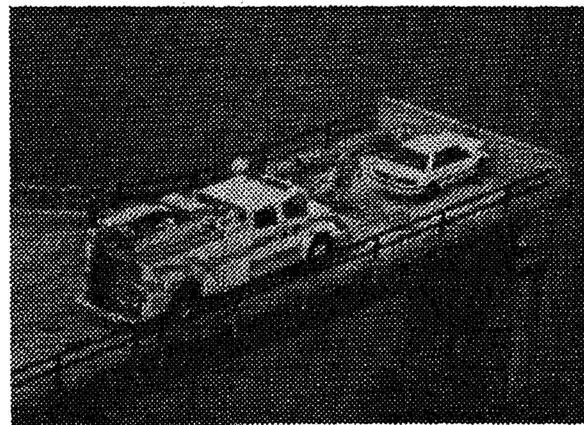


Figure 3. Case Vehicle 95-034AP Mazda Protégé, 1994

This crash occurred at approximately 5:00 AM. The vehicle was a 1994 Mazda Protégé (Figure 3). The driver entered a closed road in a sparsely traveled area and impacted a concrete barrier with a delta-V of 29 mph. The vehicle created tire skid marks for 55 feet before striking the barrier. The extent of damage was 18.5 inches of crush uniformly distributed across the front of the car. Intrusion was 9 inches for the left toepan, and 11.5 inches for the right toepan. The passenger was 32 years old, 72 inches tall, and weighed 175 lb. He was protected by a shoulder belt, and he was wearing a manual lap belt. A suitcase was jammed behind the seat, and his seat was in the full forward position. The suitcase remained wedged behind the seat after the crash, and no permanent deformation of the seat structure was observed. His injuries were as follows:

Laceration, distal arch of aorta (contained in mediastinum)	AIS-5
Contusion, heart	AIS-3
Contusion, right lung	AIS-3
Fracture, rib, right (2-7); left (4-9)	AIS-5
Laceration, spleen	AIS 2
Contusion, bowel and mesentery	AIS 2
Laceration, parietal pleura, right	AIS-2
Fracture, foot, right	AIS-2
Lacerations abrasions, and contusions (x9)	AIS-1

**complication: brain stem herniation.*

Due to the remote location of the crash, and the early hour of the morning, the crash was not discovered for more than 30 minutes. The occupant expired at the scene. More rapid discovery may have prevented this fatality.

DISCUSSION OF BELTED CASES WITH HEART INJURY

An examination of 30 mph crash tests of the Hyundai S-coup and the Mazda Protégé shows that these vehicles both produced chest injury measures well below the requirements of the federal standards. In the 30 mph tests, the 50% male dummies are positioned with the seat midway along the seat track. The shoulder belt is used, but not the lap belt. The results of 30 mph tests are shown in Table 4.

TABLE 4. Injury Measures For Passenger In 30 Mph Frontal Crash Test

VEHICLE	HIC	Chest G
Allowable	*1000*	*60*
Hyundai S-coup		40
Mazda Protégé	418	39

In the case with the Mazda Protégé, the passenger was using his lap belt. Crash tests at 35 mph are available to indicate the performance of the Protégé with the lap belt fastened. The results are: HIC - 612; Chest G - 51.

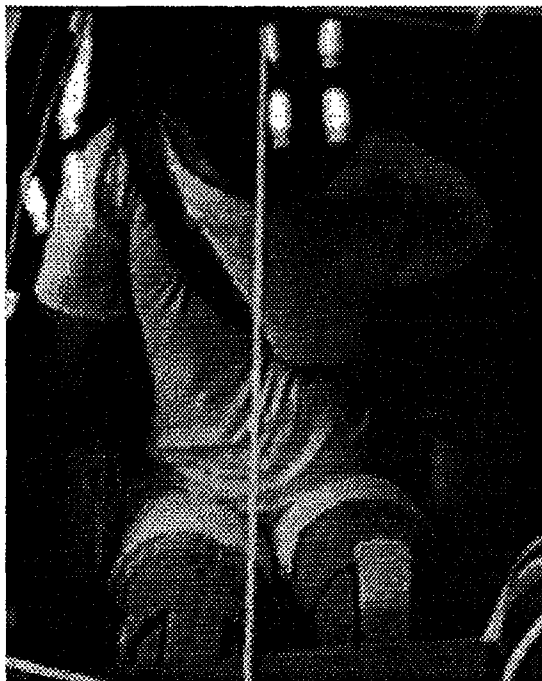


Figure 4 2-point Belt Position in Crash Test of Hyundai S-coup

All these values for chest injury are well below the 60 G's allowed in the standard. However, an examination of the test films suggest that the chest G measures may not adequately account for the concentration of load produced by shoulder belt. In the tests without lap belt, the shoulder belt rides very high on the chest.

Figure 4 is a frame taken from the crash test film of the Hyundai S-coup. In this 30 mph crash into a barrier, the position of the belt over the heart is evident. The concentrated loading of the 2-point belt, when the lap belt is not used, may increase the risk of heart injuries, as observed in the NASS/CDS data.



Figure 5 Belt Position In Mazda Protégé with Seat Full Forward

Figure 5 shows the geometry of the belt system in a Mazda Protégé, with the seat in the fully forward position. The vehicle was tested with the seat in its middle position, so that a more favorable geometry results. The lower portion of the shoulder belt is positioned higher on the chest as the seat moves forward. This change may increase the loading of the chest and the risk of heart injuries.

The observations from crash tests, supported by Lehman Center cases with passenger heart injury suggests that heart injury producing loading may be induced by the shoulder belt, under crash severities around 30 mph and higher.

In searching for clues for early detection of heart injuries, the following criteria is postulated for future application of the Lehman Center research:

- Passenger occupants in 2-point belts and crashes of 25 mph or higher and with the lap belt unfastened or the seat fully forward.

Table 5. Summary of Heart Injury Cases from Special Crash Investigations

Case	Yr.	Make	Model	DV	AIS	Obj. Struck	Age	Sex	Ht.	Wt.	Contact	Fatal	Belt	Remarks
CA9109	91	Ford	Taurus	8	5	CAR FRONT	79	F	62	120	AB	Y	Y	Side Crash
NC9307	91	Chevrolet	Corsica	8	5	TREE	78	M	70	230	AB	Y	N	Bad Heart
CA9303	91	Mercury	Capri	9	5	CAR REAR	22	F	64	Avg.	AB	N	N	Close-in
DS9422	92	Chevrolet	Corsica	11	6	CAR SIDE	73	F	59	128	AB	Y	Y	Bad Heart
IN9506	90	Lincoln	Contin.	12	5	VAN SIDE	38	F	62	110	AB	Y	N	+Brain Injury
CA9112	90	Dodge	Shadow	14	3	POLE	36	F	63	112	AB	Y	Y	Asleep
CA9309	90	Ford	Taurus	16	5	CAR REAR	71	F	62	129	AB	Y	N	+Brain Injury
DS8912	85	Ford	Tempo	18	4	CAR REAR	53	F	63	146	AB	Y	N	Braking
CA9110	91	Pontiac	Firebird	19	4	TREE	46	M	72	230	AB	Y	Y	Slumped
CA9502	94	Ford	F-150	19	4	CAR REAR	56	M	71	208	AB	Y	N	Multiple Imp
IN9407	92	Oldsmobile	Delta 88	22	5	TREE	68	M	70	250	AB/SW	Y	Y	Multiple Imp
DC9012	90	Ford	Taurus	25	3	TREE	21	M	72	210	AB/SW	N	Y	Multiple Imp
IN9201	91	Chevrolet	Caprice	32	5	BUILDING	48	M	76	260	AB/SW	Y	N	Multiple Imp
DS9108	90	Dodge	Omni	34	5	CAR FRONT	49	M	Unk	Unk	AB/SW	Y	N	Intrusion

AB = Air Bag; SW = Steering Wheel

AIR BAG CASES WITH HEART INJURY NHTSA SPECIAL CRASH INVESTIGATION CASES

The NHTSA Special Crash Investigation (SCI) file contains records of 1258 drivers who were in vehicles with air bag deployment. The criteria for inclusion in this file is varied. Consequently, it is not possible to extrapolate results to determine nationally representative frequency of occurrence. However, the file is useful for examining injury patterns among air bag protected occupants.

The file contains 56 cases with fatal injuries. Twenty-five percent of the fatal cases had severe heart injuries. Among the fatal crashes with severe heart injury, 64% were at crash speeds of less than 20 mph. Thirty-six percent were at crash severities of less than 15 mph. In addition to the fourteen fatal injuries, there were three drivers with AIS 3+ heart injuries who survived.

A summary of 14 cases of drivers with heart injury documented in the NHTSA Special Crash Investigation study is shown in Table 5. Three documented cases of driver heart injury were not included in Table 5 because they involved predominately side impacts or driver ejection. The columns in Table 5 show the case number; the vehicle year, make and model, the crash severity (mph); the heart injury AIS; the object struck; the occupant's age, sex, weight (lb.), and height (in); the object contacted by the occupant which most likely caused the injury; whether the occupant died; whether the belt was worn; and remarks which may have influenced the outcome.

The data from Table 5 indicates 11 cases of heart injury at crash speeds below 23 mph. Four of these were impacts with trees or poles. Four were impacts with the rear of another car, and two were impacts with the side of another vehicle. These types of crashes produce relatively low acceleration crash pulses, which may delay the deployment of air bags. It is probable that the occupant may have been close to the deploying air bag in these cases. Of the seven fatally injured drivers in crashes at speeds below 19 mph, four were older than 70, and six were less than 65 inches tall. Preexisting heart disease was reported in two of the older drivers. A close position of the driver relative to the air bag at the time of deployment was suspected for most of the cases. In the higher severity crashes with the occupant belted, multiple impacts of the vehicle may have contributed to the occupant's position and injuries.

LEHMAN CENTER HEART INJURY AIR BAG CASES

The Lehman Center has investigated five crashes with air bag deployment and heart injury to drivers. Four of these cases were either at crash severity's well above 30 mph, or involved massive intrusion. One case was at 23 mph and is summarized as follows.

CASE 93-026A



Figure 6. Case Vehicle 93-026A Infinity M-30, 1992

This crash occurred at 11:22 AM at a city intersection. The driver of a 1992 Infinity M-30, suddenly swerved into the opposite lane and impacted the front of a stationary vehicle (Figure 6). The delta-V was 23 mph, and the crash direction was one o'clock. The maximum crush was 20 inches, with no compartment intrusion. The driver

was a 67 year old male, 69 inches tall, weighing 137 lb. He was not wearing the 3-point belt when his air bag deployed. The pre-crash trajectory suggest that the driver may have blacked out, and he may have been close to the bag at deployment. His injuries were:

Laceration, cardiac atrium, right	AIS-5
Laceration, liver	AIS-2
Laceration, spleen	AIS-2
Lacerations abrasions, & contusions (x6)	AIS-1

The driver expired in resuscitation approximately 30 minutes after the crash. The injuries indicate that this relatively frail person was not adequately protected by the air bag.

NASS/CDS FATAL CASES

The NASS/CDS for the years 1990-93 contains 26 driver fatalities with air bag deployments. Analysis of the fatal cases which may have been caused by the air bag was conducted by Lund [Lund 1996]. According to this study, 7 of the 26 cases, involved frontal impacts as the only harmful event. Another 7 cases involved frontal impacts with multiple harmful events.

The authors found three frontal single impact cases and one frontal multiple impact case in which the air bag may have contributed to the injuries. Each case involved serious heart injuries. The suspected air bag induced injuries were as follows:

Case 917902111;
Unbelted, 35 year old female,
5'6", 160 lbs.
Delta V 46mph, Front to rear car-to-car crash

Injuries: Laceration, right ventricle	AIS-5
Contusion, interventricular septum	AIS-4

Case 930512511;
Belted , 62 year old female
5'2", 179 lbs.
DV 29mph , Pole impact

Injuries: Laceration, right ventricle	AIS-5
Flail chest, bilateral	AIS-5

Case 930600611;
Unbelted, 64 year old male
5'7", 160 lbs.
Delta V 29mph, Tree impact

Injuries: Laceration, right atrium	AIS-6
Fractures, bilateral ribs	AIS-4
Laceration, superior vena cava	AIS-3

Case 920813311;
Unbelted, 58 year old male,
6'0", 187 lbs.
Delta V unknown, Multiple impacts with shrubs

Injuries: Fractures, bilateral rib with hemothorax	AIS-5
Laceration, heart	AIS-3
Laceration, pericardium	AIS-2

DISCUSSION OF AIRBAG CASES WITH HEART INJURY

The NASS/CDS indicates that heart injuries are very rare among occupants protected by air bags. However, a number of heart injuries have been investigated at the Lehman Center, in the Special Crash Investigation Study, and in the NASS/CDS fatalities. The heart injury frequency in these data sets is much higher than seen in NASS/CDS. Of particular concern is the frequency of heart injury in low speed, low acceleration crashes. Crashes involving car-to-poles and car front-to-car-rear are over represented in these crash types. In these classes of crashes, the vehicle initially decelerates more slowly than in a barrier crash. Consequently, the crash is more difficult to detect by present day crash sensors, and the occupant may move closer to the air bag before the bag deploys.

In addition, older and short statured drivers are over represented in the low speed fatal crashes from NASS/CDS and from the SCI file. Older drivers have lower tolerance to heart injuries and present air bags may be too stiff to provide protection in low speed crashes. In low severity crashes, less powerful air bags may be beneficial to this population.

In the higher severity crashes, multiple impacts may be a factor which increases heart injury risk.

The data also suggest that severe heart injuries may be occurring more frequently than recognized in the NASS/CDS data.

CONCLUSIONS

Heart injuries are rare but serious events in the NASS/CDS data. The data from the William Lehman Injury Research Center shows a higher concentration of heart injuries, of nearly 5% of the injuries suffered by patients meeting trauma center triage criteria.

Heart injuries require immediate recognition and treatment. Clues from the crash would be desirable in aiding the detection of heart injuries. Additional clues currently being explored by the Lehman Center include the following:

- Passenger occupants in 2-point belts and crashes of 25 mph or greater with the lap belt unfastened or the seat fully forward.
- Drivers in crash conditions which delay the air bag deployment or permit the driver to be close to the air bag at the time of deployment. The presence of elderly persons and small statured persons in this type of crash adds to the index of suspicion.

In the data from Special Crash Investigation Cases of air bag crashes, heart injuries represent about 1% of the injuries but 25% of the fatalities. The majority of these fatalities occurred at crash speeds below 22 mph. Crashes involving poles and car-front to car-rear impacts are over-represented in these lower severity crashes. Older persons, and short statured persons are also over-represented. In the higher severity crashes, multiple impacts are a frequent complicating factor. Combinations of these factors may be useful in suspecting the presence of rare, but serious heart injuries.

ACKNOWLEDGMENT

This research was conducted with the support of the National Highway Traffic Safety Administration.

REFERENCES

Augenstein, J., [and others]., 1996, " Injuries Sustained by Air Bag Protected Drivers", ESV Conference, 1996.

Augenstein, J., [and others], 1995, "Occult Abdominal Injuries to Air Bag-Protected Crash Victims: A Challenge to Trauma Systems", The Journal of Trauma, Vol.38, No.4, p.502-508, April 1995.

Augenstein, J., [and others]., 1994, "Air Bag Protected Occupants - The Challenge of Identifying Occult Injuries", Proceedings of 14th ESV Conference, May 1994.

Lund, A., Ferguson, S., and Powell, M., 1996, "Fatalities in Air Bag-Equipped Cars: A Review of 1989-93 NASS Cases", SAE 960661, 1996.

Lombardo, L., and Ryan, S., 1993, "Detection of Internal Injuries in Drivers Protected by Air Bags", Research Note, NHTSA, 1993.

970393

Relationships Between Crash Casualties and Crash Attributes

A. C. Malliaris
DeBlois Associates

K. H. Digges
George Washington Univ.

J. H. DeBlois
DeBlois Associates

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This work addresses and evaluates the likelihood of human casualty in highway crashes, projected on the basis of field crash data that may become available electronically by sensors at crash time, and/or observed at the crash scene by emergency attendants. Termed collectively as a "crash signature", such data are treated as predictors and are selected from: crash severity, general area of damage, direction of force, occurrence of rollover, intrusion, vehicle crush and its specific horizontal location, collision partner, vehicle class and size, occupant age, gender, restraint use and type, seating position, and other. Crash signatures are converted into responses such as: (a) the likelihood of the most severe outcome, fatality or survived injury, by severity AIS per occupant; and (b) the same per vehicle. Cars are the vehicles selected for this investigation. A likelihood is quantified by a probability of occurrence, as a function of a string of predictors selected for maximum resolution and sensitivity, and minimum contribution to error. Likelihood determinations are performed via maximum likelihood based logistic regressions, best suited for treating dichotomous responses: "yes or no" such and such a response or outcome. Each likelihood is accompanied by a standard error or by upper and lower confidence bounds, and each procedure is evaluated by pertinent scores. All cited procedures and findings are based on the data of the National Accident Sampling System (NASS) files 1988-1995, compiled by the National Highway Traffic Safety Administration (NHTSA). This provides a nationally representative sample of about 95,000 crash involved car occupants, and 190,000 incurred injuries, all with attributes that collectively encompass as a minimum the predictors and responses cited earlier. The paper provides pertinent predictive relations which, notwithstanding complexity, are fully programmable. Probabilities of specific outcomes may vary from nearly zero to virtually 100%, depending on circumstances. Detailed and illustrative findings are presented in tabular and graphic forms.

INTRODUCTION

The advent of high volume highway accident records, many of them nationally representative and of a research caliber, addressing a broad spectrum of crash, vehicle, occupant, and casualty attributes, makes it easier now to address and evaluate several important issues in highway safety.

At the same time, the wide scope and the complexities of the available data point to the need for developing ways and means in order to

capture the essential aspects of the highway crash environment in a succinct and insightful fashion. The purpose of this paper is to take a first cut in responding to the cited need.

BASIC DATA

The data compiled in the eight years, 1988-1995, of NASS/CDS are the basic data used. The NASS weights, necessary for national projections, are used as weighing factors in any data processing procedure.

Many outcomes and their severity may be considered individually or in combinations for the purpose of human casualty prediction. Also many crash, car, occupant, and injury attributes are in principle available in the accident experience to assist, as predictors, in the derivation of said algorithms on the basis of a crash signature. The number and type of predictors are often limited by practical considerations imposed by the type of contemplated applications, the strength of probability --predictor correlation, and by the amount and quality of available data for the derivation of the algorithms.

NOMINAL PROCEDURE FOR TREATING THE RAW DATA

In view of the dichotomous nature of the outcomes at issue (e.g. "Yes" or "No" Fatality or MAIS 3+) a maximum likelihood procedure, specifically a logistic regression with weighing factors, is used to fit various algorithms to the raw data. Essentially, the probability of casualty is projected as:

P = 1 / [1 + exp(-w)] (1)

w = A0 + A1*PRED1 + A2*PRED2 + ... (2)

where PRED1, PRED2, etc are the selected predictors; and A0, A1, A2, etc are coefficients estimated by the logistic regression.

When dealing with analyses of data from the NASS, it must be taken into account that this file contains a sample as opposed to a census of national data. In order to deal with this, the applicable statistical procedures are those prescribed in "Survey Data Analysis" (SUDAAN) software, Research Triangle Institute, Research Triangle Park, North Carolina, 1992. Such procedures are applicable in the analysis of data from multi-stage sample designs, like that of the NASS.

ESTIMATION OF STANDARD ERRORS AND CONFIDENCE BOUNDS

The SUDAAN logistic procedure yields values for coefficients: A0, A1, A2, etc appearing in (2). The same procedure provides also the covariance matrix: COV(Ai, Aj). This helps in the calculation of the variance of the argument w of the probability appearing in (1). Specifically, the variance of w is given by:

var(w) = Sum[Cov(Ai, Aj)*xi*xj] (3)
over all i and j

Note that i or j assume the values: 0, 1, 2, etc, corresponding to the intercept and the predictors appearing in relation (2). When an analyst assigns desirable values to xi and xj, an application of (3) yields the variance: var(w).

Moreover, to a first approximation, the variance of the probability (1) is given by:

var(P)={exp(-2w)/[1+exp(-w)]**4}*var(w) (4)

and the standard error of P is:

seP = square root [var(P)] (5)

Also to a first approximation, the 95% confidence bounds of P are given by: P +/- (1.96 * seP)

ADDRESSED PREDICTORS

In this paper the basic data, i.e. data concerning car occupants involved in towaway crashes, are used for the derivation of algorithms that estimate: (a) the probability of a crash involved occupant being a fatality; (b) the probability of a crash involved occupant with at least one injury of maximum severity MAIS 3+; and (c) the probability of a crash involved occupant with at least one injury of MAIS 2+. Extension to other outcome populations is readily evident. Short notation for the predictors addressed in developing algorithms is given below:

ONEVEH Single Car Crash
BIGTRK Collision with Large Vehicle
ROLL Planar Crash with Rollover Occurrence
DELTAV Total Delta V, mph
GADSP Side Damage, Passenger Compart.
GADSNP Side Damage, Excluding Passenger Compartment
GADB Rear Damage
DOFS Direction of Force: 8-10 & 2-4 O'Clock
DOFB Direction of Force: 5-7 O'Clock
MAXC Maximum Crush, inches
INTRU Intrusion, 6 inches or more, in Front
CURBWT Car Curb Weight, in 100 lbs
FRPAX Right Front Seat Pasenger
RRPAX Rear Seat Pasenger
BELT Safety Belt Use
BEBA Air Bag Deployment & Belt Use
AGE Car Occupant's Age
FEMALE Occuppant's Gender
OCCWT Occuppant's Weight, lbs
OCCHT Occuppant's Height, inches
ENTRP Entrapment
EJC Complete Ejection
EJP Partial Ejection

These predictors are alternatives and operate in conjunction with a baseline that addresses unrestrained male car drivers, in primarily planar and frontal crashes with other cars. Thus, given the cited baseline, the only two alternatives addressed above for a collision partner are: ONEVEH and BIGTRK. Similarly GADSP, GADNSP, and GADB are the alternatives to frontal damage included in the baseline. Top damage is not addressed because of a relatively low incidence.

In fact it is important to note that rollover in general is not included in the development of the algorithms at issue. This is necessitated by the desire to include Delta V as a most influential parameter. This parameter is not defined in general rollover. Thus, predictor ROLL appearing in the above list covers the crashes which are initially planar, with a possible subsequent rollover.

Other implicitly understood alternatives not named in the above list are: direction of force 11 to 1 O'Clock, driver's seat, no restraint use, male occupant, and no ejection. These are baseline attributes. Most other attributes are either binary (yes or no), or continuous.

PROGRAMMABLE ALGORITHMS

Optimal algorithms are presented below in order of increasing complexity. Relation (1) is always the basic platform. The simplest way of formulating the exponent "w" is in terms of DELTA V, the most influential parameter, as shown below:

w = A0 + A1*DELTAV (6)

where DELTAV=Total Delta V in mph continuously.

The logistic regression (1) and (6) applied on the basic data, seeking the probability for fatality or MAIS 3+ or MAIs 2+, yields the coefficients A0 and A1 and the associated standard errors, with numerical values shown in the first cluster of Table I. Next we address the more complex, but still relatively simple, fit shown in (7) below:

w = A0 + A1*DELTAV + A2*DOFS + A3*DOFB (7)

where in addition to DELTAV we include:
DOFS = 1 if the direction of force is 8-10 or 2-4 O'Clock, else DOFS=0;
DOFB = 1 if the direction of force is 5-7 O'Clock, else DOFB=0; and if DOFS=DOFB=0 then the direction of force is 11 to 1 O'Clock.

Numerical values for the coefficients A0, A1, A2, and A3 appear in the second cluster of Table I.

In a similar fashion we augment the resolution of the algorithms by including additional predictors as shown in the following three progressively complex cases:

w = A0 + A1*DELTAV + A2*DOFS + A3*DOFB
+ A4*AGE + A5*BELT + A6*BEBA (8)

w = A0 + A1*DELTAV + A2*DOFS + A3*DOFB
+ A4*ROLL + A5*FRPAX + A6*RRPAX
+ A7*AGE + A8*BELT + A9*BEBA (9)

w = A0 + A1*ONEVEH + A2*BIGTRK + A3*ROLL
+ A4*DELTAV + A5*GADSP + A6*GADSNP
+ A7*GADB + A8*MAXC + A9*INTRU
+ A10*CURBWT + A11*FRPAX + A12*RRPAX
+ A13*BELT + A14*BEBA + A15*AGE
+ A16*FEMALE + A17*OCCWT + A18*OCCHT
+ A19*ENTRP + A20*EJC + A21*EJP (10)

where in addition to predictors defined earlier,
AGE = occupant's age in years continuously;
BELT = 1 if a safety belt is in use; else BELT=0;
BEBA = 1 if an air bag deploys in addition to a safety belt in use; else BEBA=0;

ROLL = 1 if car rollover occurs; else ROLL=0;
FRPAX = 1 if the occupant is in front seat right; else FRPAX=0;
RRPAX = 1 if the occupant is in rear seat; else RRPAX=0; if FRPAX=RRPAX=0 then we deal with the driver;
ONEVEH = 1 if this is a single car crash; else ONEVEH=0;
BIGTRK = 1 if the collision partner is a big truck; else BIGTRK=0; if ONEVEH=BIGTRK=0 then the collision partner is a car;
GADSP = 1 if the damage area is on the car's side and includes the passenger's compartment; else GADSP=0;
GADSNP = 1 if the damage area is on the car's side but excludes the passenger's compartment; else GADSNP=0;
GADB = 1 if the damage area is rear; else GADB=0;
MAXC = maximum crush, in inches continuously;
INTRU = 1 if 6 inches or more intrusion occurs in the front compartment; else INTRU=0;
FEMALE = 1 if the occupant is female; else FEMALE=0;
OCCWT = occupants weight in lbs continuously;
OCCHT = occupants height in inches
ENTRP = 1 if entrapment occurs; else ENTRP=0;
EJC= 1 if a complete ejection occurs; else EJC=0;
EJP=1 if a partial ejection occurs; else EJP=0; if EJC=EJP=0 then No Ejection Occurs.

Numerical values for the coefficients appearing in relations (8), (9), and (10) may be found in the third, fourth, and last cluster of Table I, respectively.

Note in the results shown in Table I that most predictors are binary, except for such predictors as Delta V, Age, etc which are continuous variables. The magnitude of coefficients for such continuous predictors must be interpreted in conjuction with the units of their measurement. Thus the coefficient value of 0.177 for Delta V in Table I (A) goes along with a Delta V in mph. It represents the increase per mph of the corresponding term in the algorithm. Similarly coefficients associated with: age, maxc, curbwt, occwt, and occht elsewhere in the cited Table must be interpreted on a basis of: per each year of age, per each inch of crush, per each 100 lbs of car curb weight, per each lb of an ocupant's weight, and per each inch of an occupant's height, respectively.

GOODNESS OF FIT AND PREDICTED V. OBSERVED OUTCOMES

A car's crash severity, Delta V, is such a strong determinant of occupant casualty outcome that it covers most of the variability observed in the field experience with relatively small errors. See for example Table I (A). Nevertheless, both the predictive resolution of an algorithm, and the goodness of the fit to the field data improve as further predictors are included, even if they are less influential than Delta V.

However beyond a certain point, diminishing returns become evident as may be seen in the progression of algorithm complexity, from (A) to (E) in Table I. As more predictors are included in the analysis, some prove to be quite marginal, given that the error of their coefficient assumes values comparable to the coefficient proper. Thus caution is recommended in order to avoid misleading results and conclusions.
Except for predictors with coefficient errors approaching the coefficient values, all predictors and all algorithms shown in Table I have been found statistically highly significant, on the basis of various statistical scores. Said algorithms account for most of the variability observed in the field experience. Discernible improvements are evident as one progresses from (A) to (C) in Table I, but beyond that point, further improvements are marginal and they may be misleading.

Quantitatively speaking, the bottom line for taking or not taking into account the influence of a predictor is the magnitute of the error relative to the coefficient of the predictor at stake.

In addressing the association of predicted probabilities of a certain outcome with actually observed outcomes we determined that the percentage of correct predictions varies from about 75% to 96%, depending on algorithm complexity and severity of predicted outcome. In addition, we used a score known as Somer's D that measures association on a 0 to 1 scale (no

association to perfect association). For the algorithms that project probability of fatality (A to E in Table I) the cited indicator was found to have values: 0.686, 0.727, 0.801, 0.803, and 0.850, respectively. The corresponding values for MAIS 3+ outcome are: 0.587, 0.626, 0.686, 0.691, and 0.748, respectively.

NUMERICAL APPLICATIONS & ILLUSTRATIONS

The probability of fatality among towaway crash involved car occupants, without any further qualification, is about 0.6%. For injured occupants at MAIS 3+, or at MAIS 2+ the probability is: 6.8% and 14.7%, respectively. These may be considered as alternative statements for casualty rates per 100 occupants.

Each of these rates may be resolved by crash severity, Delta V, with the help of algorithm (6). The results of this resolution are shown in Fig. 1. Further resolution, i.e. by direction of force in addition to Delta V, is provided when algorithm (7) is applied. Figure 2 illustrates results relevant to MAIS 3+. A similar algorithm may provide resolution by Delta V and restraint use and type. Results concerning MAIS 3+ are illustrated in Fig. 3. Resolution by Delta V and occupant's age is illustrated in Fig. 4.

The results obtained via relatively simple algorithms, as illustrated in Figs 1 to 4, could equally well and perhaps more appropriately be obtained from more complex algorithms that encompass the predictors at stake in the cited figures, plus additional predictors that could be influential.

More complex algorithms can be applied for obtaining results concerning one or a few predictors of immediate interest, with stipulated values. In practice this method is applied to a multi-predictor algorithm by assigning to predictors, other than those of immediate interest, their mean values on the basis of the field experience. The advantage of using this more cumbersome method is that the additional predictors are frozen at values common to the entire population under consideration, something that helps minimize possible confounding effects. It is also understood that no additional predictors should be entertained unless they have statistically significant coefficients, i.e. small errors.

MULTI-PREDICTOR ALGORITHMS

Very long and cumbersome algorithms may result if one insists on including many interesting and available predictors, even if all are statistically significant. Applications and results may become intractable. For this reason we developed a good approximation to the projection of a casualty probability on the basis of a known coefficient, without having to make a full application of a cumbersome algorithm. As a first approximation, the following relation holds:

DeltaP = P*(1-P)*A (11)

where DeltaP is the casualty probability increment resulting from the inclusion of a predictor with coefficient A, in any case where a probability P has been projected in the absence of said predictor. Note that in relation (11) the probability P and the increment DeltaP are on a 0 to 1 basis, as opposed to 0 to 100%. Also note that the increment becomes a reduction when coefficient A is negative.

It is evident that the increment or reduction varies from zero, at P=0, to a maximum at P=0.5, and back to zero at P=1. For convenience, relation (11) is graphically illustrated in Fig. 5, for four different values of regression coefficients: 0.25, 0.50, 1.0, and 2.0. Other values are readily obtained by interpolation.

The primary application of (11) is to assist in the quantification of effects described by the coefficients of multipredictor algorithms, such as those appearing in Table I (D) and (E). It is evident that (11) should be applied stepwise, especially when an one step application leads to the absurd result of (P+DeltaP) > 1. For the same reason (11) is applicable to individual evaluations, as opposed to an evaluation of several influences combined.

REVIEW OF THE RESULTS

It is informative to review the numerical results obtained by the algorithms in this paper, in comparison with results known from earlier and independent studies concerning factors that

influence the rates, or probabilities, of casualty. Note, for example in Fig. 1, that the probability of MAIS 3+ is about 50%, consistent with a widely held benchmark.

Resolution by direction of force, see Fig. 2, shows the 8-10 and 2-4 O'Clock impacts to dominate the threat of serious casualty, with the 11-1 O'Clock impacts second, and the 5-7 O'Clock impacts in the familiar distant third rank.

As recognized in several restraint effectiveness studies, the effectiveness of restraints is about 50% from about 10 or 15 mph DeltaV to about 30 mph. Thereafter, the effectiveness starts declining and becomes negligibly small at high crash severities. This is readily evident in the algorithm results shown in Fig. 3.

The small superiority seen for "belt & air bag" v. "belt only" in Fig.3 is actually substantially larger when the results are controlled for possible confounders beyond crash severity only. This is evident in the relevant coefficients appearing in multi-predictor algorithms such as those shown in Table I (C) to(E).

Furthermore, by comparing the relevant restraint coefficients as a function of casualty severity, it appears that the superiority of "belt & air bag" v. "belt only" is dominant at high casualty severities, e.g. fatalities or MAIS 3+, but declines at lower severities, e.g. MAIS 2+. This situation, although still confused by the accompanying statistical uncertainties, is consistent with the notion that the air bag is very effective in preventing high severity injuries although it may cause lower severity injuries.

The multipredictor algorithm (10), with coefficients shown in Table I (E), reveals quantitatively several factors of strong influence, fully consistent with qualitative notions held intuitively. Beyond Delta V, age, and restraint use or type discussed earlier, strong influences are seen here associated with rollover, ejection (whether complete or partial), with entrapment, intrusion, and maximum crush. Recall that positive coefficients are associated with an increase of casualty probability that may be approximately, and one at a time, estimated via (11).

Such an estimation is illustrated in Fig. 6 for the strong influences cited above. Most other influences, as implied by their coefficients in Table I(E), are either weak when considered in conjunction with strong influences, or are not significant in view of the accompanyig standard error.

CONCLUSIONS

Research reported in this paper, although still not fully mature, provides useful applications in succinct and insightful descriptions of the car crash environment. In adition, several further applications may be anticipated. For example, practitioners of post-crash emergency care for traffic accident casualties, from emergency vehicle dispatch centers to trauma centers and hospitals, could benefit from ways of projecting casualty severity on the basis of a crash signature.

Such adjunct information could supplement anatomical and physiological information, currently used, in order to enhance the timeliness and appropriateness of emergency care decisions. This is especially applicable if the additional information becomes available at crash time, well before the emergency care complex is fully activated. Essentially, the advent of high tech, low cost sensors and electronics that may soon be carried on cars and other vehicles, could allow the transmission, at crash time, of crucial information concerning crash circumstances, crashed vehicle(s), and crash involved occupants.

Irrespective of electronic acquisition of crash signatures at crash time, as discussed above, very similar information could be retrieved and transmitted by emergency attendants after they reach the crash scene. It becomes then a matter of translating this information into data useful to dispatch and emergency teams, either ready to attend the accident scene, or preparing for appropriate treatment(s) in emergency facilities. This paper has addressed and evaluated the state-of-the-art in translating advance notification data, either from vehicle mounted sensors at crash time, or from emergency attendants at the scene.

ACKNOWLEDGEMENT

The authors, solely responsible for the analyses, evaluations, and conclusions presented herein, acknowledge thankfully that this investigation has been supported by the Federal Highway Administration's / National Highway Traffic Safety Administration's National Crash Analysis Centet at George Washington University

Table I. Logistic Regression Coeficients and Standard Errors

(A) Rel (6)	Fatality		MAIS 3+		MAIS 2+	
Predictor	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error
Intercept	-8.252	0.326	-5.450	0.157	-3.761	0.148
DELTAV	0.177	0.011	0.178	0.010	0.136	0.011

(B) Rel (7)	Fatality		MAIS 3+		MAIS 2+	
Predictor	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error
Intercept	-9.032	0.259	-5.820	0.139	-4.029	0.115
DELTAV	0.198	0.010	0.202	0.006	0.155	0.006
DOFS	1.462	0.189	0.556	0.099	0.465	0.083
DOFB	-1.921	0.343	-2.170	0.307	-1.177	0.245

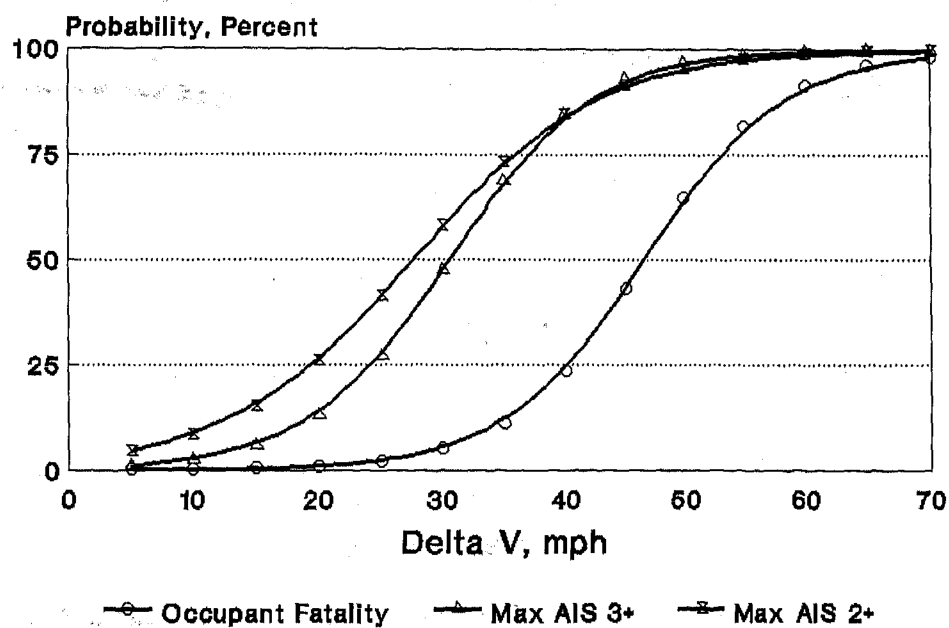
(C) Rel (8)	Fatality		MAIS 3+		MAIS 2+	
Predictor	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error
Intercept	-10.830	0.293	-6.538	0.213	-4.393	0.143
DELTAV	0.211	0.009	0.208	0.007	0.157	0.006
DOFS	1.301	0.196	0.457	0.101	0.403	0.087
DOFB	-1.716	0.303	-2.025	0.240	-1.060	0.194
AGE	0.051	0.004	0.033	0.002	0.026	0.002
BELT	-0.859	0.234	-0.826	0.098	-0.839	0.082
BEBA	-1.478	0.684	-1.564	0.325	-1.001	0.330

(D) Rel (9)	Fatality		MAIS 3+		MAIS 2+	
Predictor	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error	Ai Coeffcnt	Std Error
Intercept	-11.144	0.366	-6.710	0.211	-4.449	0.152
DELTAV	0.211	0.009	0.208	0.006	0.157	0.006
DOFS	1.298	0.203	0.432	0.102	0.397	0.087
DOFB	-1.781	0.317	-2.049	0.242	-1.068	0.195
ROLL	1.280	0.322	1.758	0.237	1.339	0.271
FRPAX	0.890	0.323	0.251	0.113	0.156	0.100
RRPAX	-0.186	0.350	0.263	0.231	0.016	0.149
AGE	0.053	0.005	0.035	0.002	0.026	0.002
BELT	-0.887	0.224	-0.811	0.095	-0.840	0.081
BEBA	-1.353	0.667	-1.518	0.327	-0.990	0.334

Table I Cont'd. Logistic Regression Coeficients and Standard Errors

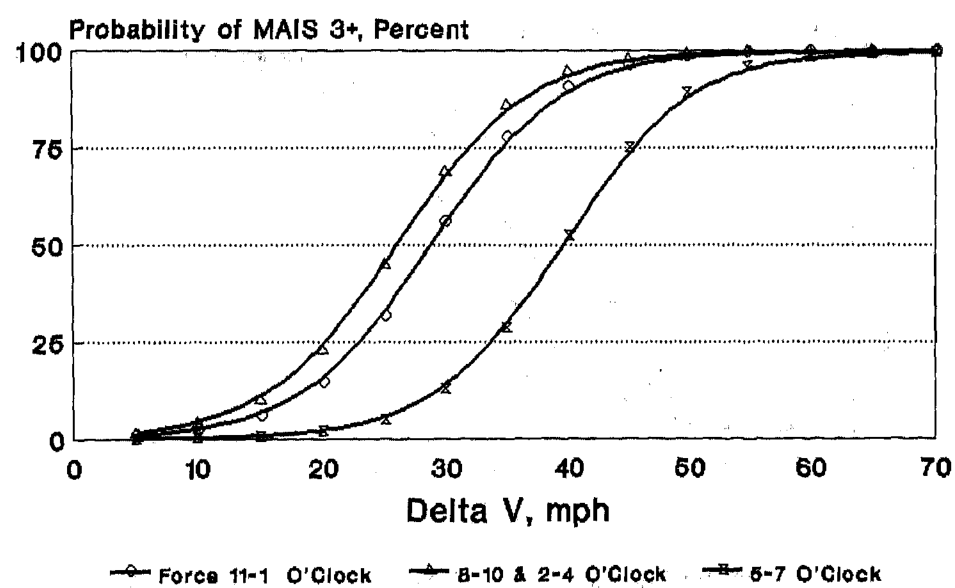
(E) Rel (10) Predictor	Fatality Ai Coeffcnt	Fatality Std Error	MAIS 3+ Ai Coeffcnt	MAIS 3+ Std Error	MAIS 2+ Ai Coeffcnt	MAIS 2+ Std Error
Intercept	-11.901	3.094	-6.118	0.703	-4.584	0.600
ONEVEH	0.259	0.230	0.322	0.152	0.155	0.134
BIGTRK	0.312	0.280	0.002	0.124	0.084	0.132
ROLL	0.764	0.357	1.157	0.284	1.086	0.316
DELTAV	0.135	0.014	0.164	0.010	0.126	0.010
GADSP	1.113	0.224	0.219	0.131	0.260	0.122
GADSNP	-0.076	0.496	0.057	0.241	0.256	0.183
GADB	-2.046	0.553	-1.793	0.254	-1.237	0.206
MAXC	0.056	0.011	0.037	0.007	0.039	0.007
INTRU	1.076	0.335	0.807	0.128	0.648	0.119
CURBWT	-0.012	0.031	-0.027	0.009	-0.031	0.008
FRPAX	1.034	0.341	0.232	0.138	0.017	0.118
RRPAX	0.348	0.347	0.103	0.203	-0.209	0.164
BELT	-0.512	0.275	-0.650	0.111	-0.691	0.093
BEBA	-1.341	0.405	-1.356	0.373	-0.698	0.400
AGE	0.060	0.005	0.042	0.003	0.030	0.003
FEMALE	0.334	0.268	0.464	0.115	0.396	0.102
OCCWT	0.002	0.004	0.003	0.002	0.001	0.002
OCCHT	-0.004	0.038	-0.014	0.011	-0.000	0.009
ENTRP	0.932	0.294	2.378	0.480	2.358	0.511
EJC	2.896	0.506	1.859	0.838	1.270	0.886
EJP	1.915	0.368	1.468	0.368	1.166	0.414

Fig. 1. Probability of Shown Outcome, as a Function of Car Crash Severity



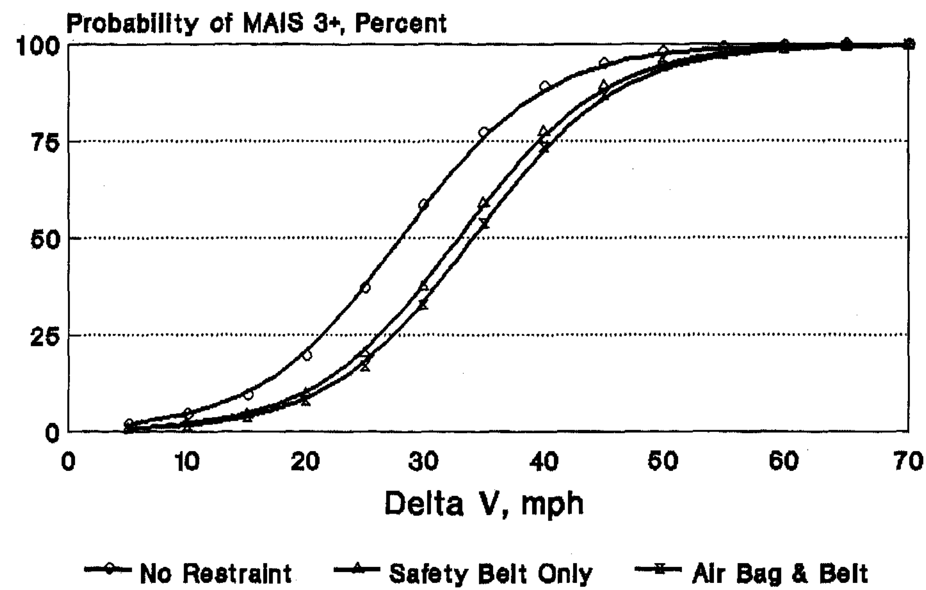
The NASS/CDS 1988-1995

Fig. 2. Sensitivity of Max AIS 3+ to the Direction of Force, as a Function of Car Crash Severity



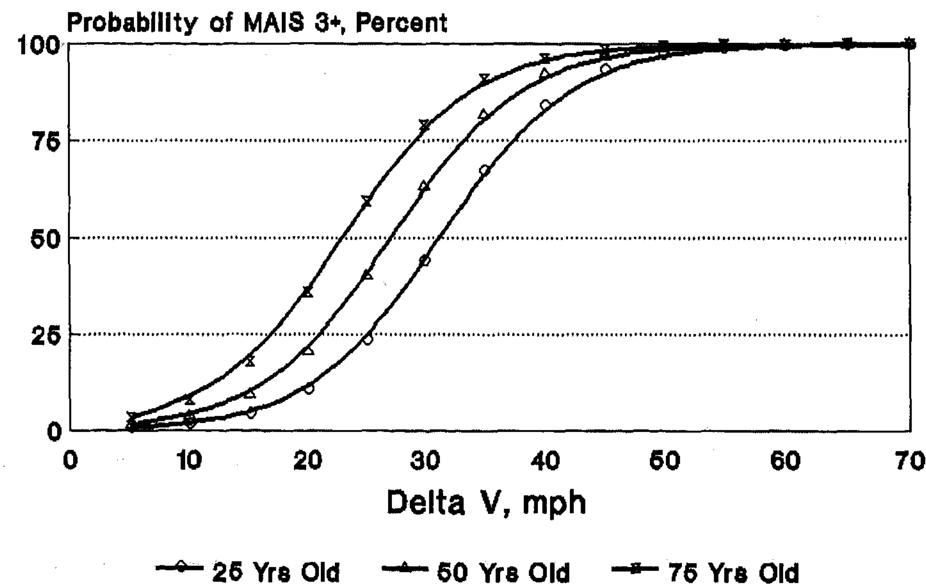
The NASS/CDS 1988-1995

Fig. 3. Sensitivity of Max AIS 3+
to an Occupant's Restraint Use and Type,
as a Function of Car Crash Severity



The NASS/CDS 1988-1995

Fig. 4. Sensitivity of Max AIS 3+
to an Occupant's Age,
as a Function of Car Crash Severity



The NASS/CDS 1988-1995

Fig. 5. Sensitivity of Casualty Probability to Shown Values of a Logistic Regression Coefficient

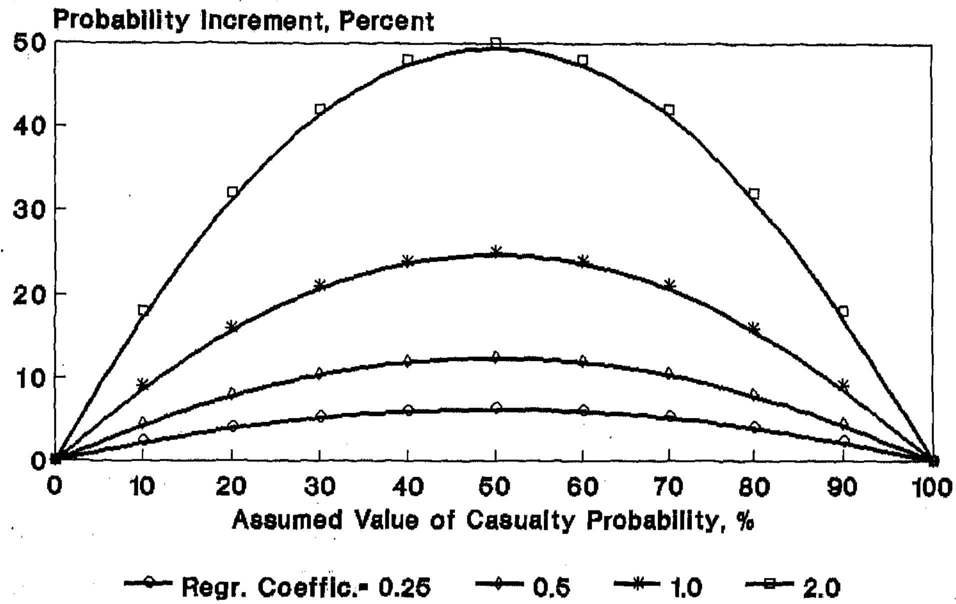


Fig. 6. Increase of MAIS 3+ Probability over Shown Base, Due to Shown Influences

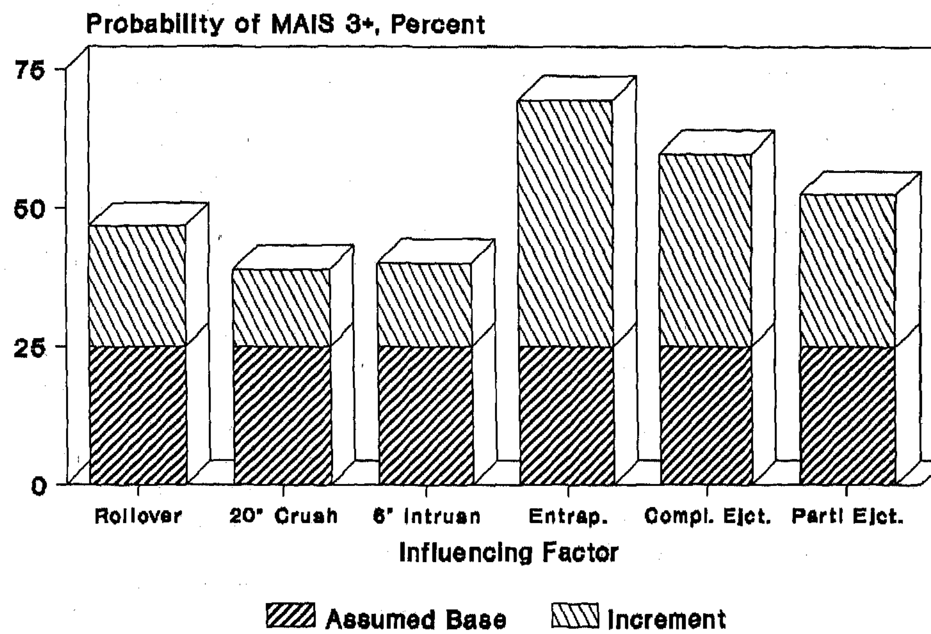


Table I(E), & Relation (11)

970396

Injury Mechanisms and Field Accident Data Analysis in Rollover Accidents

Michael B. James, Douglas "L" Allsop,
Ronald P. Nordhagen, and Robin L. Decker
Collision Safety Engineering, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Rollover accidents are responsible for a significant percentage of crash injuries. Increasing seat belt restraint use is the most effective way to reduce rollover injuries. Injuries to restrained occupants are also of interest. It has been suggested that head/neck injuries are caused by roof crush, and that modification to roof structures and seat belt systems would lead to a substantial reduction in severe rollover injuries. Field accident data and rollover testing are used to evaluate the relationship between roof crush, seat belt design, and severe rollover injuries.

INTRODUCTION

Rollover accidents are one of the most dangerous types of accidents, in terms of either fatalities or incapaciting injuries per occupant [NHTSA, 1992]. Table 1 compares the distribution of occupants, the distribution of Harm, and the Harm per occupant by accident type using 1988-1994 National Accident Sampling System (NASS) Crashworthiness Data System (CDS) data. Harm is a method developed by the National Highway Traffic Safety Administration (NHTSA) for weighting injuries according to their severity and their frequency [Mallarias, 1985]. Only 8% of occupants in this file were involved in a rollover, but their resulting injuries accounted for 17% of the total Harm. This phenomenon is attributable, at least partially, to the higher incidence of ejection in rollover accidents. Ejected occupants have a much higher likelihood of serious injury compared to occupants who are retained in the vehicle. The nature of rollover accidents result in more ejections than other accident modes.

SEAT BELT EFFECTIVENESS IN ROLLOVER ACCIDENTS

Seat belt restraints demonstrate their highest effectiveness in rollover accidents because they essentially eliminate the ejections that would otherwise occur. It has been estimated that rollover casualties could be reduced by as much as 75% if all occupants used safety belts [Felrice, 1992]. Evans, using Fatal Accident Reporting System (FARS) data, found an overall seat belt effectiveness of about 40% (in reducing fatalities) for front seat occupants [Evans, 1991]. The effectiveness in rollover accidents, however, was 68% for all vehicles that experience rollover, and almost 80% for vehicles whose first event was a rollover. The higher effectiveness in rollover accidents was attributed to the seat belt's ability to prevent occupant ejection. Table 2, (from Evans, 1991), shows seat belt effectiveness broken down into two components: 1) the portion of effectiveness attributable to ejection prevention, and 2) the balance of the effectiveness. This latter category represents the reduction of fatal impacts with the vehicle interior.

	OCCUPANTS	HARM	HARM PER OCCUPANT
FRONT	43 %	37 %	2.2
SIDE	23 %	30 %	3.4
REAR	8 %	2 %	0.8
ROLLOVER	8 %	17 %	5.9
OTHER	5 %	10 %	5.5
UNKNOWN	13 %	4 %	0.9
TOTAL	100 %	100 %	

TABLE 1
DISTRIBUTION OF OCCUPANTS, HARM, AND HARM PER OCCUPANT BY ACCIDENT TYPE [1988 - 1994 NASS CDS].

The dramatic influence of seat belt restraints on Harm reduction in rollover accidents is also evident in the NASS CDS 1988-1994 data. Table 3 presents the distribution of occupants, the distribution of Harm, and the Harm per occupant for vehicles which have rolled over. While 61% of the occupants in this sample were belted, their injuries accounted for only 27% of the Harm.

	OVERALL FATALITY REDUCTION	REDUCTION FROM EJECTION PREVENTION	REDUCTION FROM IMPACTS WITH VEHICLE INTERIOR
ALL ACCIDENTS	41%	18%	23%
ROLLOVER ACCIDENTS	68%	50%	18%
ROLLOVER AS FIRST EVENT	80%	64%	16%

TABLE 2
SEAT BELT EFFECTIVENESS "IN FATALITY REDUCTION," FARS DATA, EVANS, 1991.

	OCCUPANTS	HARM	HARM PER OCCUPANT
UNRESTRAINED	39%	73%	11.1
RESTRAINED	61%	27%	2.6

TABLE 3
DISTRIBUTION OF OCCUPANTS, HARM AND HARM PER OCCUPANT IN ROLLOVERS BY RESTRAINT USE.
[1988 - 1994 NASS CDS]

Partyka evaluated serious and fatal injuries to rollover occupants in the 1982-1989 NASS and CDS [Partyka, 1991]. As expected, the effect of seat belts was dramatic. Unrestrained occupants were 4.7 times more likely to be seriously injured than restrained occupants. In addition to preventing ejection, the data indicated that seat belt restraints also provide substantial benefits to non-ejected occupants in rollovers. Unrestrained, non-ejected occupants received serious injuries six times more often than restrained occupants. In fact, the number of serious and fatal injuries to belted occupants was so low that detailed analysis was difficult. For the 7 years of NASS data studied, there were only 77 restrained rollover occupants with AIS 3+ injuries and detailed injury data, and only 51 restrained, unejected, fatal occupants with injury information. For these 128 occupants, the most frequently injured body region was the torso, followed in order by the head, limbs, and the neck. The 5 most frequently occurring combinations of injured body region and its vehicle contact source are shown in Table 4.

Partyka noted that the restraint system itself was the highest ranked injury/contact source. However, it was noted that most of the restraint system induced injuries occurred in very severe accidents where unrestrained occupants would likely have received injuries at least as severe. Table 5 presents national annual estimates of serious injuries to belted occupants in rollovers, (from Partyka, 1991). Note that less than 8% were head or neck injuries from roof area contacts.

ROOF INTRUSION AND RESTRAINED OCCUPANTS IN ROLLOVERS

Much attention continues to be focused on the effect of roof crush and restraint system design on head/neck injuries, even though these injuries account for only a small

TORSO INJURY BY RESTRAINT SYSTEM CONTACT	13%
TORSO INJURY BY STEERING WHEEL CONTACT	13%
TORSO INJURY BY SIDE INTERIOR SURFACE CONTACT	10%
HEAD INJURY BY ROOF CONTACT	9%
HEAD INJURY BY WINDSHIELD CONTACT	7%

TABLE 4
INJURED BODY REGION AND VEHICLE CONTACT SOURCE OF INJURY SERIOUS AND FATAL INJURIES IN ROLLOVERS
[1982 - 1989 NASS, PARTYKA 1991]

	INJURED BODY REGION					
INJURY SOURCE	HEAD	NECK	LIMB	TORSO	WHOLE	TOTAL
FRONT AREAS	28	0	99	179	0	306
SIDE AREAS	94	0	82	260	0	436
INTERIOR AREAS	9	0	17	243	0	269
ROOF AREAS	120	47	10	4	0	181
FLOOR AREAS	0	0	23	0	0	23
EXTERIOR AREAS	73	0	28	0	0	101
NON-CONTACT	47	0	0	38	12	97
UNKNOWN	213	116	136	426	0	891
TOTAL	584	163	395	1150	12	2304

TABLE 5
BODY REGIONS AND INJURY SOURCE
SERIOUS INJURIES IN LIGHT VEHICLE ROLLOVERS
NON-EJECTED, RESTRAINED OCCUPANTS
[1982 - 1989 NASS CDS, ANNUAL NATIONAL ESTIMATE]

portion of the severe injuries experienced by restrained occupants in rollover accidents. It has been suggested that if restraint systems were designed to keep the pelvis tight against the seat cushion, minimizing vertical head excursion, there would generally be sufficient head room to prevent head to roof contact in rollover accidents, thus significantly reducing rollover injuries.

The actual relationship between roof intrusion and occupant injury is much more complex. Intrusion generally refers to a reduction in occupant compartment volume due to collision deformation. Intrusion can occur in all accident modes (frontal, side, rear, rollover etc,) and in any primary direction (vertical, lateral or longitudinal). Intrusion may be causally related to injury in four ways:

1. The intruding surface can move to a new location where there may be an occupant contact that would not otherwise occur.

2. The intruding surface can change shape or increase in stiffness, which may result in a more hostile contact configuration.

3. The occupant can contact the intruding surface while intrusion is taking place, increasing the local contact velocity between the occupant and the intruding surface.

4. In extremely severe cases, the intrusion can "crush" or "squeeze" the occupant between two surfaces.

While intrusion may be causally related to injury in some instances, it is somewhat misleading to talk about intrusion when the deformation is due to contact with a fixed object such as a tree or the ground. When "intrusion" occurs as a result of contact with a fixed object, the contact velocity between the occupant and the contact surface does not increase. The intruding surface does not move to a different location. It simply stops against the ground, tree, or other fixed object. When a vehicle goes sideways into a tree, the door comes to a stop against the tree, while the rest of the vehicle continues to move and deform around the tree. Thus, the door does not move or intrude into the occupant space, it is at rest. The occupant impacts and comes to a stop against the arrested door and/or the tree. The velocity of the occupant relative to the tree is simply the occupant's pre-impact velocity, regardless of whether or not the vehicle sustains "intrusion". This contact velocity is not increased as "intrusion occurs," nor is the location of the occupant contact altered by intrusion. In fact, if an injury occurs, it likely occurs before the "intrusion" takes place. The same is true of roof intrusion caused by ground contact. The roof hits the ground and stops in the vertical direction; the occupant hits the ground with the roof structure sandwiched between. Other parts of the vehicle often continue to move toward the ground resulting in "roof crush," but neither the occupant/roof contact velocity nor the location of contact are affected by the subsequent deformation.

FIELD ACCIDENT DATA, ROOF CRUSH, AND INJURIES

If there is a causal relationship between roof crush and head/neck injury, one would expect to find a correlation in the field accident data. However, establishing a <u>causal relationship</u> between intrusion and injury requires more than just finding a <u>correlation</u> [Strother, 1984]. There are many confounding factors in current field accident data bases which make it difficult to determine whether a correlation exists. The most significant limitation is the inability to account for differences in rollover crash severity. It is reasonable to assume that the likelihood of injury in a rollover accident increases as the accident severity increases. Digges and Klisch attempted to establish a measure of accident severity by evaluating the hard copy files of 140 single vehicle rollover accidents contained in the 1988-89 NASS CDS [Digges, 1992]. Factors such as roll type, trip force and roll rate at trip were evaluated. No reliable measure of rollover crash severity was found which could distinguish the 140 cases studied, or which could be used to distinguish rollover accidents in general.

Over the years, many field accident data studies have looked for, but have not found, a correlation between roof crush and injury [Mackay, 1970; Huelke, 1972; Wilson, 1972; Provensal, 1982; Strother, 1984; Plastiras, 1985; Biard, 1989]. Three recent studies claim that a correlation does exist, and claim it is evidence of a causal relationship. These studies, as well as two additional recent studies which found no correlation, will be briefly reviewed.

Friedman and Friedman performed a "limited" analysis of the 1982-1983 NASS data and claimed to have found a 4 to 1 increase in the risk of injury when the occupant is in the proximity of roof crush [Friedman, 1985]. It was acknowledged by the authors that prior evaluations of field accident data had not found a correlation; they attributed their unique findings to a methodology that matched the location of roof crush to the location of the occupant. Their methodology used the Collision Deformation Classification (CDC) to categorize roof crush location [SAE, 1980]. The intent was to isolate occupants who were seated in the area of roof intrusion and to determine whether they had a higher injury exposure than occupants who were not seated in the area of roof intrusion. However, essentially all of the occupants were categorized as "near" the roof crush. This methodology did not distinguish "near" and "not near" occupants -- they were virtually all categorized as "near." Careful analysis revealed that the "correlation" found by Friedman was not the result of matching the location of roof crush with the occupant's location, it was simply a result of failing to take into account important variables such as ejection, seat belt use, injury type, and accident severity [James, 1995].

More recently, Rains and Kanianthra evaluated head, neck and face injuries, and their relationship to headroom reduction [Rains, 1995]. The study was limited to non-ejected, properly restrained front seat occupants in vehicles which rolled at least 180 degrees and experienced no additional impacts. A total of 155 such cases were extracted from the 1988-92 NASS CDS data. Only head, neck and face injuries from contact with upper interior sources were evaluated. A methodology was developed to estimate headroom reduction, using vehicle measurements, occupant's recorded height, and roof intrusion measurements. There were too few data to evaluate the magnitude of headroom reduction vs. injury severity, therefore, occupants were categorized as either "injured" or "uninjured." Almost 80% of the "injured" occupants experienced only AIS 1 injuries. A correlation was suggested between these minor injuries and headroom reduction, although the authors noted that the confounding effects of other important factors, such as accident severity, were not accounted for and might be responsible for the correlation.

The intent of the previously-referenced 1992 Digges study was to establish some measure of crash severity for rollover accidents so that the effect of factors such as roof

crush could be evaluated. They were unable to establish a crash severity measure, but curiously found that: "There seems to be a correlation between roof crush and fatality rate, and it is likely that strengthening the roof would aid in reducing fatalities in rollovers" [Digges, 1992]. The "correlation" was established by grouping 173 non-ejected occupants from the 140 serious accident cases into 2 groups: occupants in a vehicle with top damage CDC extent 4-5, and occupants in a vehicle with top damage CDC extent 6-7. The "correlation" is based on the fact that 5% of the occupants in the CDC extent 4-5 vehicles received fatal injuries, while 20% of the occupants in the CDC extent 6-7 vehicles received fatal injuries. There were only 14 non-ejected fatal occupants in the study, and only 2 of them were restrained. 85% of the fatalities used to establish the "correlation" were not restrained. There is no information provided about the injuries, or whether the injuries might be related to roof crush. More importantly, there was no measure of accident severity with which to categorize these cases. All of these accidents involved severe roof crush, CDC extent 5 being roof crush to the level of the window sill. The most interesting finding from this study was not this alleged "correlation" but the finding that, in these vehicles with roofs crushed down to the beltline and beyond, the average injury severity for non-ejected, restrained occupants was only AIS 1.2 (minor-moderate injuries).

In a second study by Partyka, the injuries experienced by drivers in vehicles with roof intrusion were evaluated using data from the 1988-1990 NASS CDS [Partyka, 1992]. Vehicles were evaluated that rolled at least 180 degrees, had roof intrusion in the driver's seat area, and in which the driver received some injury from contact with the intruding roof. Approximately 80% of the drivers in these cases received only minor injuries. In fact, most injuries from contact with the roof, when the intrusion magnitude was less than twelve inches, were minor injuries. Only in cases when roof intrusion magnitude was greater than twelve inches were the majority of the injuries severe. In the three years of accident data evaluated, there were 918 drivers in vehicles that had rolled at least 180 degrees and had vertical roof intrusion. They represented 280,121 weighted drivers, of which only 0.6% had AIS 3+ injuries from contact with roof structures.

It was determined by Partyka that these data were insufficient to establish whether intrusion increased either the frequency of injuries or the severities of injuries. Several data limitations were noted: (1) the magnitude of roof crush likely reflects crash severity and there is no way to adjust injury rates for differences in crash severity, (2) it is not possible to estimate the number of times occupants contacted the intruding roof structures when injury did not occur, since only contacts resulting in injury are recorded, (3) it is not possible to determine the percentage of occupant contacts with non-intruded roof structures that do not result in injury, and (4) because injuries do occur when occupants contact non-intruded roofs, preventing roof intrusion may not necessarily prevent injury from roof contact. The author went on to note that "while these data describe the frequency of roof intrusion and the associated injuries, they do not indicate how reductions in intrusion frequency or severity would affect injury frequency and severity."

The most comprehensive study made to evaluate the correlation between roof crush and head/neck injury was performed by Moffatt and Padmanaban in 1995 [Moffatt, 1995]. This study examined 60,758 single vehicle rollovers, involving 96,154 occupants, using police reported data from Florida, Texas, Michigan and North Carolina. Using data from police reports avoids the problem of limited sample size in the NASS, but the cases provide less information with less reliable descriptions about the accident and the injuries. Nevertheless, what makes this study particularly unique is that the roof strength (from FMVSS 216 tests) was known for each of the accident vehicles.

Injury data from police reports is coded using the KABCO scale (K = fatal injury, A = incapacitating injury, B = non-incapacitating injury, C = possible injury, O = no injury). Each occupant injury is estimated by the police officer at the time of the accident. Only rarely is there a follow-up evaluation after the occupant has received medical treatment. Of the occupants in the police report data used by this study, 128 were also investigated in the NASS program. The reliability of police report injury data could therefore be evaluated by comparing the NASS injury data to the police report injury data for these 128 cases. The police report injury assessment agreed with the NASS injury assessment in 69% of the cases. Evaluation of the nearly 100,000 occupants found no relationship between roof strength and the likelihood of severe injury. Equally interesting was the finding that there was no relationship found between roof strength and the likelihood of severe roof damage.

CRASHWORTHINESS DATA SYSTEM (CDS) ANALYSIS

The second Partyka study [Partyka, 1992] examined the effect of roof intrusion in the first three years of NASS CDS (1988-1990) data. The CDS data contain information about the ten most severe intrusions in each vehicle. The information includes the location of the intrusion (by occupant seated position), the intruding component, the magnitude of intrusion, and the direction of intrusion (vertical, longitudinal or lateral). Injuries that are associated with any intruding component are also coded.

There are now 7 years of accident data available (1988-1994). There are 3,102 restrained, non-ejected occupants, 12 years of age or older, who were in rollover accidents. They represent a national estimate of 1,081,510 "weighted" occupants. Of these, 2744 occupants (990,363 weighted) were seated under vertical intrusion of either the roof or some other upper interior structure (roof side rail, front windshield header or back light header).

Table 6 summarizes these data and shows the number of occupants with severe (AIS 4+) injuries for each

category. The percentage of occupants with severe injuries is smaller for the subgroup of occupants who were seated under vertical roof intrusion. The Harm per occupant has also been computed as is shown in Table 6. Again, the

	RESTRAINED OCCUPANTS IN ROLLOVERS	RESTRAINED OCCUPANTS SEATED UNDER INTRUSION
ALL	1,081,510	990,363
MAIS 4+	9,555	6,743
% MAIS 4+	0.88 %	0.68 %
HARM PER OCCUPANT	2.61	1.74

TABLE 6
OCCUPANTS IN ROLLOVER ACCIDENTS (12+ YEARS OLD)
[NASS CDS 1988-1994]

occupants seated under vertical roof intrusion experienced a lower overall Harm. These findings are consistent with the numerous studies which have found no correlation between roof crush and injury. Note that these findings are based on the methodology proposed by Friedman, i. e. matching the occupant location to the intrusion location. However, in contrast to Friedman's analysis, the analysis of the new CDS data shows no correlation. Friedman's analysis did not separate restrained from unrestrained occupants, and was made using the older NASS data, which does not contain detailed intrusion information. A total of 41 occupants (6743 weighted) seated under vertical roof intrusion sustained a MAIS 4+ injuries. These cases were examined in more detail. The roof intrusion was not always coded as the source of the severe injury. In fact, only 16 occupants (1590 weighted) received their severe injury from contact with a roof that had sustained vertical intrusion. Table 7 summarizes the data for occupants seated under vertical roof intrusion. Only 0.2% received severe injuries from the intrusion. The data indicate that for restrained occupants in rollover accidents, there is no correlation between roof crush and injury. The data also show that injuries associated with vertical roof intrusion do not represent a large portion of severe rollover injuries.

	NATIONAL ESTIMATE	PERCENT
ALL OCCUPANTS	990,363	100.0 %
MAIS 0-3	983,620	99.3 %
MAIS 4-6	6,743	0.7 %
MAIS 4-6 (FROM INTRUSION)	1,590	0.2 %

TABLE 7
RESTRAINED OCCUPANTS (12+ YEARS) SEATED UNDER VERTICAL ROOF INTRUSION IN ROLLOVER ACCIDENTS
[NASS CDS 1988-1994]

TEST DATA

There are very few well-documented staged rollover tests with restrained dummies in which occupant kinematics and injury mechanisms can be evaluated. Real world rollovers include a large number of trip mechanism variations. It is difficult to simulate all of these trip mechanisms during testing. The dolly rollover procedure developed for FMVSS 208 is most often used because it provides a measure of control over the trip location and the vehicle roll direction. It does not, unfortunately, simulate the occupant motion prior to and during the trip phase. In spite of the controlled initial conditions and trip mechanism, the timing and location of specific vehicle/ground contacts and the resulting roll motion of the vehicle during staged testing are not repeatable. Another limitation of rollover testing is that the biofidelity of dummies in the rollover test mode has not been established, nor has there been development of appropriate vehicle and dummy instrumentation to measure all of the variables of interest. Nevertheless, full scale rollover tests with restrained dummies do provide valuable insights about general occupant kinematics and injury mechanisms.

The most comprehensive rollover test series was conducted by Bahling, et al. and is often referred to as Malibu II [Bahling, 1990]. A total of 8 1983 Malibu 2-door vehicles were rolled using the FMVSS 208 rollover procedure. Four tests were performed using vehicles equipped with production roof structures and four tests were performed using vehicles with roll cage reinforced roof structures. In each test, instrumented Hybrid III 50th percentile male dummies in the front outboard seats were restrained with the production 3-point belt system. Belt slack was added so that the dummy's static vertical excursion in an inverted vehicle was essentially the same as that measured when three human volunteers were inverted in the same vehicle. The average excursion measured with the volunteers was 9.9 cm (3.9 in.). All of the tests were passenger side leading, with a velocity of 51.5 km/hr (32 mph). The roll motion in each test was different, ranging from 3 to 4 rolls; the run out distance varied from 22.6 meters (74 feet) to 31.1 meters (102 feet). A pronounced difference in the roll motion was noted between the production roof vehicles and the roll cage vehicles. During the first passenger ground contact with the passenger side roof rail, the roll cage vehicle roof deformed less causing the vehicle's center of gravity to follow a higher trajectory as the vehicle continued its roll. This resulted in the driver's side roof rail of the reinforced vehicles making fewer ground contacts compared to vehicles with production roofs.

Immediately after trip, while the vehicle was airborne, centrifugal force caused the dummy to move upward and outward to the extent allowable by the lap belt and vehicle interior. Both dummies' heads tended to remain against the roof side rails, and moved away only when the vehicle's velocity was altered by significant ground contact. When the roof contacted the ground, the head adjacent to that side of vehicle generally maintained its orientation relative to the roof and roof rail, and was decelerated with the portion of the roof in contact with the ground. The head on the opposite side from the ground contact generally moved towards the impacted side of the vehicle. Wheel

contacts with the ground also altered the vehicle's velocity and caused dummy motion. For example, when the right side tires contacted the ground, the near-side passenger dummy leaned against the door, and often tried to go out the window, while the off-side driver dummy moved laterally inboard, rotated out of the shoulder belt, and loaded the lap belt. The buttocks of both dummies came off the seat cushion early in the test and never returned to the original seated position until the end of the test. There was no indication of a "squeezing" injury in any of the tests, regardless of the extent of roof crush.

In comparing the performance of the production roof vehicles to that of the roll cage vehicles, the researchers considered a peak neck compression load in excess of 2000 N as a "potentially injurious impact (PII)." Table 8 shows the distribution of PII's using this criterion. The higher frequency and severity of trailing roof rail PII's to the driver dummy are the consequence of the substantially different vehicle roll motion between vehicle types, which results in fewer trailing roof rail ground contacts to the roll cage vehicles.

The established Hybrid III neck injury criteria for compression loading is not based on peak neck load; it is a function of both the magnitude of the load and the duration of the load. See Figure 1 [Mertz, 1994]. Using the established Hybrid III injury criteria as the basis for PII produces the distribution shown in Table 10. There is no significant difference between the number of PII's with each vehicle type. However, comparing the frequency and severity of PII's in the production vehicles to the roll cage vehicles should not be used as the sole basis of evaluation. The stark difference in frequency and severity of roof contacts experienced by the two vehicle types can make direct comparison of PII frequency misleading. Table 11 shows the number of missed roof contacts and the number of roof contacts with no associated PII's for the production roof and roll cage vehicles. One would expect to see more PII's with the production roof vehicles because they experienced more roof/ground contacts.

Proper analysis of these tests would categorize the specific neck loading circumstances experienced by the dummies in both types of vehicles and determine whether

	PASSENGER LEADING ROOF RAIL	LEFT ROOF PANEL	DRIVER TRAILING ROOF RAIL	NON ROOF-TO-GRND CONTACTS	TOTAL
PRODUCTION ROOF	7	3	11	1	22
ROLL CAGE ROOF	13	2	2	1	18

TABLE 8

MALIBU II TEST SERIES:
DISTRIBUTION OF POTENTIALLY INJURIOUS IMPACTS
USING PEAK NECK COMPRESSION LOADS OF 2000 N

Friedman has presented a different analysis of the Malibu II data in which he claims that the performance of the production roof vehicles was substantially worse than the performance of the roll cage vehicles [Friedman, 1991]. He used a peak neck compression load of 4000 N as the criterion for a PII. He also evaluated the data using a criterion of 5000 N peak neck compression load plus 1000 N of concurrent shear force. Table 9 summarizes the PII's using these two criteria. Friedman claimed that his analysis of the Malibu II data demonstrates a causal relationship between neck injury and roof crush, since there are more PII's with production roof vehicles as compared to the roll cage vehicles.

roof crush made, or could potentially make, the loading circumstance more frequent or more severe. The preliminary analyses indicate that neck loadings can occur: 1) when there is no roof crush, 2) when there is only a few inches of roof crush, and 3) when there is substantial roof crush. There are only 2 instances (3L5 and 7L4) of neck loading with substantial roof crush (i. e., static crush in excess of 110 mm, 4.3 in) in the Malibu II test series. A detailed timing analysis of both of these instances was presented in the original study. The instrumentation and high speed film indicated that the neck loads occurred prior to the roof deformation. Therefore, the roof crush, even though substantial in magnitude, did not increase the frequency or severity of the loading circumstance.

	PASSENGER LEADING ROOF RAIL		DRIVER TRAILING ROOF RAIL		TOTAL	
	4000 N	5000 N + 1000 N SHEAR	4000 N	5000 N + 1000 N SHEAR	4000 N	5000 N + 1000 N SHEAR
PRODUCTION ROOF	3	0	11	5	14	5
ROLL CAGE ROOF	4	1	2	0	6	1

TABLE 9

MALIBU II TEST SERIES:
DISTRIBUTION OF POTENTIALLY INJURIOUS
IMPACTS USING PEAK NECK COMPRESSION
LOADS OF 4000 N AND 5000 N + 1000 N SHEAR

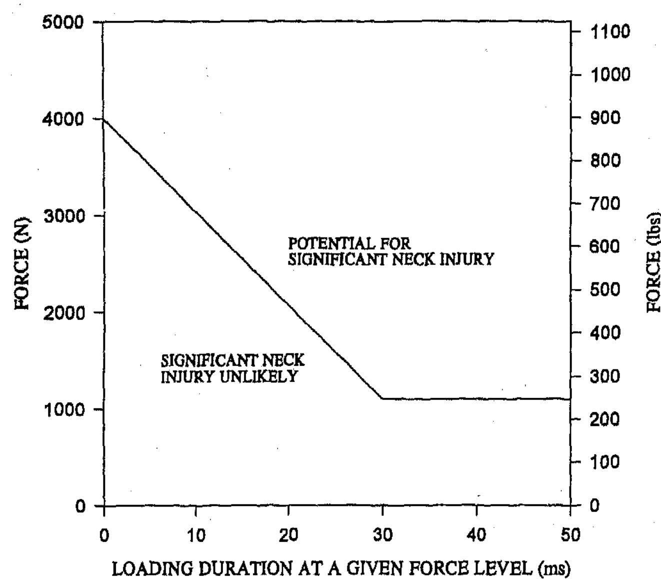


FIGURE 1
AXIAL COMPRESSION NECK
LOADING INJURY
[MERTZ 1983]

	PASSENGER LEADING ROOF RAIL	DRIVER TRAILING ROOF RAIL	TOTAL
PRODUCTION ROOF	10	13	23
ROLL CAGE ROOF	16	5	21

TABLE 10
MALIBU II TEST SERIES:
DISTRIBUTION OF POTENTIALLY INJURIOUS IMPACTS
USING ESTABLISHED H3 INJURY THRESHOLD

	PRODUCTION ROOF VEHICLES	ROLL CAGE VEHICLES
MISSED ROOF CONTACTS	5	15
ROOF CONTACTS WITH NO ASSOCIATED PII	6	6

TABLE 11
MALIBU II TEST SERIES

Another series of staged rollover tests was conducted by Sakurai et al. [Sakurai, 1991]. The series consisted of 12 ramp rollover tests using three different model vehicles. Each test was run at 50 km/hr (31 mph). The roll motion was developed by driving the right side tires onto a ramp while quickly generating right steer at the steering wheel, resulting in the vehicles rolling left side leading. Nine of the tests rolled one revolution, one test was one-half revolution, and the remaining two tests were one and one-half revolutions. The front seat outboard positions had instrumented Hybrid III dummies, restrained with three-point belts. As in the Malibu II tests, the centrifugal forces caused the dummies to move upward and outward to the extent permitted by the belts, even though the vehicle kinematics and trip mechanism were different for the two tests. The right dummy head moved into this position and remained there until the end of the test. The left dummy also moved upward and outward until its head was in contact with the roof rail. As the left roof rail contacted the ground, it stopped; the head also stopped, while the torso continued towards the head, axially loading the neck. The high speed camera and dummy instrumentation indicated that the peak neck compression load occurred prior to any substantial roof crush, a finding similar to that of the Malibu II test series.

A third series of rollover tests with restrained dummies was reported by Friedewald, [Friedewald, 1996]. These tests were conducted using the FMVSS 208 dolly rollover procedure. The vehicle instrumentation and high speed film were used to determine the contact forces between the ground and the vehicle roof structure. They found that the peak contact forces were similar in tests where the magnitude of roof crush was substantially different. The difference in roof crush magnitude was a function of contact duration, and not contact force. A series of tests were run with a production vehicle, and similar prototypes with both reinforced and weakened roof structures. The reinforced roof structures generated higher contact forces, and actually experienced larger crush magnitudes compared to the production and weakened roof structures. The roof crush magnitudes were a function of the rollover kinematics; reinforcement of the roof increased the contact forces, but had no influence on the roof deformations. (This finding is consistent with the finding by

Moffatt and Padmanaban, where the field accident data showed no correlation between roof strength and roof deformation.)

The Friedewald tests demonstrated dummy kinematics similar to that seen in both the Malibu II and Sakurai tests; the centrifugal forces caused the heads to move outward and upward until they are in contact with the roof rail above the door. The highest dummy loading occurred when the vehicle's rocker panel made ground contact, as opposed to roof/ground contact. The dummies in the two tests presented in the paper received loadings representing only minor injury.

HEAD EXCURSION WITH INVERTED VEHICLES - When a restrained occupant is inverted during a rollover accident, there will be motion of the head toward the roof. There are three forces acting upon the occupant to cause this motion: gravity, centrifugal forces, and vertical deceleration forces. The magnitude of vertical head excursions depends upon many factors including the restraint system geometry, belt slack, rotational velocity, vertical accelerations, occupant position, occupant body fat distribution, occupant spinal structure, etc. Vertical head excursion test data from two published studies, in addition to unpublished test data from Collision Safety Engineering, provide a basis for some preliminary assessments.

In 1995 Arndt et al. presented results of their efforts to minimize vertical head excursion in a rollover environment [Arndt, 1995]. Seven volunteers and a 95% percentile male Hybrid III dummy were statically tested in various configurations of lap belt angle, slack and pretension. The lap belts were not intended to represent an actual production restraint system. Rather, they were attached to an unpadded rigid seat that had no provisions for adjustablity. The lap belt was adjusted to either 2.5 cm (1.0 in.) of slack, 5.0 cm (2 in.) of slack, no slack with 64 N (14 lb) of pretension, 222 N (50 lb) pretension, or 444 N (100 lb) pretension. The force was applied by subjecting the subjects to a one G acceleration through simple inversion.

Two lap belt geometries were evaluated, one at 45 degrees and one at 90 degrees. A lap belt angle of 90 degrees is the most optimum angle for limiting vertical excursion, but it is inappropriate for frontal impacts. In fact, the FMVSS's do not allow lap belt angles of 90 degrees [CFR 571.210 S4.3.1.(b)]. The 444 N (100 lb) pretension is also "optimum" for limiting vertical excursion, but this load is much higher than what is typically used in production pretensioners. In fact, the 222 N (50 lb) pretension load caused enough pain for the human volunteers that no volunteer tests were performed with the 444 N (100 lb) pretension load. Therefore, the performance of the 90 degree lap belt with 444 N (100 lb) pretension is likely better than what can be expected of any practical production restraint system in terms of minimizing vertical excursion. Figure 2 shows the human volunteer and dummy excursion data for the 90 degree lap belt with 222 N of pretension and, the 45 degree lap belt with 222 N of pretension. The static

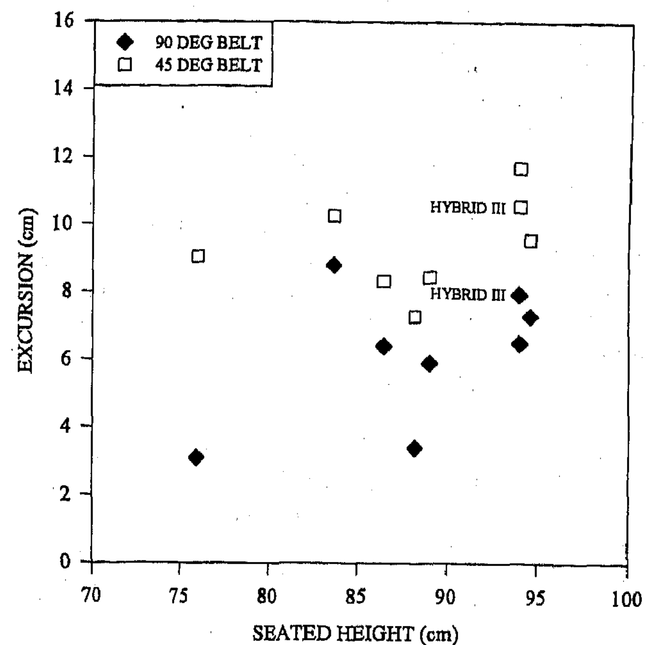


FIGURE 2
VOLUNTEER HEAD DISPLACEMENT VS. NORMAL SEATED HEIGHT
STATIC TEST, 222 N (50 lb) BELT TENSION
[ARNDT 1995]

Hybrid III dummy excursion for the 90 degree, 444 N (100 lb) pretension belt system is not given in the paper. Noting that the one G acceleration environment was unrealistically low compared to real world rollover accidents, the Hybrid III test dummy was subjected to a five G vertical acceleration pulse. The vertical head excursions were two to three times greater under these dynamic loading conditions. Figure 3 presents the dynamic head excursion for the 45 degree, 222 N system, the 90 degree, 222 N system, and the 90 degree 444 N system.

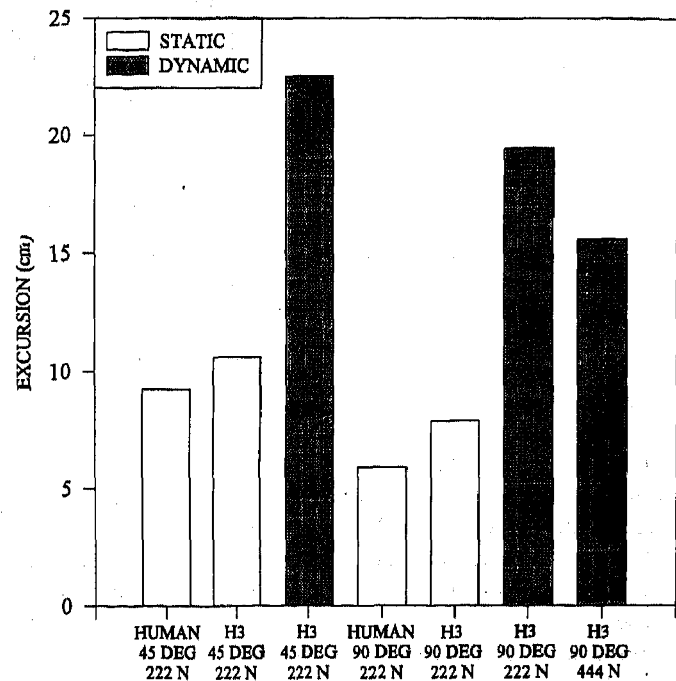


FIGURE 3
STATIC AND DYNAMIC EXCURSION
OF 95 % MALE HYBRID III DUMMY
[ARNDT 1995]

In 1996, Herbst et al. published the results of vertical head excursion testing [Herbst, 1996]. A seat buck was built to which different production vehicle seats and restraints systems could be attached. The buck was mounted in a rotisserie which inverted the seat for static excursion measurements, or the seat could be dynamically rotated at a rate of about 100 degrees per second. Three volunteers representing a 5th percentile female, a 50th percentile male, and a 95th percentile male were tested using three different production restraint systems. Belt slack was not measured, but the volunteers adjusted the belts to be comfortable. No webbing preload was used. Under the static one G acceleration condition, average occupant excursion for the small, medium and large volunteers was 14 cm (5.5 in.), 17 cm (7 in.), and 20 cm (8 in.) respectively. The geometry of each belt system was then reconfigured to minimize vertical head excursion. Figure 4 presents the static and dynamic vertical head excursion data for the production and the modified systems. Not surprisingly, the vertical head

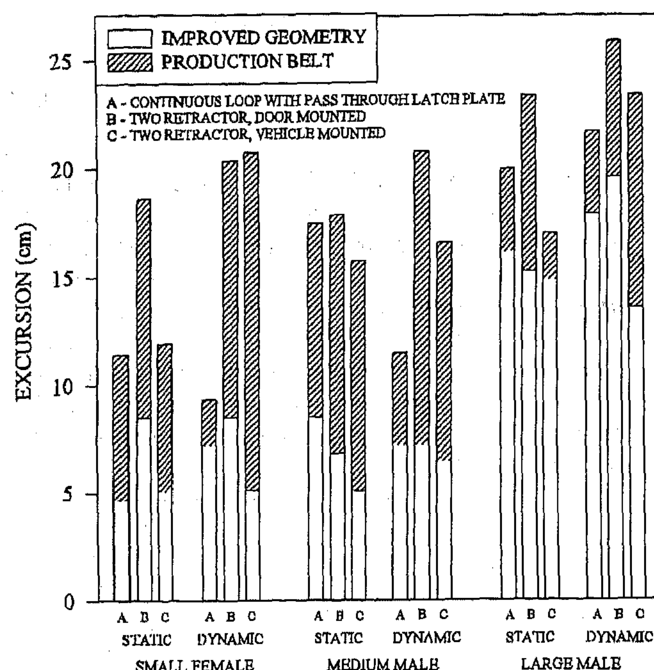


FIGURE 4
STATIC AND DYNAMIC VOLUNTEER VERTICAL EXCURSION TESTS
[HERBST 1996]

excursion magnitude was reduced by the modified system in every case. However, static excursions of 4.7 to 16.1 cm (1.9 to 6.3 in.) were still observed. There are some anomalies in the data which make the test procedure questionable. One would logically expect the dynamic excursions to be greater than the static excursions. However, the dynamic tests produced equal or less vertical head excursion than the static tests in about one third of the cases. One would also expect the larger volunteers to have greater excursions than the smaller volunteers. The 50th percentile male volunteer, however, had equal or less excursion than the 5th percentile female volunteer in one third of the tests. The dynamic data is difficult to interpret because the center of rotation and its relationship to each occupant is not given. The roll rate of 100 degrees per second is about one third of what was experienced in the Malibu II rollover testing. Due to the "squared" dependency of force upon rotation rate, one would expect the centrifugal forces to be approximately nine times higher on the occupants in a real world rollover situation. In the Arndt dynamic tests, only five times the force (5 G's) yielded two to three times the excursion. Two and a half times the modified excursion amounts in the Herbst tests yields excursion magnitudes of 12 to 40 cm (5 to 16 in.). These excursion magnitudes are sufficient for the head to make contact with the roof in almost all vehicles even without taking into account the impacts between the vehicle and the ground.

Ground Contact - Static rollover tests were conducted by Collision Safety Engineering, Inc., to analyze the relationship between occupant size, headroom, roof contact, and ground contact. The tests consisted of restraining human volunteers in several production vehicles and inverting the vehicle in a quasi-static roll motion. Test volunteers were a 5th percentile female and an 80th percentile male. It was observed that emergency locking retractors would lock before any significant occupant or seat belt movement had taken place. At the 135 degree position, an artificial ground plane was placed in contact with the vehicle roof rail as shown in Figure 5. The occupant was allowed to hang out the driver's door window, and the relationship of the occupant's head to the artificial ground

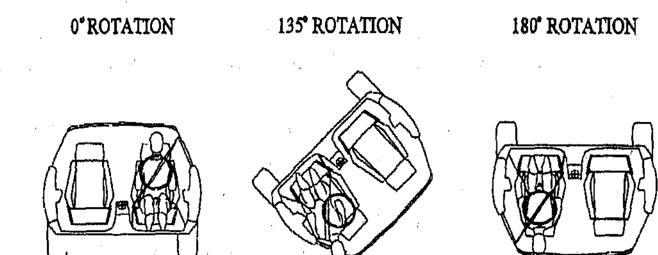


FIGURE 5
COLLISION SAFETY ENGINEERING
HEAD EXCURSION TEST PROCEDURE

was measured. In every case, the occupants head penetrated through the artificial ground plane from 1.2 cm (0.5 in.) to 6.5 cm (2.5 in.). Vehicle rotation then continued to 180 degrees, and the occupant was allowed to hang in the belts. In all cases involving the 80th percentile male, the occupant's head was in contact with the roof. The 5th percentile female contacted the roof in several instances, and was within an inch of the roof in all cases. This represented vertical head excursion of approximately 12.7 cm (5 in.) for each volunteer. Specific values, vehicle types, and measurements from the tests are listed in Table 12. Neither the occupant size nor the type of restraint system had any significant effect on the measured static excursion values. In essentially every tested configuration, the occupant's head went through the ground plane when the vehicle was at 135 degrees, and came in contact with the roof at 180 degrees, in a one G environment.

	VEHICLE ROTATION	0°	180°	135°	135°
SUBJECT*	VEHICLE	HEAD CLEARANCE	ROOF CONTACT OR CLEARANCE	HEAD / GRND CONTACT	PENETRATION THROUGH GROUND PLANE
F	83 TOYOTA TERCEL	15.2 cm	2.5 cm	YES	5.1 cm
F	83 FORD ESCORT	15.2 cm	CONTACT	YES	5.1 cm
F	83 CHEVROLET SUBURBAN	16.5 cm	3.8 cm	YES	1.3 cm
F	83 HONDA ACCORD	16.5 cm	3.8 cm	YES	1.3 cm
F	83 CHEVROLET S-10 PU	17.8 cm	3.8 cm	YES	3.8 cm
F	83 FORD F-150 PU	N/A	CONTACT	YES	2.5 cm
F	83 CHEVROLET SUBURBAN	N/A	2.5 cm	YES	5.1 cm
F	88 VOLVO 740	10.2 cm	CONTACT	YES	2.5 cm
M	87 CHEVROLET S-10 BLAZER	8.9 cm	CONTACT	YES	5.1 cm
M	87 HONDA ACCORD	2.5 cm	CONTACT	YES	5.1 cm
M	87 VW SCIROCCO	1.3 cm	CONTACT	YES	6.4 cm
M	87 FORD F-150 PU	10.2 cm	CONTACT	YES	5.1 cm
M	83 CHEVROLET SUBURBAN	8.9 cm	CONTACT	YES	3.8 cm
M	87 DODGE CARAVAN	6.4 cm	CONTACT	YES	5.1 cm

* F: 5th PERCENTILE FEMALE
M: 80th PERCENTILE MALE

TABLE 12
OCCUPANT ROOF AND GROUND CONTACT
UNDER ONE G ACCELERATION

Discussion - While it may be possible to reduce vertical head excursion through restraint system design, it does not appear likely that head contact with the roof and roof rail can be eliminated, especially under dynamic loading conditions.. Dynamic testing with dummies suggests that excursions in excess of 20 cm (8 in.) can be expected even with restraint systems designed to minimize excursion. Dynamic excursions of humans will likely be greater than those of dummies because the human spine can straighten and actually elongate under vertical loading. Pretensioners will likely have little effect on vertical excursion magnitudes. They are designed to take up slack in the belt prior to occupant loading. Typical pretensioners exert 100-200 N (22 - 45 lb) of belt tension. In a rollover accident, the pre-trip deceleration of the vehicle will most likely cause the occupant to load the lap belt prior to the actuation of any sensor triggered by vehicle roll angle. If the occupant loads the lap belt with the force level of the pretensioner prior to actuation there will be no pretensioner effect. Pretensioners are not designed to "pull" the occupant back into position. Such a pretension would produce excessive forces and input energy.

Even if it were possible to keep the head from contacting the undeformed roof and roof rail in dynamic loading conditions, some roof crush will occur even with stronger roof structures. In fact, the data suggest that strengthening roof structures may not have a significant effect on roof crush magnitudes. It must also be remembered that eliminating head/roof contact will not eliminate head contacts; the head will simply go out the window and be exposed to ground contact. The lateral motion of the head cannot be ignored.

It is important to assess the performance of restraint systems in accident modes other than rollover before modifications are proposed to reduce vertical head excursions. Considerations must include the overwhelming greater frequency of frontal impacts, occupant size variations, comfort and convenience and its effect on usage rates.

CONCLUSIONS

Prior evaluations of field accident data have found no correlation between roof crush and injury. Recent studies have claimed to find a correlation and suggest that it indicates the existence of a causal relationship. These studies are flawed by either failing to distinguish restrained from unrestrained occupants, or by failing to distinguish between minor and severe injuries. Analysis of the recent NASS CDS field accident data are consistent with the finding that no correlation exists between roof crush and occupant injury. The CDS field accident data also suggest that there are few serious injuries associated with roof intrusion.

Rollover testing with restrained instrumented dummies indicates that centrifugal forces cause the dummies' heads to move upward and outward until they are in contact with, or in close proximity to, the roof prior to roof contact with the ground. When the roof contacts the

ground, peak neck compression loads occur prior to any substantial roof deformation.

The magnitude of roof crush is a function of the vehicle roll motion, which is different in every rollover event. Both field accident data analysis, and rollover testing have suggested that stronger roof structures do not necessarily result in reduced roof crush.

Current production seat belt systems can be modified to reduce vertical head excursion, but it is unlikely that head contact with the roof and roof rail can be eliminated. If head to roof contact could be eliminated, lateral motion of the head out of the window opening will expose the head to contact with the ground.

BIBLIOGRAPHY

Arndt, M.W.; Mowry, G.A.; Dickerson, C.P.; Arndt, S.M.; Evaluation of Experimental Restraints in Rollover Conditions. SAE 952712. 39th Stapp Car Crash Conference. November 1995.

Bahling, G.S.; Bundorf, R.T.; Kaspzyk, G.S.; Rollover and Drop Tests - The Influence of Roof Strength on Injury Mechanics Using Belted Dummies. SAE 902314. 34th STAPP Conference. November 4, 1990.

Baird, R.; Cesari, D.; Block, J.; Rollovers of Passenger Cars in France. 12th International Technical Conference on Experimental Safety Vehicles, NHTSA, 1989.

Digges, K.; Klisch, S.; Final Report - TASK II (DTRS-57-90-C-00092) Crashworthiness Systems-Effectiveness In Rollover Crashes. January 2, 1992.

Evans, L.; Traffic Safety and the Driver. Van Nostrand Reinhold. 1991.

Felrice, B.; Letter to John Martin, National Correspondent, ABC News, Washington Capital Cities/ABC, Inc., 1717 DeSales Street, NW, Washington, D.C. 20036. March 18, 1992.

Friedewald, K.; An Analysis of Body Loads During Rollover Tests; Roof Crush and Occupant Protection.. Paper Number 96-S5-O-09. 15th Experimental Safety Vehicle Conference. May 13, 1996.

Friedman, K.; Friedman, D.; Statistical Analysis of Injured Occupants by Seating Position and Location of Roof Deformation. (Included with Docket Submission 92-28-No2-031-01); Liability Research. 1985.

Friedman, D.; Friedman, K.; Roof Collapse and the Risk of Severe Head and Neck Injury. 13th Experimental Safety Vehicle Conference. November 4, 1991. Number No. 91-S6-0-11.

Herbst, B.; Forrest, S.; Wang, P.; Chng, D.; Friedman, D.; Friedman, K.; The Ability of 3 Point Safety Belts To Restrain Occupants in Rollover Crashes. Paper Number 96-S5-O-12. 15th Experimental Safety Vehicle Conference. May 13, 1996.

Huelke, D.F.; Marsh, J.C.; Sherman, H.W.; Analysis of Rollover Accident Factors and Injury Causation. 16th Conference of the American Association for Automotive Medicine. 1972.

James, M.B.; Allsop, D.L.; Smith, G.C.; Comments on Claims of Causal Relationship Between Roof Crush and Injury. Docket Submission No. 92-28-N02-085. June 27, 1995.

Mackay, G.M.; Tampan, I.D.; Field Studies of Rollover Performance. SAE 700417. 1970.

Malliaris, A.C.; Hitchcock, R.; Hansen, M.; Harm Causation and Ranking in Car Crashes. SAE 850090. SAE International Congress and Exposition. February 25, 1985.

Mertz, H.; Anthropomorpic Test Devices. (Reprinted with permission from Accident Injury: Biomechanics and Prevention, Ed., Alan M. Nahum and John W. Melvin, Springer-Verlag, 1993) Hybrid III: The First Human-Like Crash Test Dummy. SAE PT-44. 1994.

Moffatt, E.A.; Padmanaban, J.; The Relationship Between Roof Strength and Occupant Injury in Rollover Accident Data. Failure Analysis Associates. May 1995

Partyka, S.C.; Roof Intrusion and Occupant Injury in Light Passenger Vehicle Towaway Crashes. Office of Vehicle Safety Standards, Rulemaking National Highway Traffic Safety Administration. February 18, 1992.

Partyka, S.C.; Light Passenger Vehicle Rollovers with Serious Occupant Injury. Office of Vehicle Safety Standards, Rulemaking National Highway Traffic Safety Administration. April 15, 1991.

Plastiras, J.K.; Lange, R.C.; McCarthy,R.L.; Padmanaban, J.A.; An Examination of the Correlation between Vehicle Performance in FMVSS 216 versus Injury Rates in Rollover Accidents. SAE 850335. 1985.

Provensal, J.; Discussion of SAE 820244. Crash Protection. SAE SP-513. 1982.

Rains, G.C.; Kanianthra, J.N.; Determination of the Significance of Roof Crush on Head and Neck Injury to Passenger Vehicle Occupants in Rollover Crashes. SAE 950655. International Congress & Exposition. February 27, 1995.

Sakurai, T.; Takigawa, Y.; Ikeno, H.; Study on Passenger Car Rollover Simulation. 13th International Technical Conference on Experimental Safety Vehicles. S6-O-10. November 4, 1991.

Strother, C.E.; Smith, G.C.; James, M.B.; Warner, C.Y.; Injury and Intrusion in Side Impacts and Rollovers.

SAE 840403. Advances in Belt Restraint Systems: Design, Performance and Usage. February 1984.

Strother, C.E.; Smith, G.C.; James, M.B.; Warner, C.Y.; Injury and Intrusion in Side Impacts and Rollovers. SAE 840403. Advances in Belt Restraint Systems: Design, Performance and Usage. SAE P-141, International Congress and Exposition. February 27, 1994.

Wilson, R.A.; Gannon, R.R.; Rollover Testing. SAE 720495. May 22, 1972.

970397

Measuring Head Restraint Force and Point of Application During Low-Speed Rear-End Automobile Collisions

Jonathan M. Lawrence, Gunter P. Siegmund, and Jeff S. Nickel
MacInnis Engineering Assocs.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

In order to assess the head and neck kinetics of human subjects exposed to low-speed rear-end impacts, a method for measuring the magnitude and line of action of the force between the head and the head restraint was required. In addition to being accurate and repeatable, the design was required to maintain original seat back and head restraint geometry, mass, stiffness, and height adjustment. This paper presents a design using strain gauges applied to the head restraint tubes, upper seat back, and custom replacements for brackets attaching the head restraint to the seat back. The background theory and free-body analysis, the analog math circuitry, and a dynamic calibration procedure are presented. Overall force magnitude and line-of-action errors are quantified, and a sample output from a human subject undergoing a rear-end collision with a speed change of 8 km/h is presented.

INTRODUCTION

As part of an investigation into the kinematic, kinetic and muscular response of the human head and neck during a low-speed rear-end automobile impact, a means of measuring the force between the head and head restraint was required. To solve for reaction forces and moments at the atlanto-occipital joint, the head restraint contact force and its line of action were required. Mertz[1] conducted a similar analysis using a test sled with a specifically-constructed seat made of rigid planar seat and seat-back panels joined by a rigid hinge connection capable of rotating after a pre-set torque was reached. Mertz used two axes of a single tri-axial load cell to measure the head contact force and rigidly mounted the load cell to the seat back. As noted by Mertz himself, the shortcoming of this instrumentation arrangement was the difficulty determining the point on the head through which the head restraint contact force acted. He visually estimated the contact point from high speed video, but noted that the error in this method was "...large, and hence the computed values of head torque for cases in which the headrest was used are dubious...."

A specific goal of our study was to test human subjects in an actual seat in an actual automobile. Head contact with the head restraint was expected in most tests and therefore a method of predicting head restraint force magnitude and line of action relative to the head was needed. Figure 1 shows free-body diagrams of the human head and head restraint in the sagittal plane.

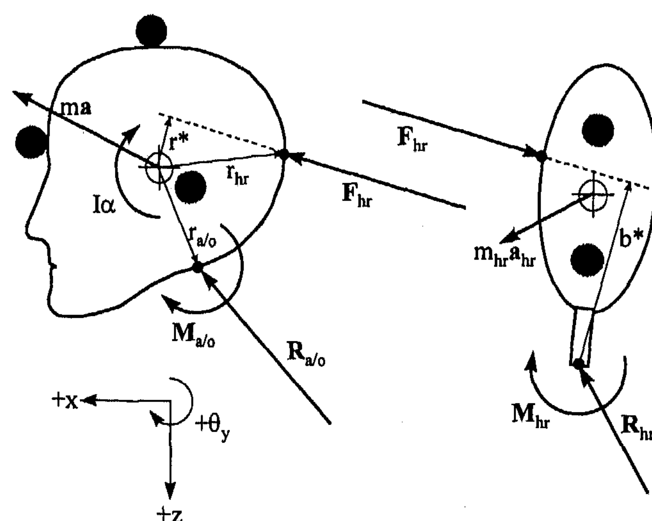


Figure 1. Free-body diagram of the head and restraint

To calculate $\mathbf{M}_{a/o}$ and $\mathbf{R}_{a/o}$, the following quantities are required: head mass m, head moment of inertia I, head linear acceleration at the center of mass $\mathbf{a}$, head angular acceleration α, the head restraint contact force $\mathbf{F}_{hr}$, and the position vectors from the center of mass to the head restraint contact point $\mathbf{r}_{hr}$ and the atlanto-occipital joint $\mathbf{r}_{a/o}$. Head accelerations can be determined from accelerometry and high speed video and resolved to an estimated position for the head center of mass. The head restraint force $\mathbf{F}_{hr}$ and the offset b* of its line of action is determined from measurements of the reaction force $\mathbf{R}_{hr}$ and reaction moment $\mathbf{M}_{hr}$ after inertial compensation for the head restraint mass. Using video targets (shown as gray circles in Figure 1) on both the head and head restraint, the line of action of the head restraint force can be mapped onto the head to determine $\mathbf{r}_{hr}$ or more importantly the moment arm r*. The other position vector $\mathbf{r}_{a/o}$ can be estimated from magnetic resonance (MR) films.

This paper describes a method of measuring the magnitude, direction and location of a force applied to an automobile head restraint. Although the kinetic analysis will be limited to the sagittal plane, the system was designed to measure three-dimensional forces and moments to confirm that non-sagittal-plane components were negligible. The design used strategically-applied strain gauges on the head restraint, head restraint supports, and upper seat back to build a custom 6-degree-of-freedom load cell using the seat structure. This type of design minimized changes to the head restraint geometry, mass, and stiffness, and allowed the head restraint height to remain adjustable.

The first part of this paper outlines the background and theory behind the design. Equations are then developed for two-dimensional and three-dimensional configurations as well as a least-squares calibration procedure. The accuracy and repeatability of the completed design are then assessed, followed by a sample data set from a rear-end collision with a speed change of 8 km/h.

Vehicle and Seat - The vehicle chosen for our tests was a 1990 Honda Accord LX sedan. Test subjects sat in the right front seat. A schematic of the Accord's upper seat back and head restraint is shown in Figure 2. The head restraint consists of a fabric cover over foam padding and a rigid plastic core. The top of an inverted U-shaped, chromed-steel tube is embedded in the rigid core and its two legs slide in plastic sleeves inserted into mounting brackets welded to the top of the seat back frame. The plastic sleeves clip into the metal brackets with tapered barbs on their lower ends. The outboard plastic sleeve is also screwed to the mounting bracket and has a spring-loaded latch which engages four notches in the U-shaped tube to lock the head restraint height. From the lower three latched positions, the head restraint can be adjusted up without depressing the latch. The inboard leg of the U-shaped tube was free to slide vertically in its plastic sleeve, although this sleeve was replaced with a locking sleeve for our tests.

BACKGROUND

Strain gauges are primarily sensitive to strains along one axis. However, strains along the gauge axis at the gauge location may result from a combination of internal stresses, i.e., axial and bending stresses in a beam (see Figure 3). Using strain gauges on the top and bottom of the beam, the strains due to the axial force and in-plane bending moment can be differentiated.

The axial force causes strains of equal magnitude and direction at both gauge locations, i.e., tension or compression. The bending moment causes strains of equal magnitude (assuming symmetrical cross-section), but opposite directions at each gauge location, i.e., one tensile and the other compressive. Adding the two strain gauge signals cancels the bending moment strains and leaves a combined strain proportional to the axial force. Subtracting the two signals cancels the axial force strains and leaves a differential strain proportional to the bending moment. Strains caused by out-of-plane bending can be minimized by mounting the gauges near the beam's neutral axis for out-of-plane bending.

Addition of strain gauge signals is achieved by wiring the two gauges to opposite arms of a Wheatstone bridge circuit (Figure 4a). Alternatively, subtraction is achieved by wiring the gauges to adjacent arms (Figure 4b). Therefore, pairs of gauges can be used to build analog axial-force or bending-moment transducers.

2-D Theory - Consider a two-dimensional head-restraint subjected to a force of unknown magnitude and direction applied at an unknown point on the head restraint. This configuration is shown in Figure 5 with strain gauges at positions 1 through 6. Free-body diagrams for sections through A and BC are shown in Figure 6. Equations for F_z and F_x as a function of strain are as follows:

$$F_z = k_1(\varepsilon_1 + \varepsilon_2) \quad (1)$$

$$F_x = k_2(\varepsilon_3 + \varepsilon_4) + k_3(\varepsilon_5 + \varepsilon_6) \quad (2)$$

where k_i's are constants which depend upon the cross-sectional area and the stiffness of the structure, and ε_i is the strain at gauge i. If the geometry and stiffness of the beam are the same at B and C, then $k_2 = k_3$.

The total applied force on the structure is the vector sum of F_x and F_z and the angle θ_y of the force is

$$\theta_y = \tan^{-1}\left(-\frac{F_z}{F_x}\right) \quad (3)$$

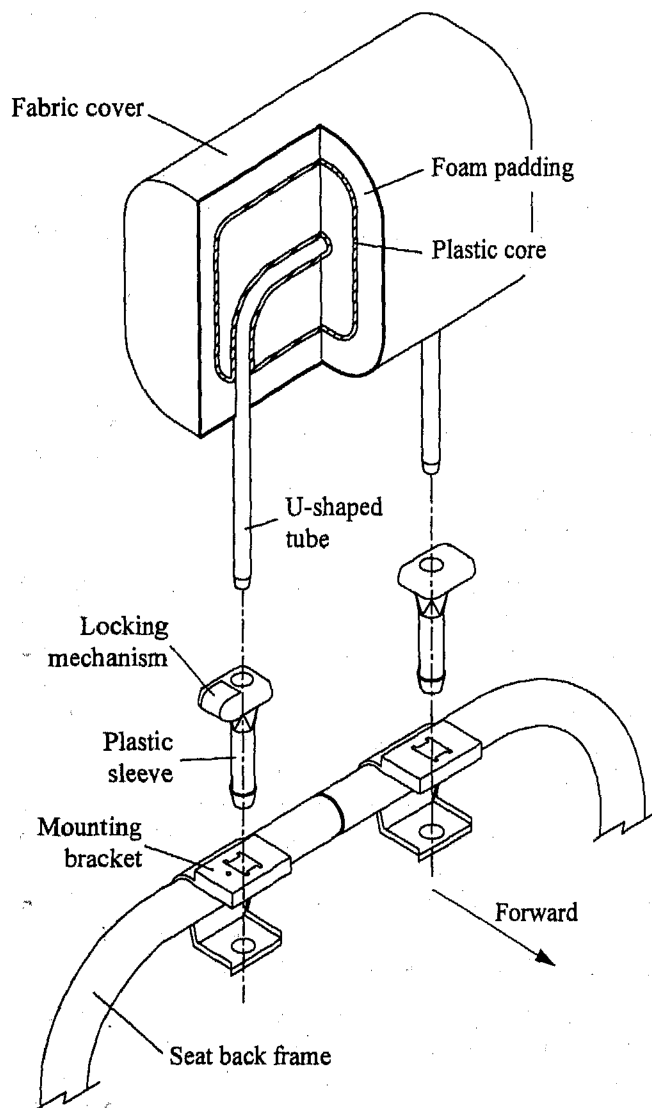


Figure 2. Honda Accord head restraint configuration. Locking height adjustment is present on the outboard side only.

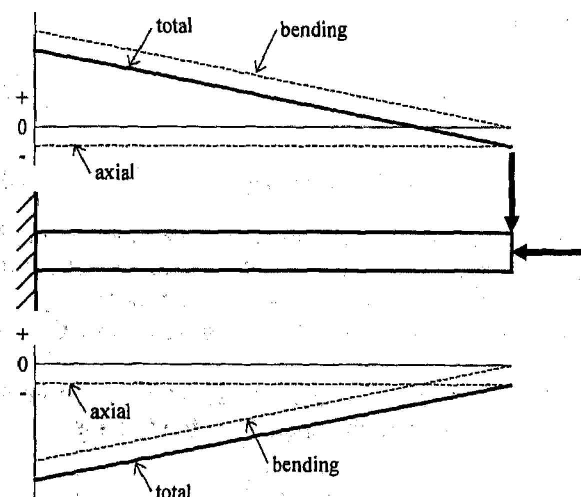
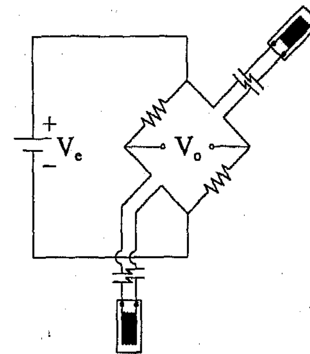


Figure 3. Strains on the top and bottom surfaces of a cantilevered beam subjected to combined compression and bending (+ve strain corresponds to elongation).



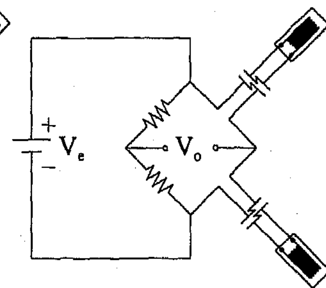


Figure 4a.
Axial Force Transducer

Figure 4b.
Moment Transducer

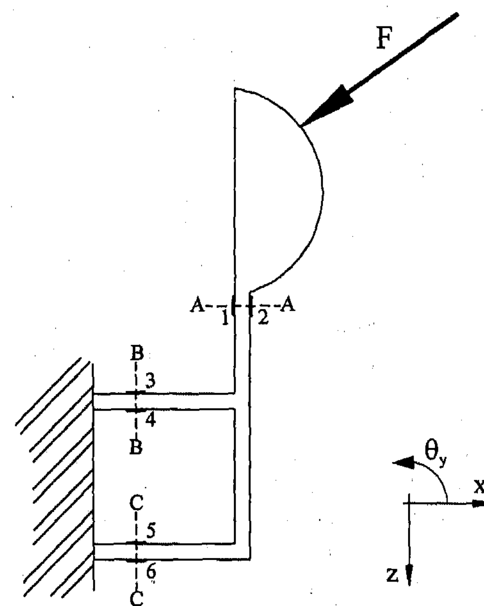


Figure 5. Two-dimensional schematic of Accord's head restraint

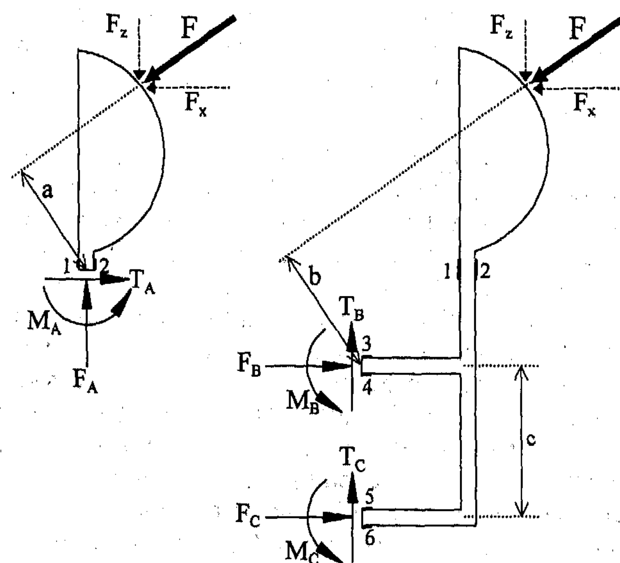


Figure 6. Sections at A and BC. Axial forces F, shear forces T, and bending moments M are shown.

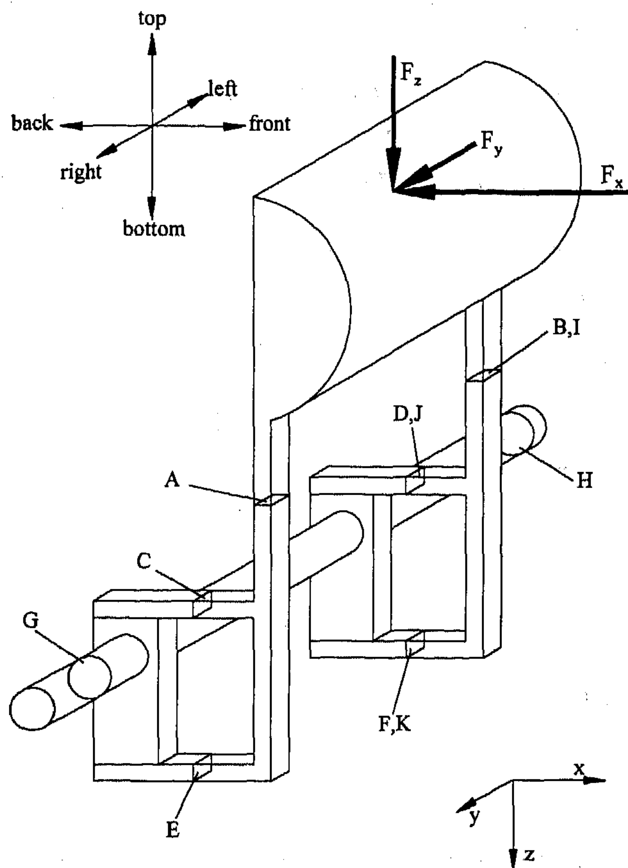


Figure 7. Three-dimensional schematic of the Accord's head restraint

Table 1. Position, sensitive axis, gauge number, and surface for each strain gauge mounted at positions A through K in Figure 7.

Pos'n	Axis	Gauge No.	Surface	Pos'n	Axis	Gauge No.	Surface
A	z	1	right	B	z	5	right
		2	left			6	left
		3	front			7	front
		4	back			8	back
C	x	9	right	D	x	13	right
		10	left			14	left
		11	top			15	top
		12	bottom			16	bottom
E	x	17	right	F	x	21	right
		18	left			22	left
		19	top			23	top
		20	bottom			24	bottom
G	y	25	front	H	y	29	front
		26	back			30	back
		27	top			31	top
		28	bottom			32	bottom
I	z	33	right	J	x	37	right
		34	left			38	left
		35	front	K	x	39	right
		36	back			40	left

Summing moments about points A and B and setting them to zero yields the following equations:

$$Fa = k_4(\varepsilon_1 - \varepsilon_2) \tag{4}$$

$$Fb = k_5(\varepsilon_3 - \varepsilon_4) + k_6(\varepsilon_5 - \varepsilon_6) + k_7(\varepsilon_5 + \varepsilon_6) \tag{5}$$

Note that moments are proportional to the difference between a strain gauge pair and forces to the sum of a strain gauge pair. The third term in equation 5 should be multiplied by the distance c, however because c remained constant for our test series, it can be incorporated into k_7.

3D Theory - The same theory can be applied to a three-dimensional applied force by reducing the problem to a series of two-dimensional problems. The number of strain gauges necessary to handle the three dimensional case increases significantly (see Figure 7). At each of positions A through I there are four gauges, while at both J and K there are 2 gauges. The gauge number, sensitive axis and mounted surface of each gauge is given in Table 1. Redundant strain gauges are needed at locations B, D, and F because gauge pairs are required in both sum and difference bridges, which cannot be accomplished with a single pair of gauges. Free-body diagrams from a number of different views are given in Appendix A. These free-body diagrams illustrate the force and moment balance necessary to relate strain to applied force.

From the derived equations in Appendix A, seven equations in thirteen variables (D_1 to D_{13}) can be written:

$$\begin{aligned} F_x &= k_3(D_3) + k_4(D_4) \\ F_y &= k_5(D_5) + k_6(D_6) \\ F_z &= k_1(D_1) + k_2(D_2) \\ F_{xz}a &= k_7(D_7) \\ F_{xz}b &= k_8(D_8) + k_9(D_9) \\ F_{yz}d &= k_{10}(D_{10}) + k_{11}(D_{11}) \\ F_{xy}f &= k_{12}(D_{12}) + k_{13}(D_{13}) \end{aligned} \tag{6}$$

where D_1 through D_{13} are combinations of strain sums and differences. Note that moments involving F_{xz} appear twice in the equation set. This redundant measure was used to increase confidence in the line-of-action estimate for the force in the sagittal plane (our analysis plane). The above system of equations can be expressed in matrix form as follows:

$$\mathbf{F} = \mathbf{DK} \tag{7}$$

where:

$$\mathbf{F} = \begin{bmatrix} F_x & F_y & F_z & F_{xz}a & F_{xz}b & F_{yz}d & F_{xy}f \end{bmatrix}$$

= force components and moments

$$\mathbf{D} = \begin{bmatrix} D_1 & D_2 & \cdots & D_{13} \end{bmatrix}$$

= combined strain gauge outputs

$$\mathbf{K} = \begin{bmatrix} 0 & 0 & k_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_2 & 0 & 0 & 0 & 0 \\ k_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_5 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & k_{11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_{12} \\ 0 & 0 & 0 & 0 & 0 & 0 & k_{13} \end{bmatrix}$$

= theoretical calibration matrix

Calibration and Error Estimates - The value of k_1 to k_{13} in the theoretical calibration matrix relate the combined strain gauge outputs to the internal axial forces and bending moments and could be determined analytically. However, misalignment of strain gauges and the head restraint structure itself will cause strain gauges to respond to forces perpendicular to their sensitive axes. This cross-talk results in non-zero elements throughout the $\mathbf{K}$ matrix which cannot be determined analytically. Therefore, an empirical calibration procedure was used to determine the elements of the calibration matrix $\mathbf{K}$:

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & \cdots & k_{1,7} \\ k_{2,1} & k_{2,2} & k_{2,3} & \cdots & k_{2,7} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{13,1} & k_{13,2} & k_{13,3} & \cdots & k_{13,7} \end{bmatrix} \tag{8}$$

A minimum of 13 sets of calibration data in which $\mathbf{F}$ and $\mathbf{D}$ are known are needed to determine $\mathbf{K}$. However for a robust estimate of $\mathbf{K}$, more than 13 data sets should be used and a least-squares solution for $\mathbf{K}$ determined. A least-squares estimate of $\mathbf{K}$ is calculated from the following equation:

$$\mathbf{K} = \left(\mathbf{d}^{\mathrm{T}}\mathbf{d}\right)^{-1}\mathbf{d}^{\mathrm{T}}\mathbf{f} \tag{9}$$

where:

$$\mathbf{d} = \begin{bmatrix} D_{1,1} & D_{2,1} & D_{3,1} & \cdots & D_{13,1} \\ D_{1,2} & D_{2,2} & D_{3,2} & \cdots & D_{13,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_{1,n} & D_{2,n} & D_{3,n} & \cdots & D_{13,n} \end{bmatrix}$$

= strain gauge data from n calibration runs

$$\mathbf{f} = \begin{bmatrix} F'_{x,1} & F'_{y,1} & F'_{z,1} & \cdots & \left(F_{xy}f\right)'_1 \\ F'_{x,2} & F'_{y,2} & F'_{z,2} & \cdots & \left(F_{xy}f\right)'_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F'_{x,n} & F'_{y,n} & F'_{z,n} & \cdots & \left(F_{xy}f\right)'_n \end{bmatrix}$$

= applied forces and moments from n calibration runs

To assess the error in the calculated forces and moments, a standard deviation (σ) for each force and moment component was calculated:

$$\sigma_{F_x} = \sqrt{\frac{\sum_{i=1}^{n}\left(F_{xi} - F'_{xi}\right)^2}{n-p}}$$

$$\sigma_{F_y} = \sqrt{\frac{\sum_{i=1}^{n}\left(F_{yi} - F'_{yi}\right)^2}{n-p}} \tag{10}$$

$$\vdots$$

$$\sigma_{F_{xy}f} = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(F_{xy}f\right)_i - \left(F_{xy}f\right)'_i\right]^2}{n-p}}$$

where:

F_{xi} = calculated force in x-direction for calibration i
F'_{xi} = actual force in x-direction for calibration i
n = number of points in the calibration data set
p = number of calculated parameters = 13

IMPLEMENTATION

STRUCTURAL MODIFICATIONS - To simplify the impact response of the head restraint, the mounting arrangement of the inboard post was modified to match the locking outboard post. This modification eliminated vertical motion of the inboard post during the impact and prevented the head restraint from asymmetrical distortion. The steel mounting brackets were replaced by steel and aluminum components to improve strain gauge sensitivity (see Figure 9). The position and orientation of the head restraint were not altered by these modifications. The steel mounting block was welded onto the seat back frame and the aluminum "dog-bones" were screwed into the new mounting blocks. A sheet-metal guard was attached to the front of the seat-back frame to protect the mounts and head restraint posts from occupant loading through the seat back foam.

STRAIN GAUGES - Strain gauges were applied to the head restraint tubes, "dog-bones" and seat back tube at the locations shown in Figures 7 and 9 and listed in Table 1. A total of 40 strain gauges were applied. The strain gauges were manufactured by JP Technologies. Pace 06-062AH-350 gauges were used on the steel seat back and head restraint tubes and Pace 13-062AH-350 gauges were used on the aluminum "dog-bones". Both types of gauges had a grid length of 1.6 mm, an overall length of less than 5 mm, and a nominal resistance of 350 ohms. Insulated, solid-copper, 34-gauge leads were soldered to the strain gauge terminals and run to terminal strips located within 3 cm. Shielded, 22-gauge, four-conductor, stranded wire was run from the terminal strips to a signal conditioning board mounted behind the seat. Two gauges were run in each four-conductor cable.

Bean-brand CA-200SL adhesive and catalyst were used to fix the strain gauges and terminal strips to prepared metal surfaces. This cyanoacrylate-based glue required neither baking nor extended cure. Gauges and lead wires were coated with Bean Gagekote #7 - a two-part epoxy material which protects the gauges and lead wires from mechanical loads and the effects of humidity. A strain-gauged "dog-bone" is shown in Figure 10.

BRIDGE CIRCUITS - Each gauge had a nominal resistance of 350 ohms, however after mounting its resistance was slightly altered by strains induced by adhesive cure or structure pre-load. This initial offset in each Wheatstone bridge was balanced using an external resistor network immediately prior to each human subject test or series of calibration tests. The output signals from the bridge circuits were amplified 82 times before entering the analog math circuits. The bridge balance and amplifying circuit is shown in Figure 11.

ANALOG MATH CIRCUITS - The signals from the forty strain gauges (ε_1 to ε_{40}) were reduced to thirteen output signals (D_1 to D_{13}) using board-based math. Low signal levels, coupled with the necessity to record the differences between some gauges and bridge outputs dictated that analog algebraic circuitry be used. Analog math improved the signal-to-noise ratio and minimized quantization errors in the digital data acquisition system. Operational amplifiers were configured for analog signal addition and subtraction as shown in Figure 12. These circuits were arranged on a single printed circuit board to produce the appropriate combinations of signals at the thirteen outputs (Figure 13).

CALIBRATION PROCEDURE - To solve for the calibration matrix $\mathbf{K}$, a series of known forces were applied at known locations in known directions and the combined strain gauge outputs measured. A dynamic load pulse similar to that expected during an actual head impact was chosen to generate the calibration data. The seat back was rigidly supported at mid-height to eliminate rearward deflection at the seat hinge for the calibration procedure.

To apply a load to the head restraint, three eyebolts were fixed to the rigid plastic core through the back of the head restraint. A cable/pulley arrangement with an in-line load cell (Artech 20210 S-type load cell) was used to apply a dynamic loading pulse to the back of the head restraint. The pulley was attached to a frame with multiple pulley attachment points. By combining 15 pulley positions and 3 eyebolt locations, 45 loading conditions could be applied to the seat back (Figure 14), of which 32 were used for each calibration run.

The direction of the applied force and point of application for each pulley/eyebolt combination was measured with a three-dimensional digitizing arm with an accuracy of ±0.76 mm. This resulted in calibration force direction estimates within ±0.2 degrees. The locations about which the moment equations were developed (shown in Appendix A) were also recorded. With this data, the magnitude of the force components (F_x, F_y, and F_z) and offsets (a, b, d, and f) for each of the 32 calibration loading conditions were calculated. Forces varied between 37.5 degrees either side of horizontal and 30 degrees above and 55 degrees below the horizontal rearward loading direction (Figure 14).

RESULTS

Eleven sets of calibration tests were performed over the eight week testing schedule. For each calibration test, the load cell output and thirteen combined strain gauge outputs were recorded. A sample set of time-varying signals is shown in the top row of graphs in Figure 15. In order to determine the calibration matrix $\mathbf{K}$, the $\mathbf{d}$ and $\mathbf{f}$ matrices were generated from the 32 dynamic data sets.

Because of the dynamic nature of the calibration load, each time-varying data set represented many redundant measures of the same loading condition. These redundant data allowed us to examine the effect of using a number of time slices per calibration data set. The improved fit between the applied force (thin line) and

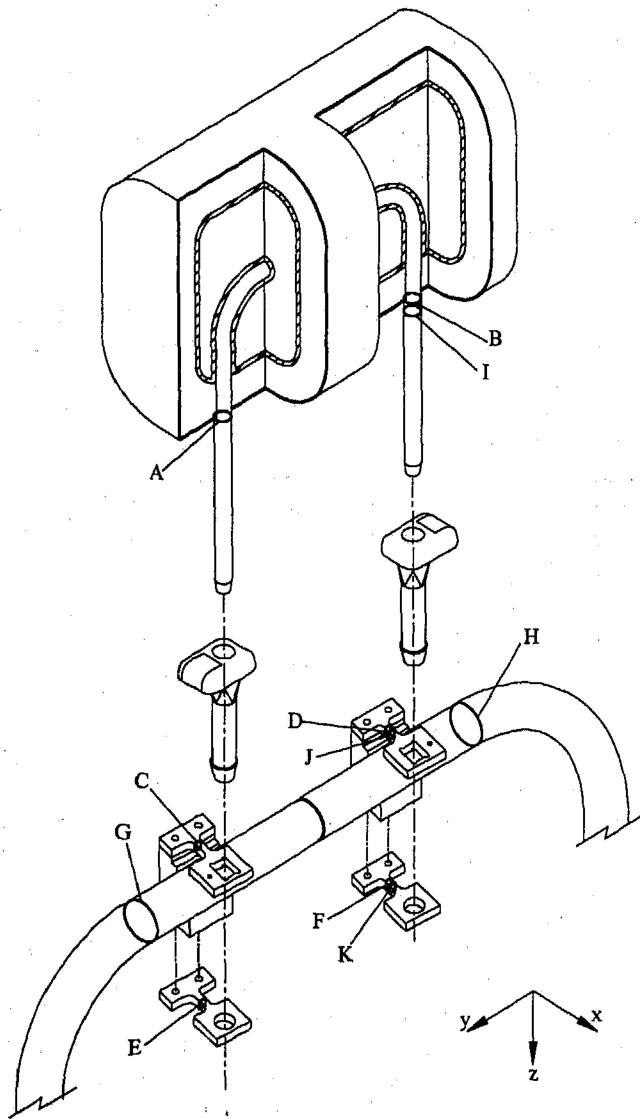
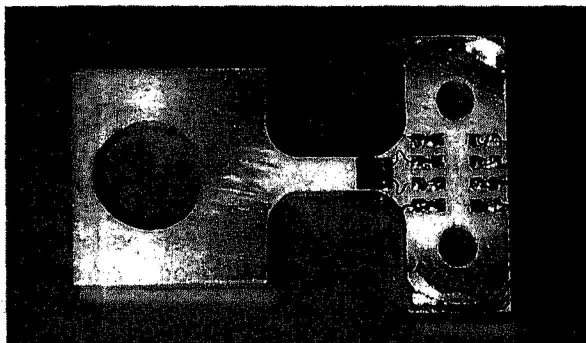


Figure 9. Modified head restraint mounts showing four strain gauged "dog-bones" and locking mechanism on both head restraint posts.


Figure 10. A strain-gauged aluminum "dog-bone"

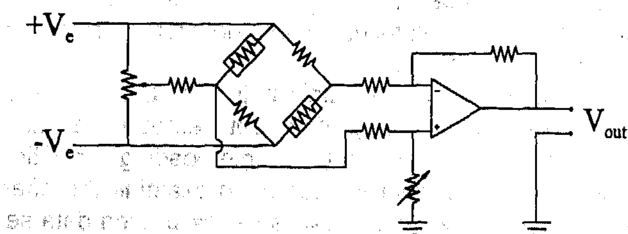
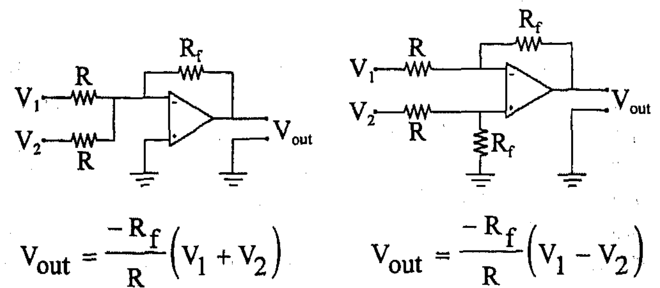


Figure 11. Wheatstone bridge with balance and amplifier circuits



$$V_{out} = \frac{-R_f}{R}\left(V_1 + V_2\right) \qquad V_{out} = \frac{-R_f}{R}\left(V_1 - V_2\right)$$

Figure 12. Addition and subtraction circuits

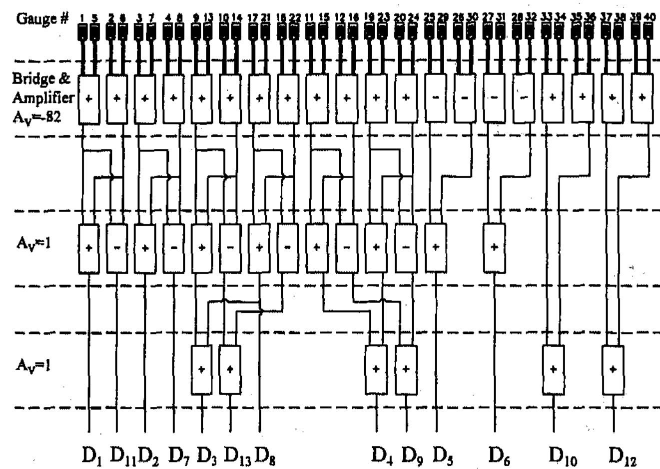


Figure 13. Bridge and math circuit board

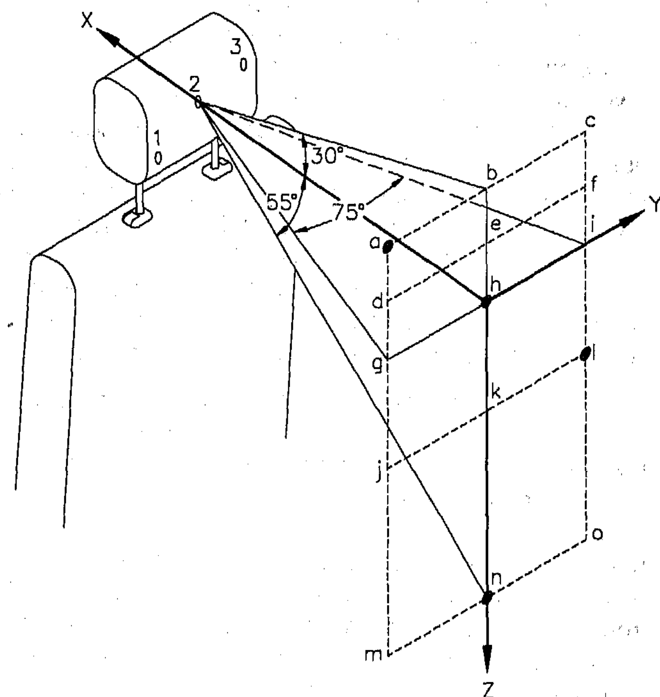


Figure 14. Calibration loading configuration showing the three eyebolt locations on the head restraint and the 15 pulley attachment locations on the frame located behind the seat. Calibrations loading conditions are described using the letter and number corresponding to the pulley and eyebolt used, i.e., h2 is horizontal rearward force. Circled pulley positions correspond to data presented in Figures 17 and 18.

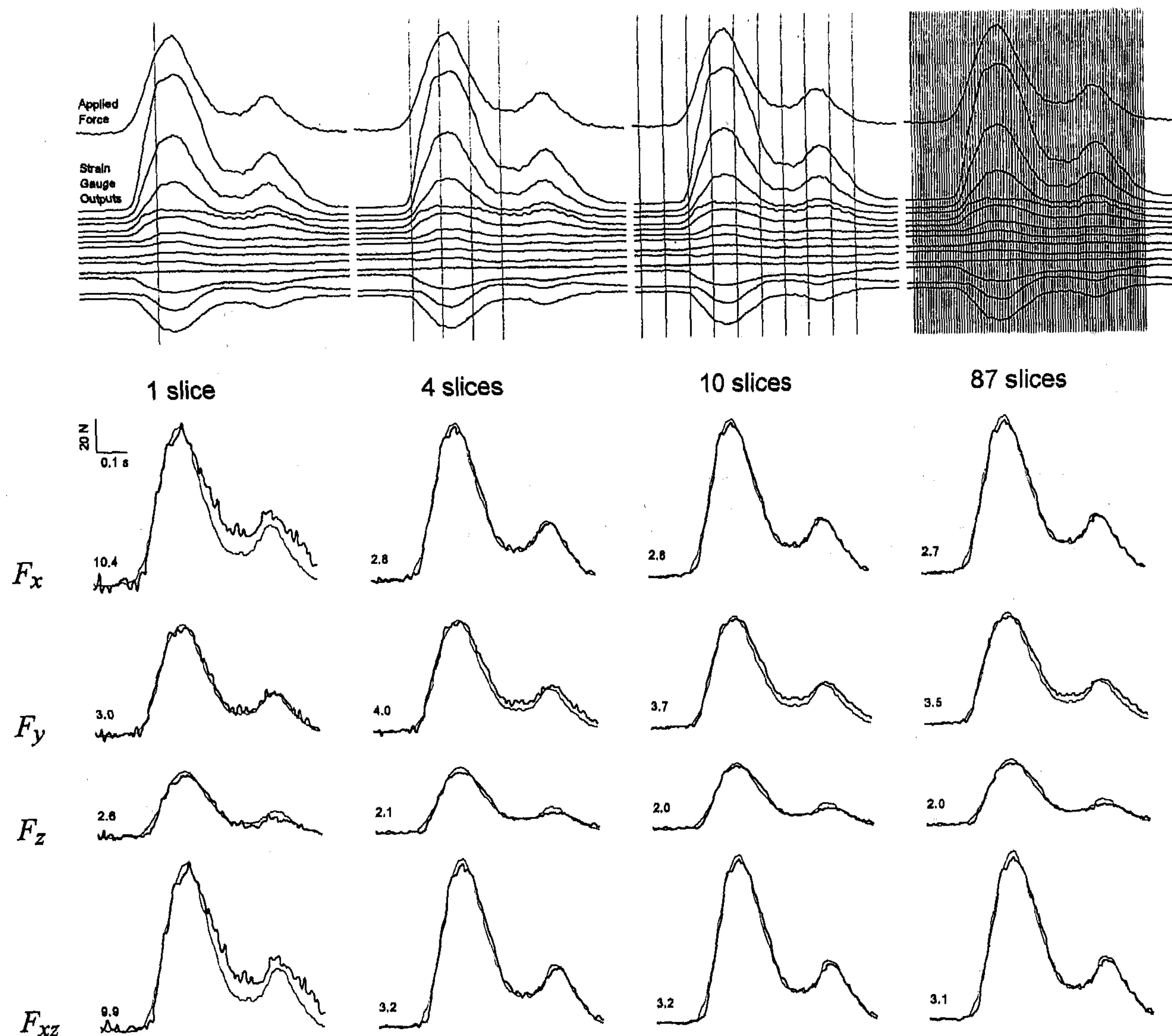


Figure 15. Calibration results using a different numbers of observations (slices) from each time-varying load applied during a single calibration set of 32 applied loads. The load shown is from pulley/eyebolt location I2 during calibration set 11. The top row depicts the same data set with 1, 4, 10, and 87 slices used for the calibration. Results for each force component F_x, F_y, F_z and the combined force F_{xz} in the sagittal plane are shown in the four lower rows for the four different number of observations used. In the four lower graphs, the actual applied load (thin line) and the calculated load (thick line) are shown. Standard deviations (shown at the left of each graph) between the signals indicate diminishing improvement with increased number of slices.

Table 2. Standard deviation of the calculated force components F_x, F_y and F_z and angle θ_y of the combined force F_{xz} in the sagittal plane for each calibration data set.

Calibration	F_x (N)	F_y (N)	F_z (N)	θ_y (deg)
1	1.9	1.6	1.6	3.1
2	2.8	2.8	2.0	2.3
3	2.3	2.0	1.9	2.9
4	2.4	1.9	1.8	3.0
5	2.0	1.9	1.7	2.9
6	2.5	2.0	1.7	2.7
7	3.1	2.2	2.1	3.1
8	2.6	2.1	1.8	3.0
9	2.9	2.0	1.9	3.2
10	3.0	2.1	1.7	3.2
11	3.0	2.2	2.0	3.3
All	2.6	2.1	1.8	3.0

Table 3. Standard deviation of the calculated line-of-action offsets *a*, *b*, *d* and *f* from the points about which the moments were calculated in Appendix A.

Calibration	*a* (mm)	*b* (mm)	*d* (mm)	*f* (mm)
1	3.5	4.3	23.9	16.1
2	4.1	5.0	22.1	14.5
3	3.9	4.8	25.5	15.3
4	4.0	5.0	24.7	15.9
5	4.4	5.4	23.7	12.9
6	3.8	4.8	25.6	15.0
7	4.2	5.0	26.4	15.9
8	4.7	5.6	27.0	14.5
9	5.2	6.5	25.1	16.3
10	5.1	6.3	26.0	17.0
11	5.1	6.0	27.4	16.2
All	4.4	5.3	25.1	15.3

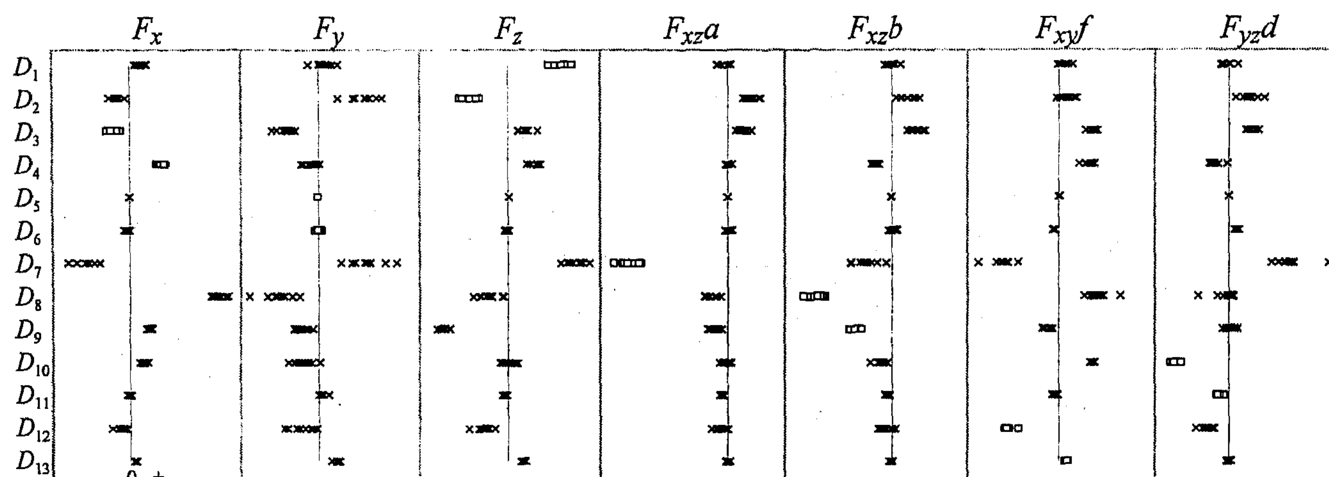


Figure 16. Comparison of calibration matrices from data sets 3 through 11. Each horizontal grouping of points represents the variation in a common element across all 9 matrices. The elements are arranged identically to the K-matrix in equation 7 to allow direct comparison between actual and theoretical quantities, i.e., the top left group corresponds to $k_{1,1}$ and the bottom right group corresponds to $k_{13,7}$ from equation 8. Elements represented by a square symbol correspond to elements k_1 through k_{13} in equation 7.

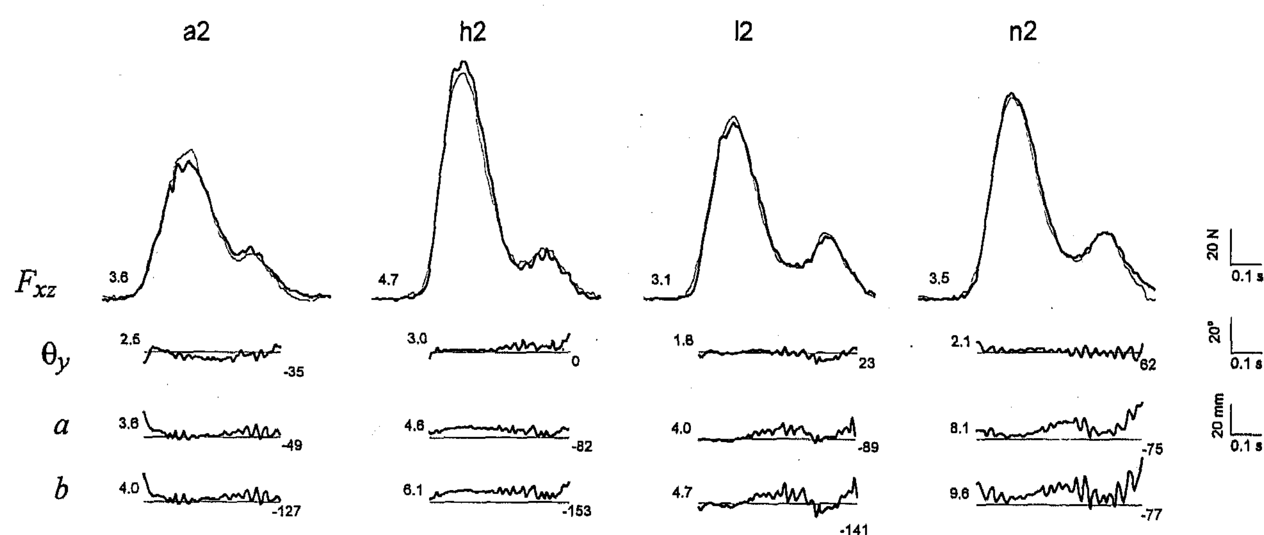


Figure 17. Comparison between actual (thin line) and calculated (thick line) parameters F_{xz}, θ_y, a and b for $\mathbf{K}_{11}$ and calibration data set 11. From left to right, the data represents pulley/eyebolt combinations a2, h2, l2 and n2 in Figure 14. The graphs for θ_y, a and b are truncated because the denominator used in their calculation tends to zero when no load is applied and therefore becomes unstable. Data is shown for loads greater than 10 N. Standard deviations between the two signals are shown to the left of each graph. The value of the actual parameter (thin line) for θ_y, a and b is shown to the right of each graph.

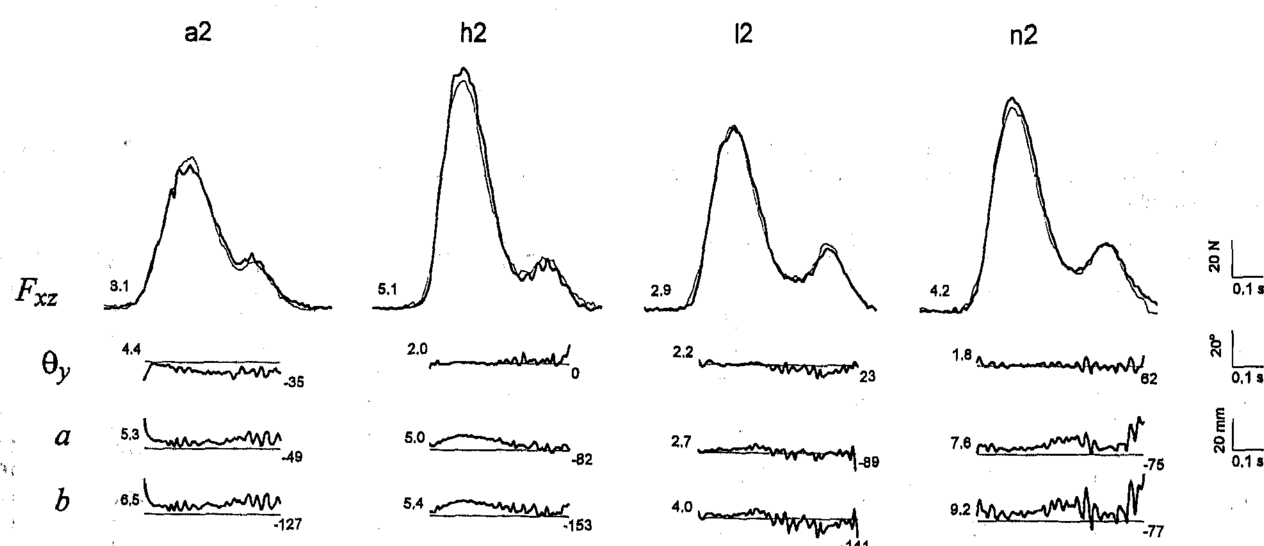


Figure 18. Same as Figure 17, except calculated values obtained by applying $\mathbf{K}_3$ to calibration data set 11.

calculated force (thick line) by increasing the number of time slices is shown in the bottom four rows of Figure 15. The standard deviations in Figure 15 show that increasing slices improves the fit between actual and calculated forces, but that this effect diminishes above 10 slices.

ERROR ESTIMATES - The standard deviations of the calculated force components were determined with Equation 10 and are summarized in Table 2. The standard deviations in the line-of-action offsets are shown in Table 3. Standard deviations of about ±2 N were achieved for the force components and ±3 degrees for the force direction in the xz-plane. The standard deviations for the line-of-action offsets were about ±5 mm in the xz-plane (moments about the y-axis), however they were three to five times larger for moments about the x- and z-axes. These larger deviations may be a result of a calibration procedure which focused on sagittal plane loading rather than a uniform distribution of loading in 3-space.

Calibration matrices were calculated for each of the 11 calibration runs using 87 slices. Nine of the 11 matrices are compared in Figure 16 (the first two calibration matrices can not be compared to the final nine due to a change in amplification gain made to one of the strain gauge output channels). Figure 16 indicates variation between the calibration matrices. To examine the effect of this variation, two calibration matrices ($\mathbf{K}_3$ and $\mathbf{K}_{11}$) were applied to the data set for calibration run 11 and differences in the ultimate calculated parameters for our sagittal plane analysis (F_{xz}, θ_y, a, and b) were examined (see Figures 17 and 18). Differences between $\mathbf{K}_3$ and $\mathbf{K}_{11}$ were representative of variations between the nine matrices and were used for this comparison because they were chronologically the furthest apart.

Figure 17 represents the best possible result because the calibration matrix was calculated from the same data to which it is being applied. Figure 18 shows what $\mathbf{K}_3$ predicts for the same data. A comparison of the standard deviations between the applied and calculated quantities indicates repeatable results despite differences between calibration matrices.

SAMPLE DATA - The head restraint data acquired during a rear-end automobile collision was processed with the calibration matrix from the closest chronological calibration data set. For this test, a female subject (162 cm, 65 kg) was exposed to an 8 km/h speed change. The resulting head restraint forces are shown in Figure 19. The rise in head restraint force in advance of head contact is caused by inertial loading and must be subtracted from the calculated force before it is mapped to the subject's head reference frame. The net force in the xz-plane is also shown, along with the direction θ_y of the force and the line-of-action offsets a and b. Note that the loading direction remains within the calibration envelope shown in Figure 14.

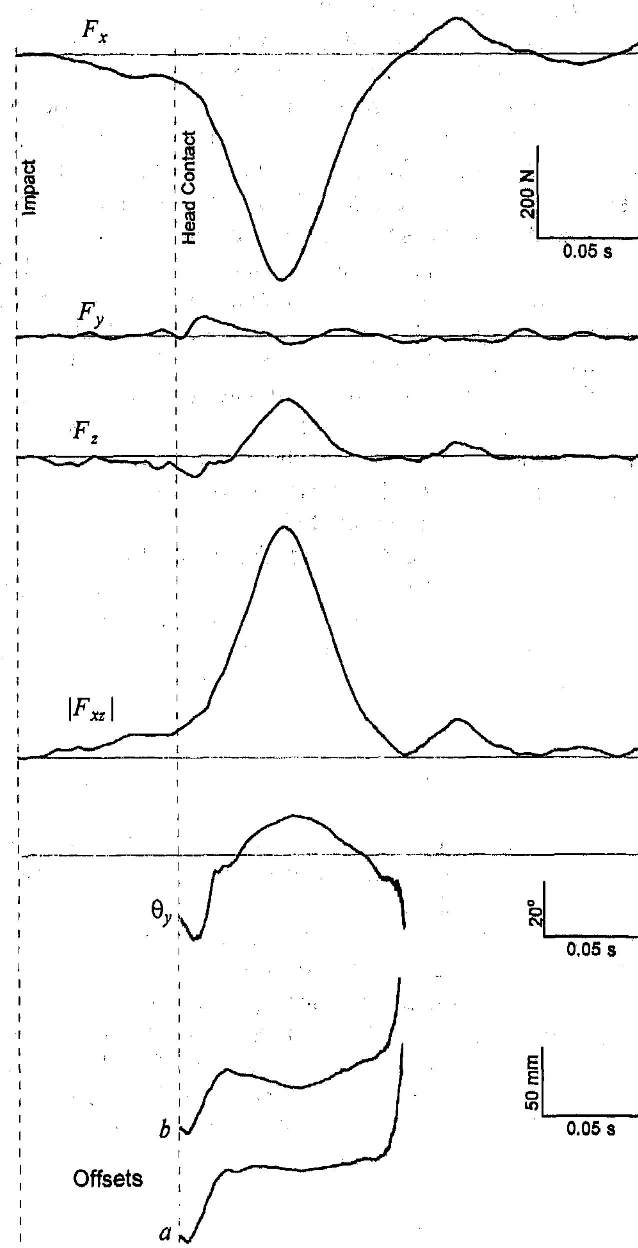


Figure 19. Sample data from rear-end collision with speed change of 8 km/h. Rise in F_x between impact and head restraint contact is due to inertial loading. Data for θ_y, a, and b not shown for forces less than 10 N.

DISCUSSION

The only problem with the strain gauges during the testing schedule involved the lead wires. The Gagekote material cracked at a point of flexure and caused an imbedded lead wire to break. All of the lead wires on the head restraint gauges were replaced and were not re-treated with Gagekote.

The strain gauges became noticeably warm during operation. To minimize the transient effects of heating and cooling during the test, the gauges were allowed to reach a constant temperature before each test.

Two of the combined strain gauge outputs (D_6 and D_{11}) occasionally responded non-linearly to the applied force. This is believed to be due to slight shifting of the head restraint posts within the plastic sleeves during loading. This shifting is a result of normal play in the mounting system. The plastic sleeves were left in the system for two reasons: they contain the height adjustment mechanism, and our objective was to keep as many of the original head restraint components in the system as possible. As a result of the non-linear behavior in signals D_6 and D_{11}, near-zero coefficients in the calibration matrix were generated (see Figure 16).

Some of the strain gauge outputs remained very low throughout the calibration force applications. A least squares solution for such a channel produced a near-zero coefficient since there was little information about the applied force in the output signal.

Future work could include eliminating the play due to the plastic sleeves and making use of the data currently ignored because of non-linearities. Alternatively, these and other low-output channels could be removed from the calibration data and smaller $\mathbf{K}$-matrices calculated.

CONCLUSIONS

The use of strain gauges to measure the magnitude, direction and line of action of a force applied to a head restraint has been demonstrated. A dynamic calibration procedure was developed and least-squares calibration matrices were calculated.

Standard deviations for predicting the force components F_x, F_y and F_z and line-of-action offsets a, b, d and f are estimated to be ±2.6 N, ±2.1 N, ±1.8 N, ±4.4 mm, ±5.3 mm, ±25.1 mm and ±15.3 mm respectively. The standard deviation for predicting θ_y is estimated to be ±3.0 degrees.

ACKNOWLEDGMENT

Funding for this project was provided by MacInnis Engineering Associates, and Biomechanics Research & Consulting of El Segundo, CA. Additional funding was provided by a Technology BC "Investments in Research and Development" grant (#132) administered by the Science Council of British Columbia.

REFERENCES

1. Mertz, Harold F. J., The kinematics and kinetics of whiplash, Ph.D. Thesis, Wayne State University, Detroit, Michigan, 1967.

APPENDIX A 3-D FREE-BODY DIAGRAMS

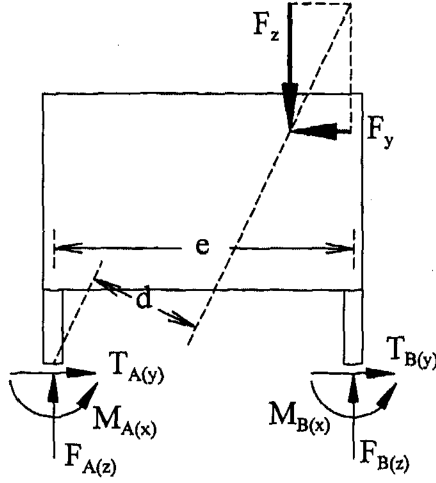


Figure A1. yz-plane, sectioned at A and B (front view)

$$F_z = F_{A(z)} + F_{B(z)}$$
$$F_{yz}d = F_{B(z)}e + M_{A(x)} + M_{B(x)}$$
$$F_z = k_1(\varepsilon_1 + \varepsilon_2 + \varepsilon_5 + \varepsilon_6)$$
$$F_z = k_2(\varepsilon_3 + \varepsilon_4 + \varepsilon_7 + \varepsilon_8)$$
$$F_{yz}d = k_{10}(\varepsilon_{33} + \varepsilon_{34} + \varepsilon_{35} + \varepsilon_{36}) + k_{11}[(\varepsilon_1 - \varepsilon_2) + (\varepsilon_5 - \varepsilon_6)]$$
$$D_1 = \varepsilon_1 + \varepsilon_2 + \varepsilon_5 + \varepsilon_6$$
$$D_2 = \varepsilon_3 + \varepsilon_4 + \varepsilon_7 + \varepsilon_8$$
$$D_{10} = \varepsilon_{33} + \varepsilon_{34} + \varepsilon_{35} + \varepsilon_{36}$$
$$D_{11} = (\varepsilon_1 - \varepsilon_2) + (\varepsilon_5 - \varepsilon_6)$$

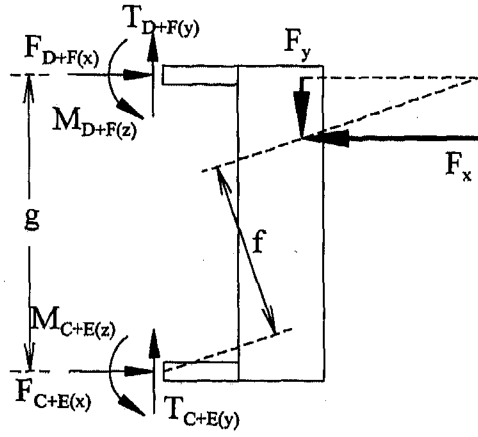


Figure A2. xy-plane, sectioned at C, D, E and F (top view)

$$F_x = F_{D(x)} + F_{F(x)} + F_{C(x)} + F_{E(x)}$$
$$F_{xy}f = (F_{D(x)} + F_{F(x)})g - M_{D(z)} - M_{F(z)} - M_{C(z)} - M_{E(z)}$$
$$F_x = k_3(\varepsilon_9 + \varepsilon_{10} + \varepsilon_{13} + \varepsilon_{14} + \varepsilon_{17} + \varepsilon_{18} + \varepsilon_{21} + \varepsilon_{22})$$
$$F_x = k_4(\varepsilon_{11} + \varepsilon_{12} + \varepsilon_{15} + \varepsilon_{16} + \varepsilon_{19} + \varepsilon_{20} + \varepsilon_{23} + \varepsilon_{24})$$
$$F_{xy}f = k_{12}(\varepsilon_{37} + \varepsilon_{38} + \varepsilon_{39} + \varepsilon_{40})$$
$$+ k_{13}[(\varepsilon_{13} - \varepsilon_{14}) + (\varepsilon_9 - \varepsilon_{10}) + (\varepsilon_{17} - \varepsilon_{18}) + (\varepsilon_{21} - \varepsilon_{22})]$$
$$D_3 = \varepsilon_9 + \varepsilon_{10} + \varepsilon_{13} + \varepsilon_{14} + \varepsilon_{17} + \varepsilon_{18} + \varepsilon_{21} + \varepsilon_{22}$$
$$D_4 = \varepsilon_{11} + \varepsilon_{12} + \varepsilon_{15} + \varepsilon_{16} + \varepsilon_{19} + \varepsilon_{20} + \varepsilon_{23} + \varepsilon_{24}$$
$$D_{12} = \varepsilon_{37} + \varepsilon_{38} + \varepsilon_{39} + \varepsilon_{40}$$
$$D_{13} = (\varepsilon_{13} - \varepsilon_{14}) + (\varepsilon_9 - \varepsilon_{10}) + (\varepsilon_{17} - \varepsilon_{18}) + (\varepsilon_{21} - \varepsilon_{22})$$

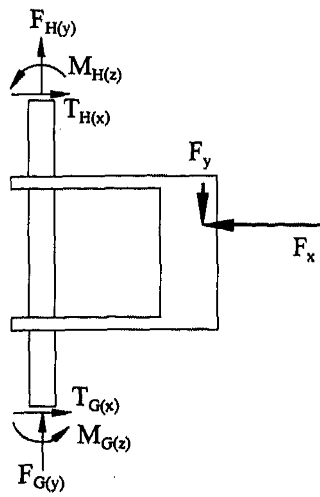


Figure A3. xy-plane, sectioned at G and H (top view)

$$F_y = F_{H(y)} + F_{G(y)}$$
$$F_y = k_5(\varepsilon_{25} + \varepsilon_{26} - (\varepsilon_{29} + \varepsilon_{30}))$$
$$F_y = k_6(\varepsilon_{27} + \varepsilon_{28} - (\varepsilon_{31} + \varepsilon_{32}))$$
$$D_5 = \varepsilon_{25} + \varepsilon_{26} - (\varepsilon_{29} + \varepsilon_{30})$$
$$D_6 = \varepsilon_{27} + \varepsilon_{28} - (\varepsilon_{31} + \varepsilon_{32})$$

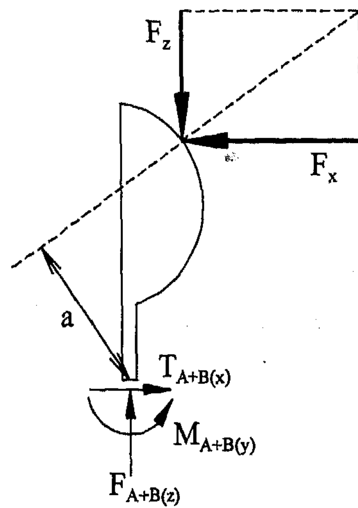


Figure A4. xz-plane, sectioned at A and B (side view)

$$F_{xz}a = -M_{A(y)} - M_{B(y)}$$

$$F_{xz}a = k_7\left[\left(\varepsilon_3 - \varepsilon_4\right) + \left(\varepsilon_7 - \varepsilon_8\right)\right]$$

$$D_7 = \left(\varepsilon_3 - \varepsilon_4\right) + \left(\varepsilon_7 - \varepsilon_8\right)$$

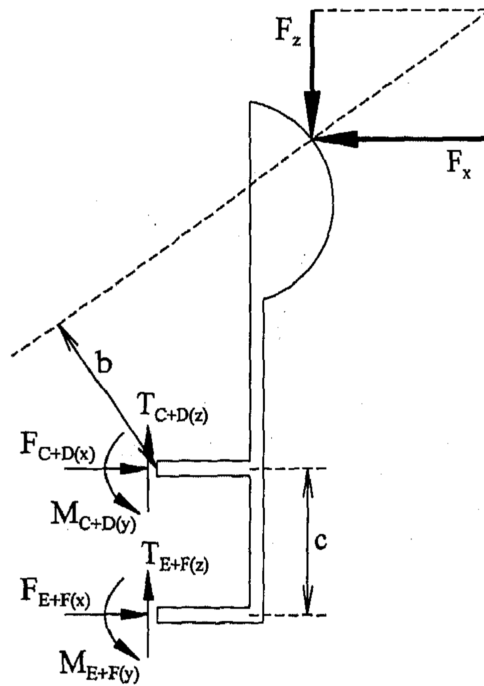


Figure A5. xz-plane, sectioned at C, D, E and F (side view)

$$F_{xz}b = -c\left(F_{E(x)} + F_{F(x)}\right) - M_{C(y)} - M_{D(y)} - M_{E(y)} - M_{F(y)}$$

$$F_{xz}b = k_8\left(\varepsilon_{17} + \varepsilon_{18} + \varepsilon_{21} + \varepsilon_{22}\right) + k_9\left[\left(\varepsilon_{11} - \varepsilon_{12}\right) + \left(\varepsilon_{15} - \varepsilon_{16}\right) + \left(\varepsilon_{19} - \varepsilon_{20}\right) + \left(\varepsilon_{23} - \varepsilon_{24}\right)\right]$$

$$D_8 = \varepsilon_{17} + \varepsilon_{18} + \varepsilon_{21} + \varepsilon_{22}$$

$$D_9 = \left(\varepsilon_{11} - \varepsilon_{12}\right) + \left(\varepsilon_{15} - \varepsilon_{16}\right) + \left(\varepsilon_{19} - \varepsilon_{20}\right) + \left(\varepsilon_{23} - \varepsilon_{24}\right)$$

970398

Multi-Body Model of Upper Extremity Interaction with Deploying Airbag

E. M. Sieveka, S. M. Duma, J. Pellettiere, J. R. Crandall, C. R. Bass, and W. D. Pilkey
University of Virginia Automobile Safety Lab.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Three-dimensional simulation models of a driver's right upper extremity interacting with a deploying airbag have been set up and run with the Articulated Total Body program. The goal of this study is to examine the significance of various occupant and airbag parameters during deployment, such as grip strength, upper extremity position, shoulder compliance, flap position, flap aggressivity, and deployment speed. Given a range of 250 N to 650 N, the grip strength did not affect the resultant loads. Also, the contact force and torque at the c.g. of the forearm are not sensitive to shoulder joint compliance. The flap aggressivity and the position of the airbag module relative to the upper extremity are most important in affecting the interaction. This study is used to justify cadaveric experiments involving disarticulated upper extremities.

INTRODUCTION

Motivation for this research stems from the high incidence of upper extremity injuries from airbags as seen in the literature [1-6]. The bones of importance in the upper extremity are the humerus of the upper arm and the radius and ulna of the forearm [7]. In the literature cases most fractures occur at the distal third in both the humerus and radius. Kuppa et al analyzed several accident data bases including NASS, NHTSA Special Investigations, and NHTSA sponsored studies and Jackson Memorial, University of Maryland and the University of New Jersey [8]. Kuppa found that 1.1% of drivers who were restrained only by a seatbelt experienced an upper extremity injury; however, in the presence of a deploying airbag, 4.4% of drivers showed upper extremity injuries. Two modes have been suggested for these injuries. The first is direct contact with the airbag, the type examined in this paper, and may occur while executing a left turn with the right hand, thus placing the forearm directly over the module. The second, indirect contact injuries, involves upper extremity contact with another object in the automobile subsequent to interaction with the airbag.

Saul et al. used an instrumented Hybrid III upper extremity to examine this interaction [9]. Using strain gauges and accelerometers, they found that bending moments and accelerations could be accurately recorded. Moreover, a correlation was found between these values and the airbag's inflator properties, flap, and steering wheel orientation.

Several studies examine the joint kinematics within the upper extremity and provide useful modeling data [10-16]. Additionally, the Research Arm Injury Device (RAID) was developed by Conrad Technologies Inc. and NHTSA to examine the upper extremity-airbag interaction [8]. This device determined that the two most critical factors were the orientation of the arm with respect to the airbag module and the separation distance between the two.

MODEL DESCRIPTION

Using the Articulated Total Body (ATB) multi-body dynamics program, a three-dimensional model was created to study the interaction between a driver's upper extremity and a deploying airbag. The model's initial conditions are shown in Figures 1 and 2.

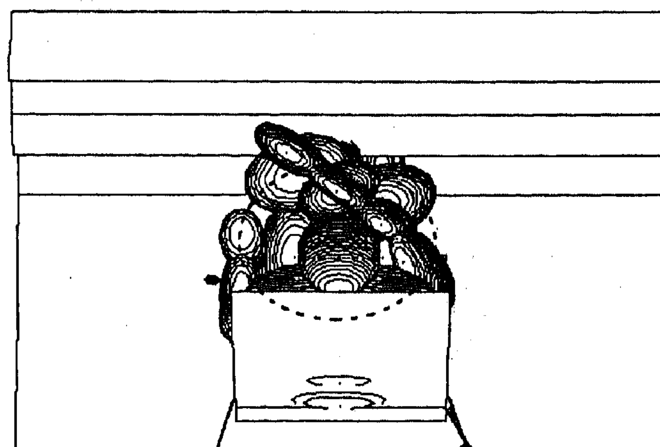


Figure 1. Rear View of Model Showing Upper Extremity Across Airbag Flap

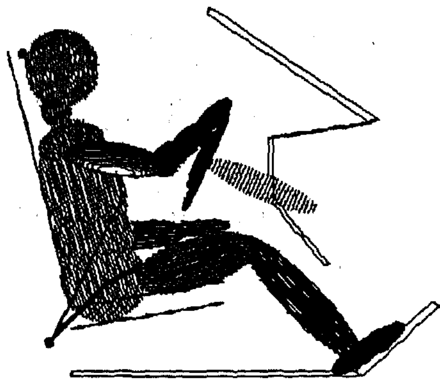


Figure 2. Side View of Model with Initial Upper Extremity Orientation

OCCUPANT JOINTS - The full occupant body has been retained for simulations involving a vehicle deceleration pulse, but the model can be restrained to allow movement of the right upper extremity only. The shoulder is modeled as a ball joint with some front-to-back (X-direction) compliance. This additional degree-of-freedom has been provided to simulate the flexibility associated with the clavicle, but it can also be locked to simulate a test with the shoulder fixed in space.

Friction in the joints is currently set to a 1 g level, but it can be increased to simulate muscle tension. The elbow is treated as a pin joint and the wrist as a ball joint. The right forearm has been divided into two segments connected by a locked joint at the forearm c.g. location. This permits force and torques experienced at the c.g. to be "measured" since ATB outputs the forces and torques that it applies to the associated segments in order to maintain the joint's locked condition. In principle, this joint can be made to unlock, or "break", when the applied torque exceeds a pre defined level; however, this would confound the prediction of potential peak loads for the various initial conditions.

GRIP STRENGTH - Gripping of the wheel rim by the driver's hand is simulated by a spring between the hand segment and the upper wheel segment. The spring force is defined by a tabulated function which reaches a maximum value that is associated with the grip strength. If the spring is stretched further, the force drops to zero to simulate the hand releasing from the rim. At present, the steering wheel is being treated as rigid.

AIRBAG MODEL - The airbag model is comprised of seven segments. Five of these segments, a center segment and four side segments, are soft, with contact properties approximating the fully deployed bag. Another segment represents the steering wheel "flap" that covers the bag module. Finally, the last segment is a special "punch" that is used to open the flap. See Figure 3 for a sketch of the model. Bag deployment is simulated by applying time-dependent external force functions to the five, soft bag segments. The center segment is driven backward toward the driver while the side segments, which are initially inside the center segment, are deployed sideways. Punch-through begins around 20 ms after simulated impact so that times can be related to times in a full sled or vehicle test. The functions are presently chosen so that the bag is almost completely deployed in 45 ms. The motion of each bag segment is finally arrested when a spring-damper, acting as a tether reaches equilibrium with the forcing function.

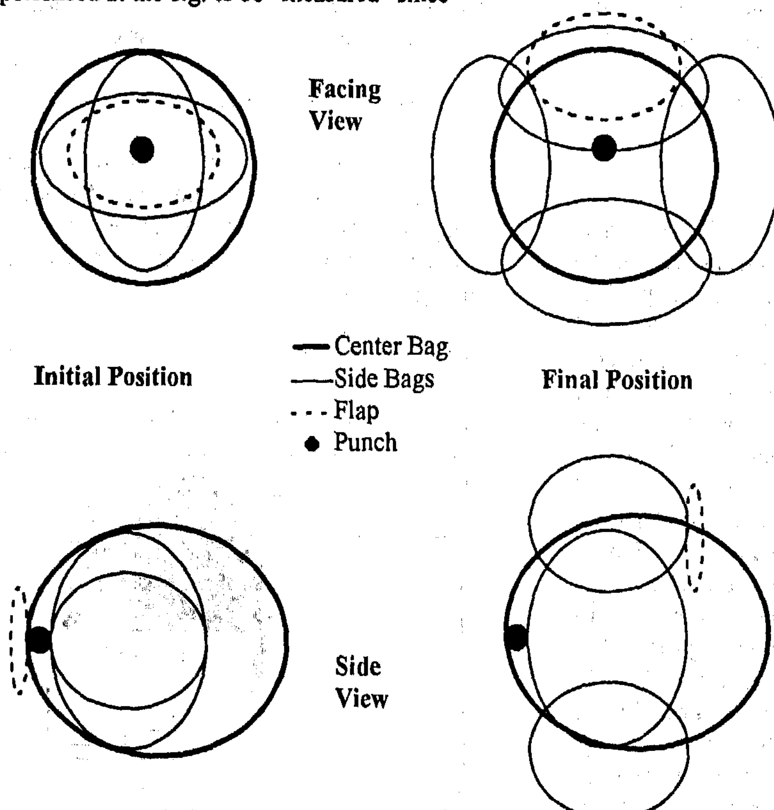


Figure 3. Airbag Model Consisting of Seven Segments: One Center Bag, Four Side Bags, One Flap, and One Punch

The punch segment offers considerable flexibility when studying the opening of the flap. It is connected to the center segment by a sliding, or "slip", joint and contacts the forward side of the flap as the center segment moves rearward. Relative motion between the center and punch segments is prevented for punch-to-flap force levels below the slip joint's locking force; after which, restraint is governed by a spring-damper. This design provides the capability to model a release, or plateau, of the bag-to-flap pressure as the flap opens. Initially the bag and flap move as one, but once a sufficient gap has opened the bag can begin billowing out sideways even though the flap may be restrained by the driver's upper extremity. Tests are needed, but it is reasonable to assume that the bag force against the flap does not increase significantly after punchout. The flap in the present model is fairly stiff, but the contact force between the flap and the upper extremity can be modified to simulate flap compliance.

EXAMPLE - Figure 4 shows the baseline model in action at the beginning, middle, and end of airbag deployment. The bag module is centered relative to the steering wheel and the grip strength of 650 N represents a relatively strong male driver [17].

The pictures in Figure 4 demonstrate that the model behavior is qualitatively realistic as compared to laboratory tests. Two results that are readily obvious are the lack of shoulder movement during the flap contact phase and the fact that even a "strong" grip is quickly released. With the basic functionality of the model established, it can be exercised to study the sensitivity of upper extremity loading to various parameters which are Table 1.

The parameters are listed in order of estimated importance. Upper extremity position is placed first since this study is only concerned with upper extremities in front of a deploying bag. However, it is not a parameter to be subjected to much variation since the upper extremity must be assumed to be in a high risk, worst case position that places it across the center of the wheel, in front of the deploying bag.

PARAMETER STUDY RESULTS

Simulations have been run with variations made to parameters 3, 4, 6, 7, 8 and 9. One of the initial goals of the simulation project was to determine if upper extremity-only component testing, with a shoulder joint fixed in space, could adequately replicate the upper extremity-to-airbag interaction. Figure 5 clearly demonstrates that, since significant upper extremity-to-airbag contact happens early in the crash event, the results obtained from the component test and the full-body, full-motion test are almost identical. Furthermore, Figures 6 and 7 show that there is also very little sensitivity of the upper extremity-to-flap contact force, or the C.G. torque, to a fixed or translating shoulder joint. In animations made from the simulations, motion of the translating shoulder joint is clearly visible, but the peak loading between the upper extremity and the flap has already occurred. These results strongly support the validity of upper extremity-only component testing, with non-translating shoulder joints.

Table 1. Summary of Available Model Parameters

1. Upper extremity position
2. Airbag deployment speed
3. Flap aggressivity (upper extremity contact stiffness and "punch through" behavior)
4. Bag module position (centered, high, low, etc.)
5. Flap size
6. Grip strength
7. Wrist and elbow stiffness (via friction variable)
8. Shoulder motion (locked or compliant in X-direction)
9. Torso motion relative to vehicle
10. Bag stiffness (for upper extremity contact)

Figures 6 and 7 also verify the relatively low importance assigned to grip strength and wrist torque. Of some interest for experimental testing, are the sharp peaks that appear around 37 ms in the C.G. torque plot. These peaks, which exceed the magnitude of the initial flap-contact peaks, are an artifact of having a hard-stop model for the wrist joint in order to represent the metal-to-metal contact of the actual Hybrid-III wrist stop. The human wrist has a distributed mass over several joints and would not show these spikes. In Figure 8, it is demonstrated that a much softer joint stop, as expected in a human wrist, eliminates this peak.

As expected, upper extremity loads show a much higher sensitivity to flap aggressivity (Figure 9), which is represented here by the flap-to-bag force plateau. Other factors which directly affect this occupant parameter are bag deployment speed and flap stiffness. While upper extremity-to-airbag loads can also be several thousand Newtons, it is the localized nature of flap contact which is significant and which increases the probability of breakage.

The results do show some sensitivity to bag module position (Figure 10), but further study is needed in order to make definite conclusions. A higher module position, in which the airbag seam is moved distally towards the wrist, reduces the contact force, but laboratory tests have demonstrated that bone fracture still occurs in this case. This may be due to narrowing of the upper extremity bones near the wrist. A low module position, on the other hand, produces little change in contact force, but the probability of injury may decrease due to increased bone cross-sectional area near the elbow, as well as additional support from elbow joint.

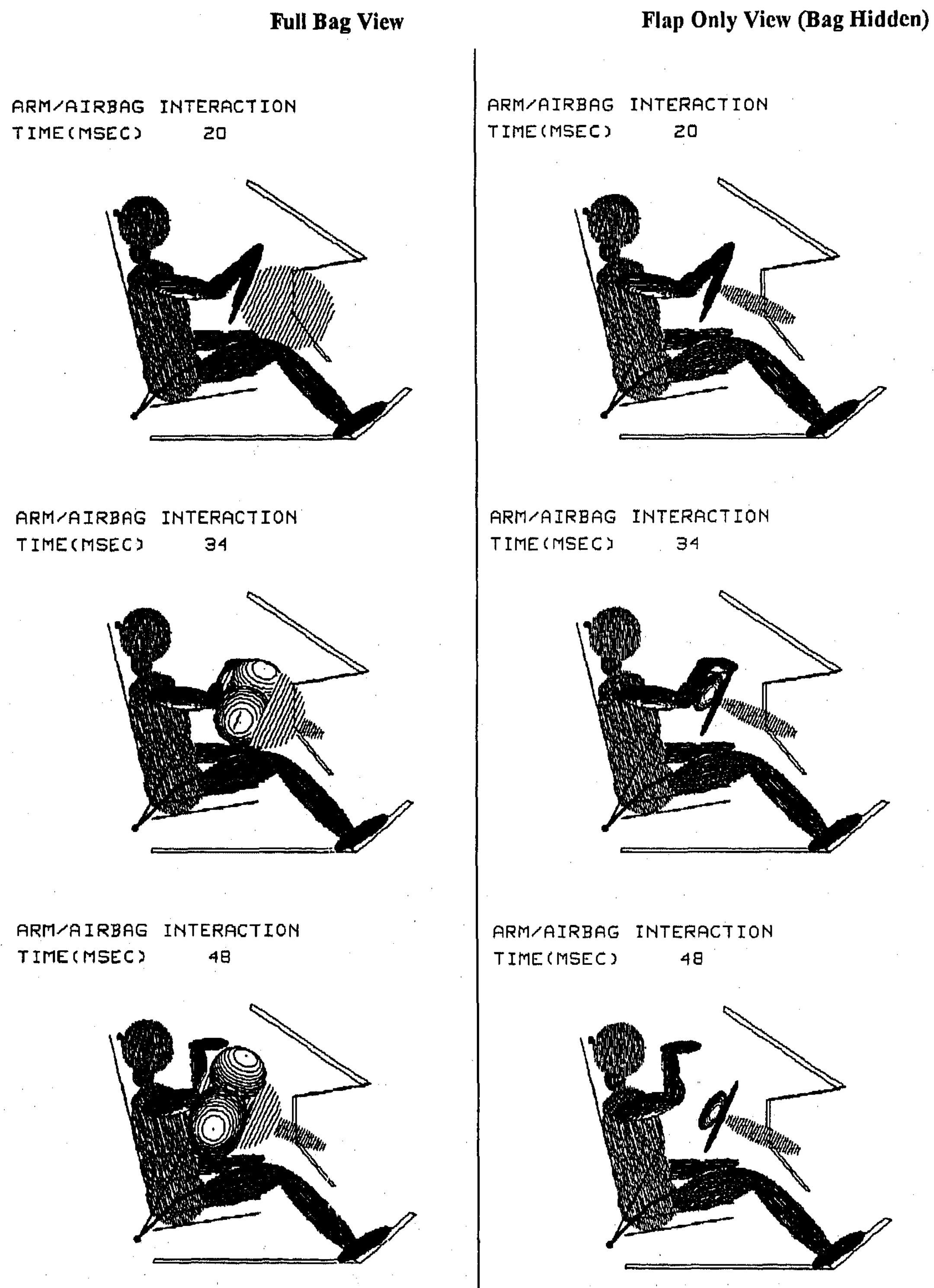


Figure 4. Upper Extremity/Airbag Interaction Sequences
Given the Full Bag View and Flap Only View

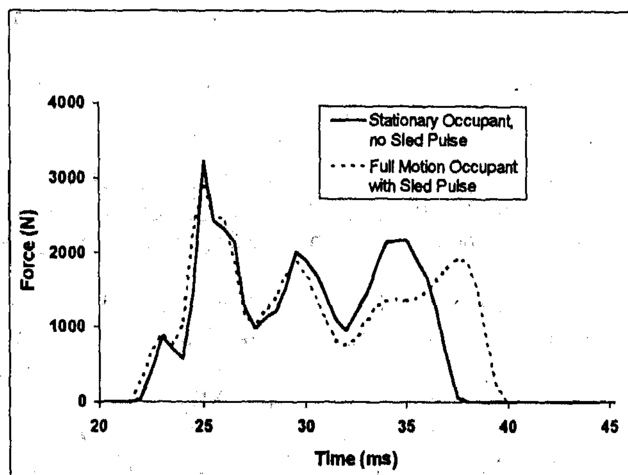


Figure 5. Upper Extremity-to-Flap Contact Force for Stationary and Moving Occupant

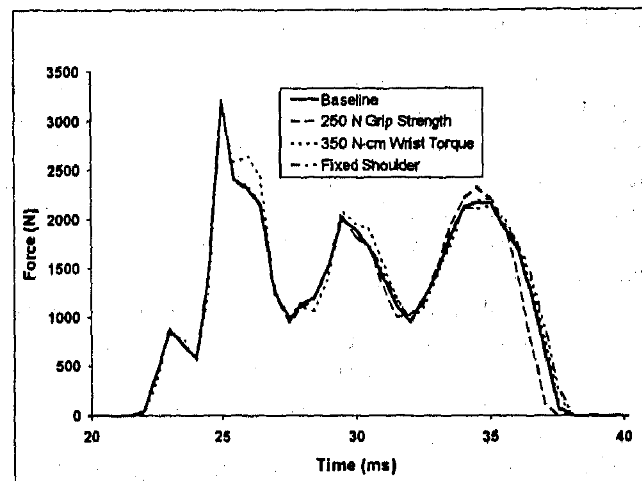


Figure 6. Upper Extremity-to-Flap Contact Force Sensitivity (200 N Flap-to-Bag Plateau) (Baseline Parameters: 650N Grip strength, 35N-cm (1G) Wrist Torque, Front-to-Back Shoulder Clavicle Compliance)

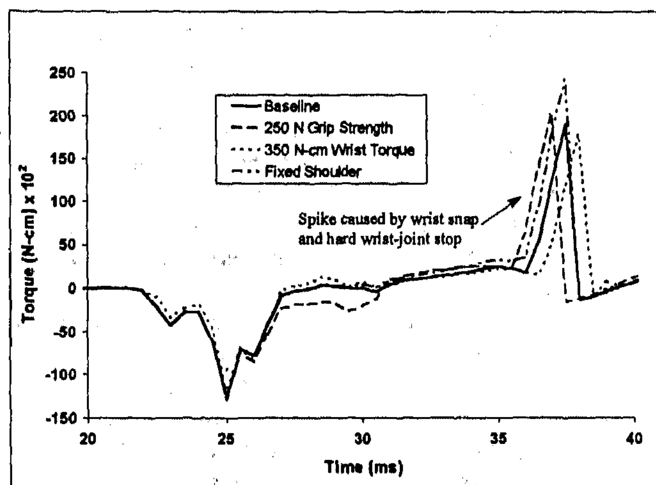


Figure 7. Forearm C.G. Torque Sensitivity (200 N Flap-to-Bag Plateau) (Baseline Parameters: 650N Grip strength, 35N-cm (1G) wrist torque, front-to-back shoulder clavicle compliance)

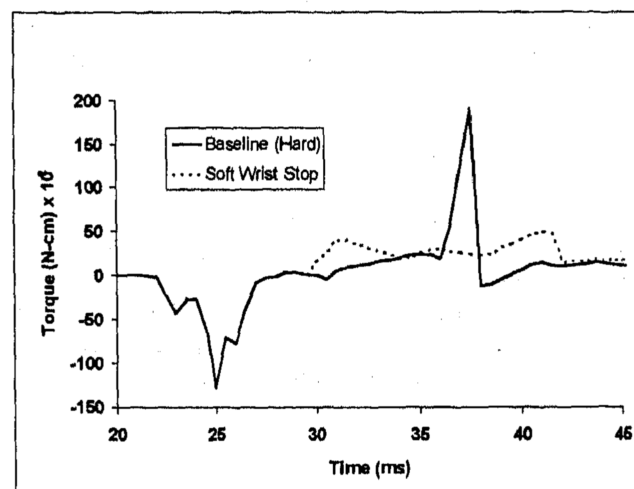


Figure 8. Forearm C.G. Torque for Hard and Soft Wrist Joint Stops

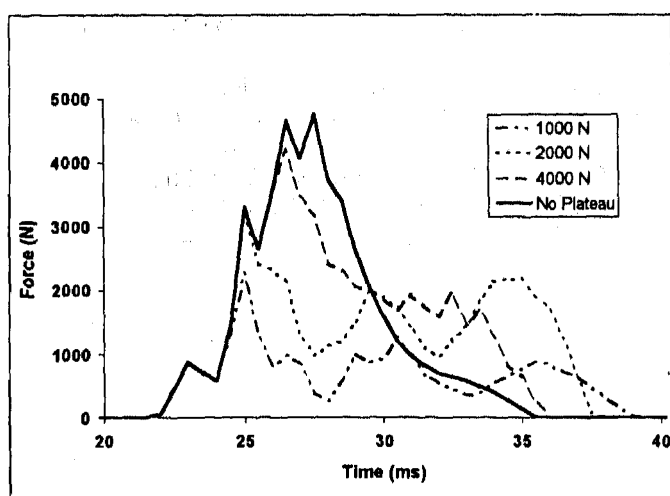


Figure 9. Upper Extremity-to-Flap Contact vs. Flap-to-Bag Force Plateau

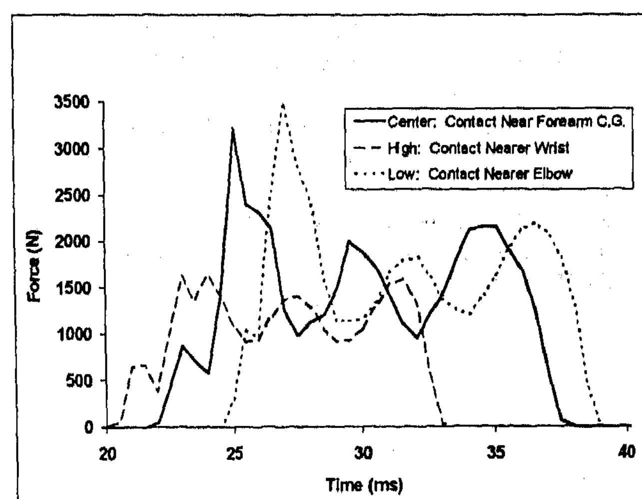


Figure 10. Upper extremity-to-Flap Contact Force vs. Bag Module Location (200 N Flap-to-Bag Plateau)

CONCLUSIONS AND RECOMMENDATIONS

The results obtained in this study support the use of disarticulated at the shoulder upper extremity component tests when conducting research into upper extremity-airbag interactions. The use of more complicated, full-body, full-motion tests is not expected to yield significantly different results, but it will increase the cost and data-variability associated with each experiment.

Peak loading of the upper extremity is associated with the airbag module flap, and occurs within 5 msec of punch-out. From this observation it seems that the most fruitful area for future research is the minimization of this contact force. Thus far, the variables which have been studied that are most significant include the coupling between the bag and flap, and the position of the bag module. Over a reasonable range of values, grip strength and wrist torque play a negligible role with regard to the flap contact force.

This ATB model in conjunction with laboratory testing can be used to examine other variables in the future. Additional tests may study various airbag deployment speeds, different flap tear-seam orientation (vertical instead of horizontal), or flap size (two small flaps versus one large flap).

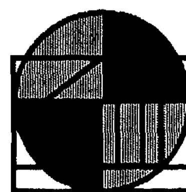
ACKNOWLEDGMENTS

The authors would like to acknowledge the support and guidance of Nopporn Khaewpong (NHTSA), Rolf H. Eppinger (NHTSA), and Shashi Kuppa (Conrad Technologies). This work was supported in part by DOT NHTSA Cooperative Agreement DTNH22-95Y-07029 and by the University of Virginia School of Engineering and Applied Science Department.

REFERENCES

1. Marco F, Garcia-Lopez A, Leon C, et al: Bilateral Smith Fracture of the Radius Caused by Airbag Deployment. J Trauma 40:663, 1996
2. Freedman EL, Safran MR, Meals RA: Automotive Airbag-Related upper Extremity Injuries: A Report of Three Cases. J Trauma 38:577, 1995
3. Huelke DF, Moore JL, Compton TW, et al: Upper Extremity Injuries Related to Airbag Deployments. J Trauma 38:482, 1995
4. Kirchhoff R, Rasmussen SW: Forearm Fracture Due to the Release of an Automobile Airbag. Acta Orthop Scand 66:483, 1995
5. Roth T, Meredith P: Hand Injuries From Inflation of an Airbag Security System. J Hand Surgery 18B:520, 1993
6. Smock WS, Nichols GR: Airbag Module Cover Injuries. J Trauma 38:489, 1995
7. Backaitis SH, Hicks M: Musculoskeletal System of Human Shoulders and Arms. In Biomechanics of Impact Injury and Injury Tolerances of the Extremities. Warrendale, PA, Society of Automotive Engineers, 1996, pp845-857
8. Kuppa SM, Yeiser CW, Oslon MB, et al: RAID - An Investigation Tool to Study Airbag/Upper Extremity Interactions. SAE Paper (to be published), SAE International Congress and Exposition, Detroit, MI, 1997
9. Saul RA, Backaitis SH, Beebe MS, et al: Hybrid III Dummy Instrumentation and Assessment of Arm Injuries During Air Bag Deployment. SAE Paper 962417, 40th Stapp Car Crash Conference, Albuquerque, New Mexico, 1996
10. Lemay MA, Crago PE: A Dynamic Model for Simulating Movements of the Elbow, Forearm, and Wrist. CS 29:1319, 1996
11. Raikova R: A General Approach for Modeling and Mathematical Investigation of the Human Upper Limb. J Biomechanics 25:857, 1992
12. Amis AA, Dowson D, Wright V: Analysis of Elbow Forces Due to High-Speed Forearm Movements. J Biomechanics 16:825, 1980
13. Amis AA, Dowson D, Wright V: Elbow Joint Force Predictions for Some Strenuous Isometric Actions. J Biomechanics 13:765, 1980
14. Andrews JG, Youm Y: A Biomechanical Investigation of Wrist Kinematics. J Biomechanics 12:83, 1979
15. Youm Y, Dryer RF, Thambyrajah K, et al: Biomechanical Analysis of Forearm Pronation-Supination and Elbow Flexion-Extension. J Biomechanics 12:245, 1979
16. Chai EY, Morrey BF: Three-Dimensional Rotation of the Elbow. J Biomechanics 11:57, 1978
17. Humanscale, Chart No. 5a, produced by Henry Dreyfus Associates

CONTACT FOR ADDITIONAL INFORMATION



University of Virginia
Automobile Safety Laboratory
1011 Linden Avenue
Charlottesville, VA 22902
Phone (804) 296-7288
Facsimile (804) 296-3453

970399

RAID - An Investigative Tool to Study Air Bag/Upper Extremity Interactions

Shashi M. Kuppa, Mitchell B. Olson, Charles W. Yeiser, and Lynne M. Taylor
Conrad Technologies, Inc.

Richard M. Morgan and Rolf H. Eppinger
National Highway Traffic Safety Administration

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A study of frontal collisions using the NASS data base showed that there were four times as many arm injuries to belt restrained drivers who had an air bag deploy than for the drivers who were simply belted. By far, the distal forearm/hand was the most commonly injured region. Hard copy review identified two modes of arm injury related to the deploying air bag: 1) The arm is directly contacted by the air bag module and/or flap cover, and 2) The arm is flung away and contacts an interior car surface.

Based on the field studies, a mechanical device called the Research Arm Injury Device (RAID) was fabricated to assess the aggressivity of air bags from different manufacturers.

Results from static air bag deployment tests with the RAID suggested that the RAID was able to clearly distinguish between the aggressive and non-aggressive air bags. Maximum moments ranging between 100 Nm and 650 Nm, and hand fling velocity ranging between 30 and 120 km/h were measured on the RAID in these tests. In general, the aggressive air bags imparted a maximum moment on the RAID above 300 Nm and a hand fling velocity above 70 km/h.

Two factors were identified as critical to the test setup. The first was the orientation of the arm with respect to the air bag module. The second was the distance of the arm from the plane of the air bag module face. The maximum moment and fling velocity increased when the initial distance between the RAID and the air bag module was reduced.

INTRODUCTION

Though recent studies [1] suggest that air bags have reduced the number of fatalities, preliminary reports by Huelke [2,3] from a limited number of accident cases involving deployed air bags suggested that air bags may cause upper extremity injuries. These accident cases also suggested that the intensity and frequency of air bag related upper extremity injuries may be related to the velocity of impact, occupant stature and gender, and various other parameters. Since these case studies were anecdotal and limited in number, no conclusive statement could be made.

The objective of this research effort was to get a better understanding of the problem related to upper extremity trauma induced by the air bag and develop a Research Arm Injury Device (RAID) to study this phenomenon. This objective was achieved in two phases.

Phase I: In order to better understand upper extremity injuries associated with a deploying air bag various accident data bases were examined. In addition, detailed review of accident cases in the NASS data base and NHTSA sponsored hospital and special investigation data bases were conducted.

Phase II: The objective in this phase was to develop the Research Arm Injury device and to study arm/air bag interaction and assess the relative aggressivity of different air bags.

PHASE I: REVIEW OF ACCIDENT DATA

The objective of this phase was accomplished by analyzing various accident databases, including NASS, NHTSA Special Investigations and from NHTSA-sponsored hospital studies at Jackson Memorial, and the MD and NJ Trauma Centers. Examining the arm trauma to drivers whose air bags deployed provided valuable information for the development of the RAID.

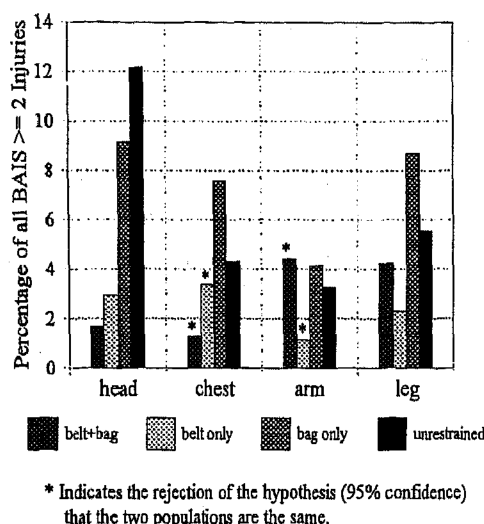


Figure 1: Distribution of BAIS≥2 injuries. 1988-94 weighted NASS files.

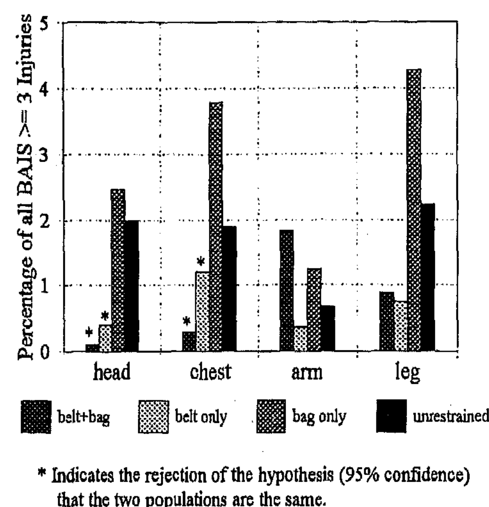


Figure 2: Distribution of BAIS≥3 injuries. 1988-94 weighted NASS files.

ANALYSIS OF NASS DATABASE - First, a general overview of the incidence of injury was performed. The National Accident Sampling System (NASS) was used to generate the distribution of injuries to various body regions of drivers under different restraint conditions. Computer records were examined from the years 1988-1994 to quantify the prevalence of injuries to drivers (over the age of 15) in passenger cars. Frontal impacts with a primary direction of force between 11 and 1 o'clock were considered. In addition, the selected drivers must not have been ejected from the vehicle or been involved in a rollover collision.

Four restraint categories were considered - 1. Unrestrained: The driver was not restrained by any belt and the air bag was not present or did not deploy, 2. Bag only: The driver was not restrained by any belt and the air bag deployed, 3. Belt only: The driver was restrained by the shoulder and lap belt and an air bag was not present or did not deploy, 4. Belt+bag: The driver was restrained by the shoulder and lap belt and the air bag deployed.

Four body regions were considered in the analysis - 1. Head (head, face, neck), 2. Chest (abdomen, chest, back, shoulder), 3. Arm (wrist, hand, upper limb, elbow, upper arms, forearm, and humerus), 4. Leg (pelvis, knee, lower limb, thigh, ankle/foot).

The data were analyzed using the maximum injury to each body region (BAIS) based on the Abbreviated Injury Scale. For example, if there were two injuries to a defined body region, then the injury with the larger AIS would be the BAIS value. The BAIS can be thought of as a modified-MAIS (Maximum Injury to a Body Region) such that for each driver, there are four maximum AIS values instead of one value. Drivers who were not injured at all had a BAIS value equal to zero in each of the body regions. The frequency of injury at BAIS≥2 and BAIS≥3 was computed as

$$Freq. = \frac{no.\ of\ drivers\ with\ injuries}{total\ no.\ of\ drivers\ in\ each\ restraint\ system}$$

The percentage frequency of injury at BAIS≥2 and 3 for various restraint conditions are presented in Figures 1 and 2. The percentage of BAIS≥2 (and BAIS≥3) injury for a particular body region was compared between two restraint systems and statistical significance at the 95 percent confidence level were computed. It was found that 1.1% of drivers restrained by a 3-point belt only experienced an upper extremity injury; however, 4.4% of drivers restrained by a deploying air bag and 3-point belt sustained a BAIS≥2 upper extremity injury. The difference in these proportions was found to be statistically significant.

DETAILED CASE REVIEWS - Analysis of the NASS data base suggested that belt restrained drivers are four times more likely to experience upper extremity trauma when an air bag deploys than when no air bag deploys. In order to investigate in detail the cause of these arm injuries, individual cases of drivers who sustained arm injuries while restrained by air bags were culled from the database sources. These selected files were then examined in detail. About one hundred cases (65 from NASS) were selected for in depth review. Each of these cases met the selection criteria that an air bag deployed and there was an arm injury of AIS≥2.

Injury mechanisms were determined based on information from: (1) interview with the driver, (2) exterior and interior vehicle damage, and (3) all available medical records. From the examination of these select accident cases, this study linked three principle injury mechanisms involving air bag deployment. The first mechanism consisted of injuries resulting from direct contact of the arm with the module cover and/or bag. The other two resulted from the arm being flung against interior car parts: either projected upwards toward the front windshield/A-pillar, or flung by the air bag toward the left side door. These three mechanisms are depicted in Figure 3. Of the 65 selected accident cases in NASS, approximately half (32 cases) were judged to be related to the three principal injury mechanisms involving air bag deployment, as shown in Figure 4.

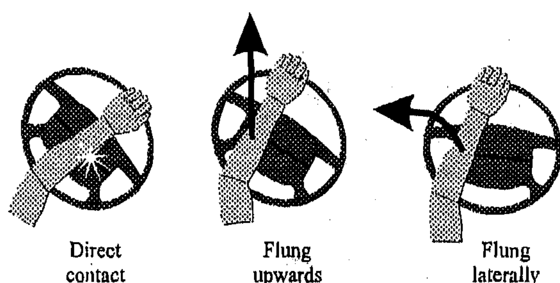


Figure 3: Three principal mechanisms of upper exremity injury due to air bag deployment.

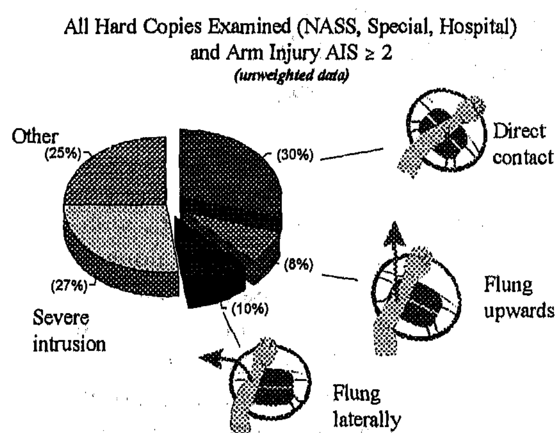
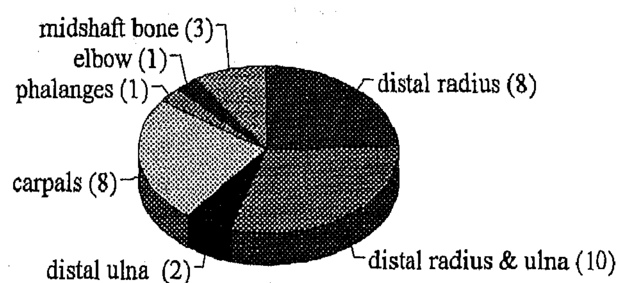


Figure 4: Distribution of injury mechanisms in the selected accident cases from NASS.

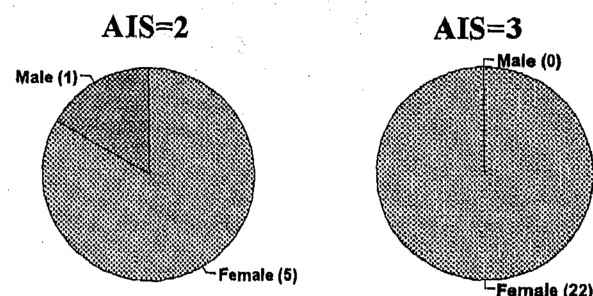
In the 32 accident cases in NASS related to the three principal injury mechanisms, the vast majority of the injuries were fractures of the distal portion of the limb, as shown in Figure 5. Wrist injuries were more common in mechanisms where the forearm was flung by the deploying air bag. In accidents cases with $\Delta V \leq 24$ km/h (from all data bases where the three primary mechanisms were judged to be the cause of AIS≥2 upper extremity trauma), injury occurred primarily in females, as can be seen in Figure 6. Almost all these drivers were female who were under 5'6" and less than 135 lbs.



87% of all fractures were to the distal end.

Contains 32 unweighted NASS (88-94) cases where the injury could be related to the deploying bag. There is one case with multiple (2) entries in the pie.

Figure 5: Location of fractures in the 32 NASS cases related to the three principal injury mechanisms.



Vast majority were under 5'6" and under 135lb.

Figure 6: Effect of gender on arm injuries induced by air bag deployment at low ΔV. (NASS, hospital, and special investigation data bases)

PHASE II: THE RESEARCH ARM INJURY DEVICE

The detailed study of NASS cases suggested two distinct types of upper extremity injury mechanisms related to air bag deployment. The first type of mechanism is where the arm is directly contacted by

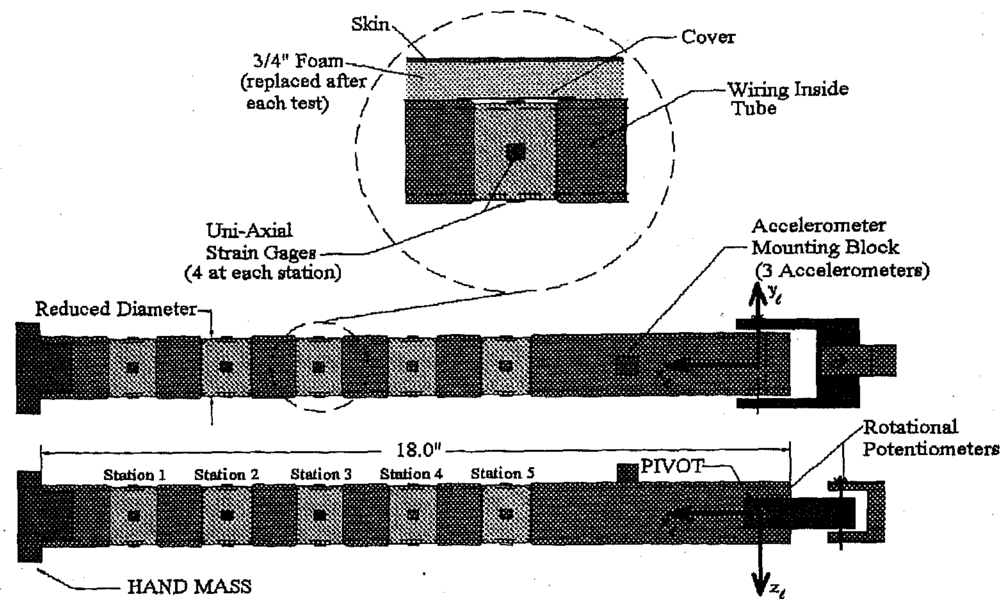


Figure 7: The Research Arm Injury Device

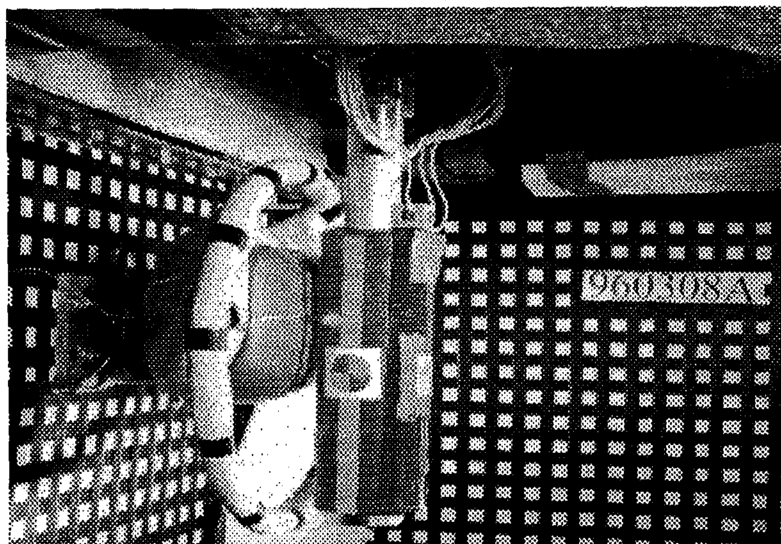


Figure 8: Side view of the RAID test setup.

the deploying bag and/or module flap cover. This results in high bending moments on the arm causing injury in some instances. The second type of mechanism involves the arm being flung away at a high velocity and impacts some portion of the vehicle interior or the driver body. The objective in this phase of the research program was to develop a simple research tool to evaluate the relative severity of air bags from different manufacturers taking into consideration these established modes of upper extremity trauma.

DESCRIPTION OF THE RAID - The Research Arm Injury Device or RAID was developed as an investigative tool to study arm-air bag interactions. The RAID is made from an aluminum tube with double pivot attachment to allow motion along two axes. It has the approximate cross-sectional properties and weight (1.6 kg) of a human forearm. A small mass (0.5 kg) is attached to the free end of the RAID to simulate the hand. The length of the RAID was extended to 46 cm to protect the pivot attachments from the deploying air bag. The RAID instrumentation includes five stations of diametrically opposed strain gages to measure moments along two axes. In addition, the pivot rotations are measured by two angular potentiometers and the accelerations are measured by a triaxial accelerometer block attached approximately at mid-length of the RAID. The RAID is covered with 2 cm of foam and rubber

skin similar to that on the Hybrid III mid-forearm. Figure 7 illustrates the details of the RAID.

EXPERIMENTAL SETUP - A side view of the test setup with the RAID is shown in Figure 8. The RAID was hung vertically down in front of the steering wheel and was allowed to rotate at the pivots in two directions as shown in the figure. It was possible to move the RAID in any direction to achieve the correct positioning with respect to the air bag module. The distance from the surface of the RAID to the plane of the steering wheel rim was changed according to test specifications. The steering wheel was rotated to achieve the desired test configuration (the orientation and position of the module tear seam with respect to the RAID). A load cell was located behind the steering wheel to measure reaction forces. The time of air bag cover opening was determined using break wires over the tear seam. A backstop with foam padding was used to stop the RAID after the air bag propelled it. The tests with the RAID were captured on film at 1000 frames per second by two high speed cameras for the front and side views.

SELECTION OF AIR BAGS - Six different air bag modules were selected for testing. The selection of the air bags was based in part upon the results from the NASS analysis and also upon tank test data. The System L was identified as a less aggressive bag from the NASS studies. The System J, System T, and the System P air bags were identified as less aggressive by the manufacturer. The System H and the System K were identified as aggressive air bags from the NASS studies as well as from tank test data. System H has a large heavy asymmetric module cover and System K has a very high tank pressure. These features make the Sytems H and K modules aggresssive in nature.

Table 1 presents various characteristics of these air bag modules. Figure 9 shows roughly scaled sketches of the different modules and covers. All modules except the System J air bag have an "H" tear pattern. The System J has an "I" tear pattern.

TEST CONFIGURATION - Thirty-four tests were conducted with the RAID using different air bag modules in different test configurations and distances from the plane of the steering wheel rim. A test configuration is defined as the orientation and position of the air bag module tear seam with respect to the RAID. After experimenting with various configurations, three main configurations, shown in Figure 10, were selected to compare air bag aggressivity.

Configuration I was designed to represent a person turning the wheel with their arm in front of the steering wheel. This configuration consists of a hand at 11 o'clock and an elbow at 4 o'clock, with the steering wheel rotated counter clockwise by 90 degrees from the neutral position.

Configurations V and VI were designed to study the most severe interaction of the arm with the air bag. Configuration V consists of the forearm located over the center of the module and oriented perpendicular to the tear seam of the module. Configuration VI consists of the forearm located over the center of the module and oriented parallel to the tear seam of the module. As a driver in a vehicle makes various maneuvers and adjustments, his or her arms are in continuous motion, so these configurations are thought to be plausible, although possibly less likely.

In addition to varying the test configurations, the distance of the RAID from the plane of the steering wheel rim was varied. Tests were conducted with the distance of the RAID from the plane of the steering wheel rim at 1.3 cm, 7.6 cm, and 12.7 cm.

TEST RESULTS - Results from static air bag deployment with the RAID suggested that the RAID was able to clearly distinguish between aggressive and non-aggressive air bags. The maximum moments, hand velocity, resultant acceleration, and steering column force were consistently higher for the aggressive air bags than the air bags identified as non-aggressive, as shown in Figures 11 to 14. The hand velocity was computed using the rotational potentiometer data.

Maximum moments ranging between 100 Nm and 650 Nm, and hand fling velocity ranging between 30 and 120 km/hr were measured on the RAID in these tests.

All the air bag module covers opened between 3 - 5 msec after the firing pulse. The maximum moments and accelerations occurred very early in the event - 5 to 20 msec (during the punch-out phase.) The maximum hand velocity occurred later when the RAID was propelled by the air bag and it was no longer in contact with the bag. The maximum velocity occurred between 30 - 40 msec. The maximum moments and velocity occurred earlier for the aggressive air bags than for the non-aggressive air bags.

When the initial distance of the RAID from the plane of the steering wheel rim was increased from 1.3 cm to 7.6 cm, the maximum moments, maximum accelerations, and maximum hand velocity decreased considerably. Further increasing the initial distance of the RAID from the air bag module to 12.7 cm did

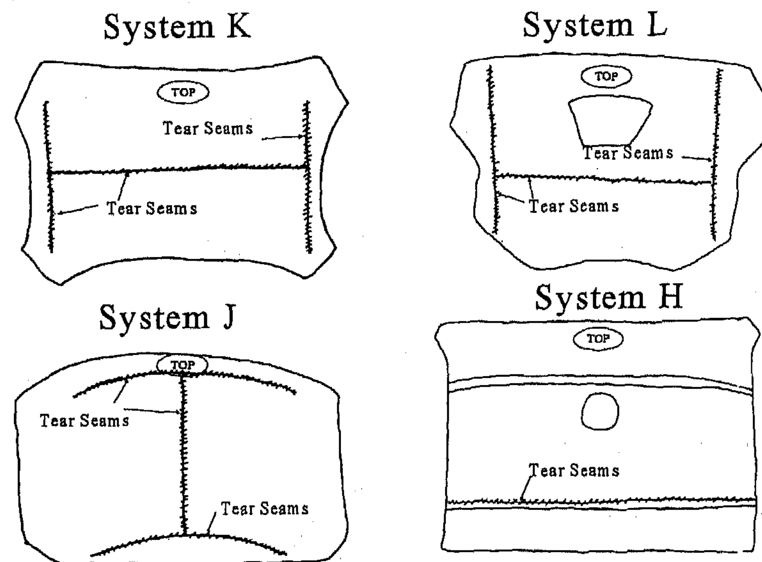


Figure 9: Sketches of air bag module covers indicating the tear patterns.

Table 1. Characteristics of air bag modules
(all measurements in centimeters)

Air Bag Characteristics	System H	System L	System K	System J
Diameter of steering wheel	38	38	38.7	39.7
Location of module plane wrt. To wheel	0.32 above	0.95 above	0.64 above	0.64 above
distance from top of rim to tear line	25.9	23.8	22.2	20.0
Vertical height of module	17.8	15.2	15.2	15.2
Horizontal width of module at seam	20.3	17.1	19.1	21.6
Distance from top of module to seam	13.8	9.1	7.8	10.8
Thickness of flaps	0.32	0.44	0.51	0.32
Vertical height of air bag	68.6	69.9	63.5	66.0
horizontal width of air bag	68.6	63.5	66.0	63.5
number of tethers	4	2	3	none
length of tethers	26.7	27.9	31.8	--

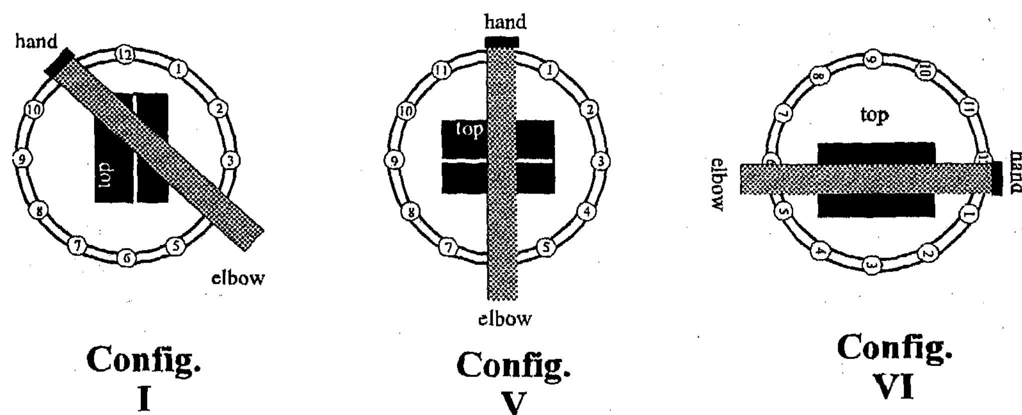


Figure 10: Three test configurations with the RAID.

not change the maximum values of moments, accelerations, and velocity.

The maximum moment measured was located at station 4 which is 20 cm from the hand mass of the RAID. This station was approximately located at the air bag tear seam where the greatest effect of the punch out due to module cover opening was experienced. Figure 15 presents the peak moments at each station for the tests with the System L air bag in various configurations. As can be noted from Figure 15 the moments are lower for the tests where the initial distance of the RAID from the air bag module was increased so that the cover flaps did not impinge it during the punch out phase.

Two major parameters measured on the RAID were identified to clearly distinguish between the aggressive and non-aggressive air bags. They are the maximum resultant moment and the maximum hand velocity. Figure 16 presents the maximum moment versus maximum hand velocity for all the tests in Configurations I, V, and VI. This figure suggests that the aggressive air bags imparted a maximum moment on the RAID above 300 Nm and a hand fling velocity above 75 km/h. In general, the maximum moments and hand velocity decreased when the initial distance between the air bag module cover and the RAID was increased.

The orientation and the distance of the arm with respect to the air bag module were identified as critical to the test setup. Of all the test configurations examined, configuration V was most severe. Repeatability tests were conducted with the System J and System K air bags. Good repeatability in the acceleration, moments, and rotation signals were evident.

ESTIMATION OF FRACTURE TOLERANCE OF THE HUMAN FOREARM

The relative severity of the various air bag modules can be determined from the RAID tests. However, in order to ascertain whether a particular air bag module would cause injury to the forearm, the fracture tolerance of the human forearm had to be determined.

As a first step, the fracture tolerance of the human forearm was estimated from static 3-point bending test data of isolated bones reported by Yamada [4] and Melvin [5]. Melvin reported the breaking load in bending of the radius and ulna of male and female human post-mortem specimens. Assuming the bending strength of the radius and ulna to be additive and assuming a dynamic factor of 1.5 as suggested by Melvin, the dynamic fracture tolerance of the human forearm was estimated to be 115 Nm for male specimens and 90 Nm for female specimens. This is similar to the forearm injury reference value of 120 Nm suggested by Saul [6].

Figure 11 suggests that all the air bags tested produced bending moments on the RAID which exceeded the estimated fracture strength of the forearm. The RAID was designed as a simple investigative tool and not necessarily a biofidelic device. Hence, the responses measured on the RAID may not represent human response. In order to identify whether a certain air bag would injure the human forearm, it would be necessary to scale the RAID response to represent that of a human forearm.

CONCLUSIONS

Detailed study of the NASS database suggested that arm injuries occur with greater frequency in

drivers whose vehicle is equipped with a deploying air bag. In addition, results of the study suggested that belt restrained drivers who had an air bag deploy are four times more likely to experience arm injuries than belt restrained drivers without a deployed air bag. The three most common arm injury mechanisms due to air bag deployment are: (1) Direct contact with the air bag and/or flap, (2) The arm is flung upwards and hits the A-pillar or front windshield, and (3) The arm is flung to the side. The distal portion of the upper extremity was the most commonly encountered injury in these cases. At low velocities (less than 15 mph), AIS=3 injuries have thus far only occurred in females, most of whom were small in stature and weight.

A mechanical device called the Research Arm Injury Device was developed to test air bag aggressivity. Preliminary testing with the RAID suggested that this device was able to distinguish between aggressive and non-aggressive air bags with reasonable accuracy. Maximum moments ranging between 100 Nm and 650 Nm, and hand fling velocity ranging between 30 and 120 km/hr were measured on the RAID in these tests.

The orientation of the arm with respect to the air bag module and the distance of the arm from the air bag module were identified as critical to the test setup. The most severe case occurred when the RAID was directly over the center of the module and the tear seam was perpendicular to the RAID. When the initial distance between the RAID and the air bag module was increased, the maximum moment, acceleration, steering column loads, and hand velocity decreased.

The punch out phase when the deploying air bag/module cover first interacts with the RAID during its deployment was when the RAID experienced maximum moments and velocities. The punch-out phase is therefore the most critical period of air bag deployment for causing trauma to the upper extremities.

FUTURE RESEARCH

The estimated dynamic fracture tolerance is based on broad assumptions using static 3-point bending tests of isolated bones. In order to get more accurate estimates of dynamic fracture tolerance, dynamic 3-point bending tests of the whole forearm should be conducted. Preliminary data from such tests suggest an average dynamic fracture tolerance of 125 Nm.

Since the RAID is not biofidelic, the moments measured on the RAID may not be the moments which a human arm would experience when at close proximity to the air bag module in a realistic situation. Hence, in order to interpret the results from the RAID tests to determine the injury causing potential of particular air bag modules, testing with human cadaver forearms should be conducted in conjunction with the RAID tests to associate forearm trauma with the moments measured on the RAID. Preliminary data from upper extremity component tests suggest that the air bags identified as more aggressive from the RAID tests do indeed cause fractures to the cadaver forearm.

ACKNOWLEDGMENT

This study was supported by DOT NHTSA Contract No. DTNH22-92-D-07092. All findings and views reported in this manuscript are based on the opinions of the authors and do not necessarily represent the consensus or views of the funding organization.

REFERENCES

[1] Lund, Adrian k, and Ferguson, Susan A. "Driver Fatalities in 1985- 93 Cars with Air Bags." Journal of Trauma: Injury, Infection and Critical Care, Vol 38, No. 4, April 1995, pp. 469- 475.

[2] Huelke, Donald et al., "Upper Extremity Injuries Related to Air Bag Deployments,"Journal of Trauma, Injury, Infection and Critical Care, Vol 38, No. 4, April 1995, pp 459 -75.

[3] Huelke, Donald et al., "Upper Extremity Injuries Related to Air Bag Deployments," Journal of Trauma: Injury, Infection and Critical Care, Vol. 38, No. 4, April, 1995, pp. 469 - 475.

[4] Yamada, Hiroshi, "Strength of Biological Materials," The Williams & Wilkins Company, Baltimore, 1970.

[5] Melvin, J., Evans, F. G., "The Biomechanics of Trauma", Chapter 21, Edited by Nahum, A.M., Melvin, J., Appleton-Century-Crofts/Norwalk, Connecticut.

[6] Saul, R. A., et al., "Hybrid III Dummy Instrumentation and Assessment of Arm Injuries During Air Bag Deployment," Proceedings of the Fortieth Stapp Car Crash Conference, November, 1996, SAE No. 962417.

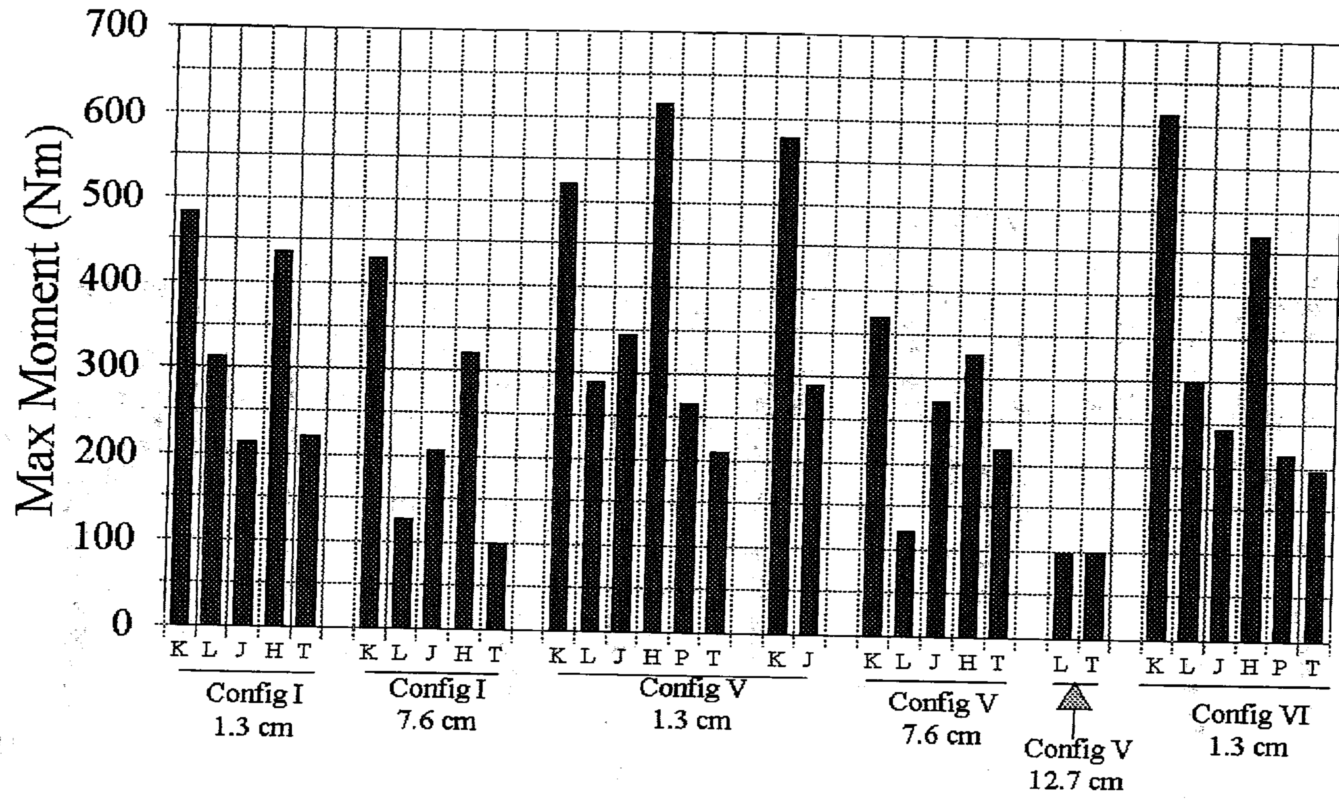


Figure 11: Maximum resultant moment by air bag module and test configuration. K- System K, H-System H, L-System L, J-System J, P - System P, and T - System T. I, V, VI, indicates the test configuration.

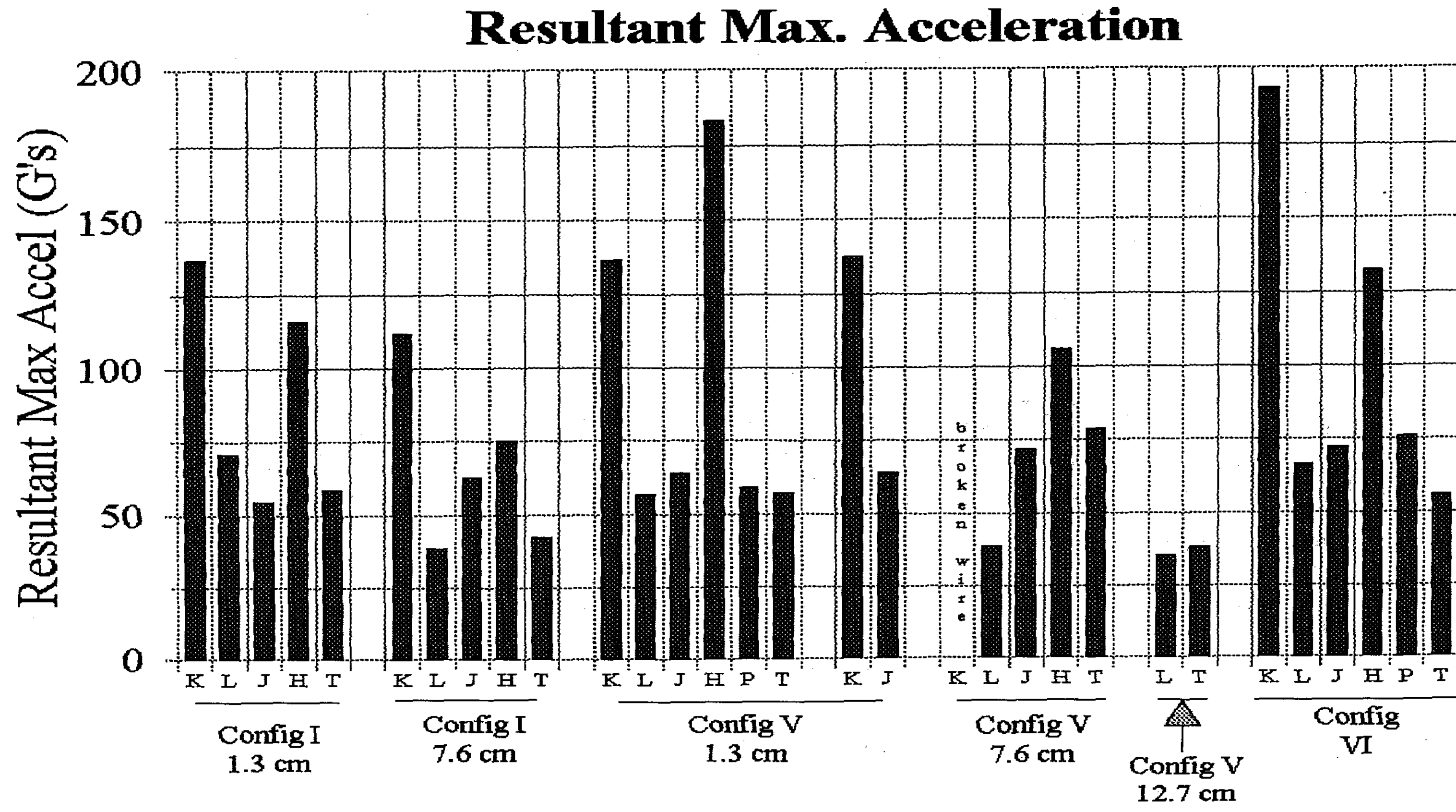


Figure 12: Maximum resultant acceleration by air bag module and test configuration. K- System K, H-System H, L-System L, J-System J, P - System P, and T - System T. I, V, VI, indicates the test configuration.

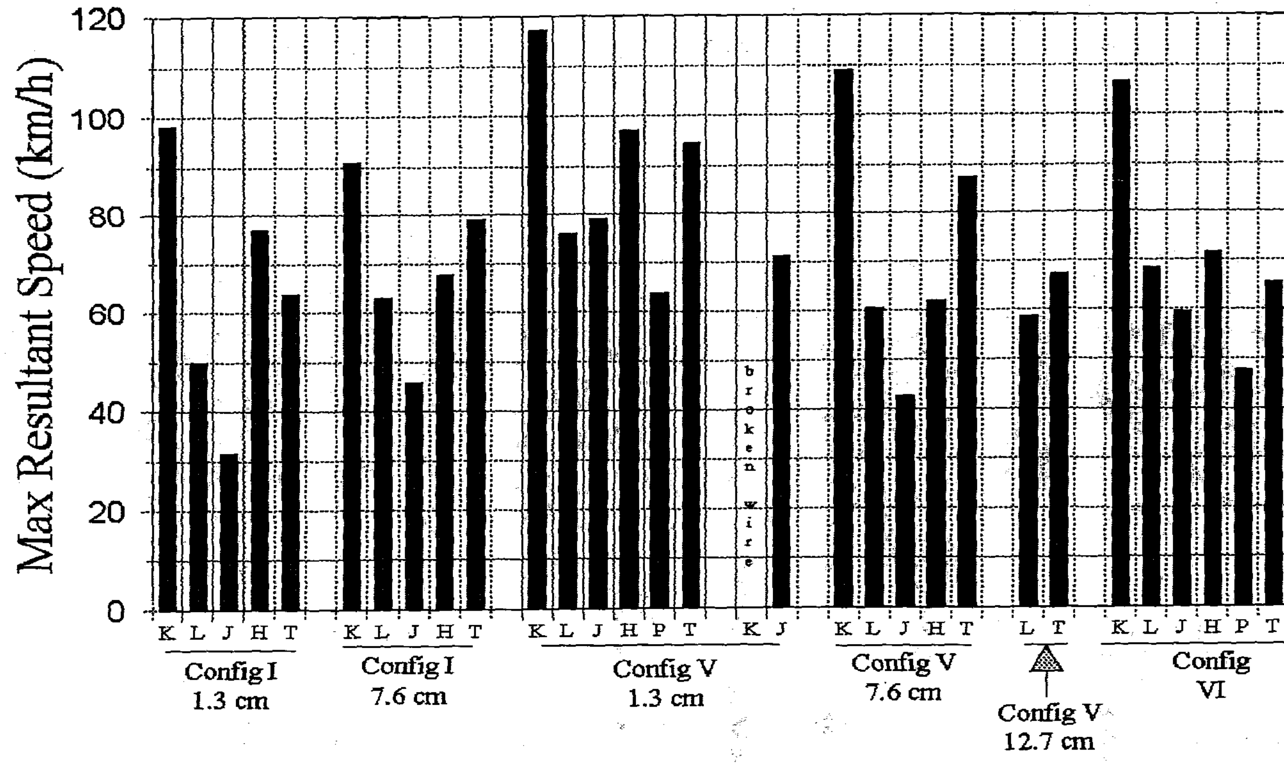


Figure 13: Maximum hand velocity by air bag module and test configuraton. K- System K, H-System H, L-System L, J-System J, P - System P, and T - System T. I, V, VI, indicates the test configuration.

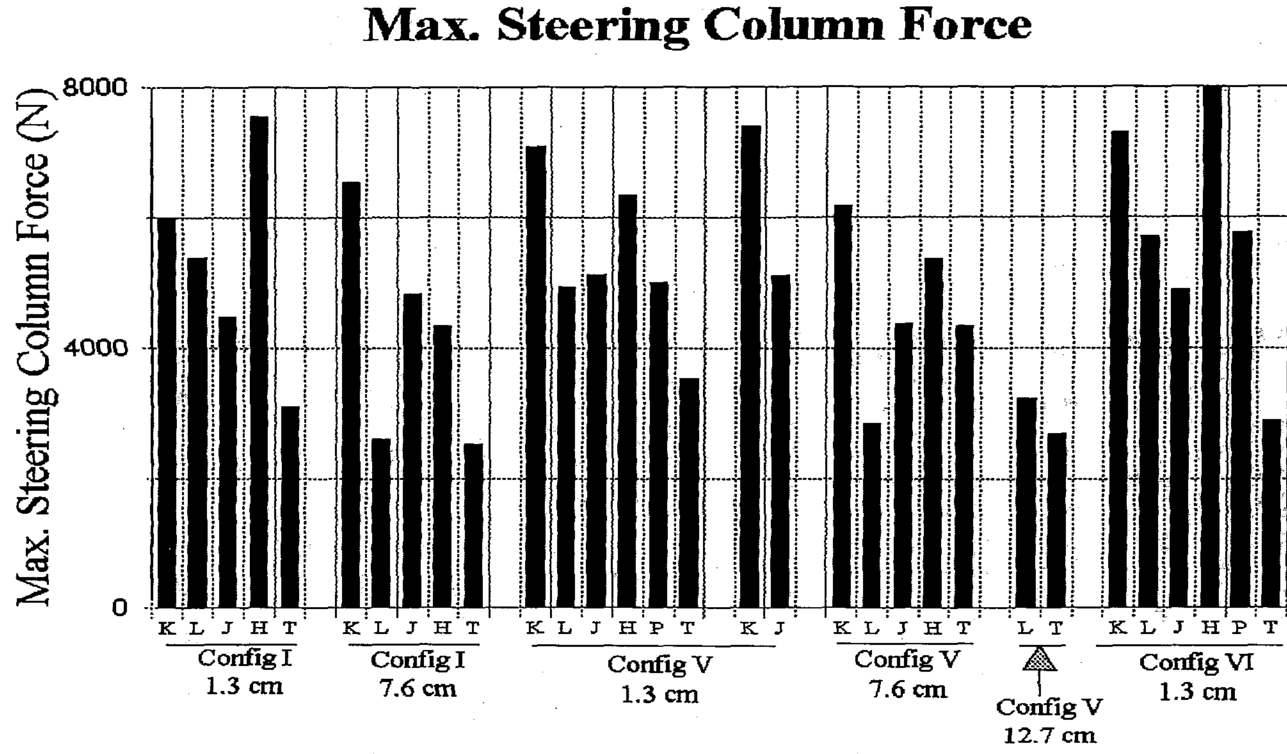


Figure 14: Maximum steering column loads by air bag module and test configuraton. K- System K, H-System H, L-System L, J-System J, P - System P, and T - System T. I, V, VI, indicates the test configuration.

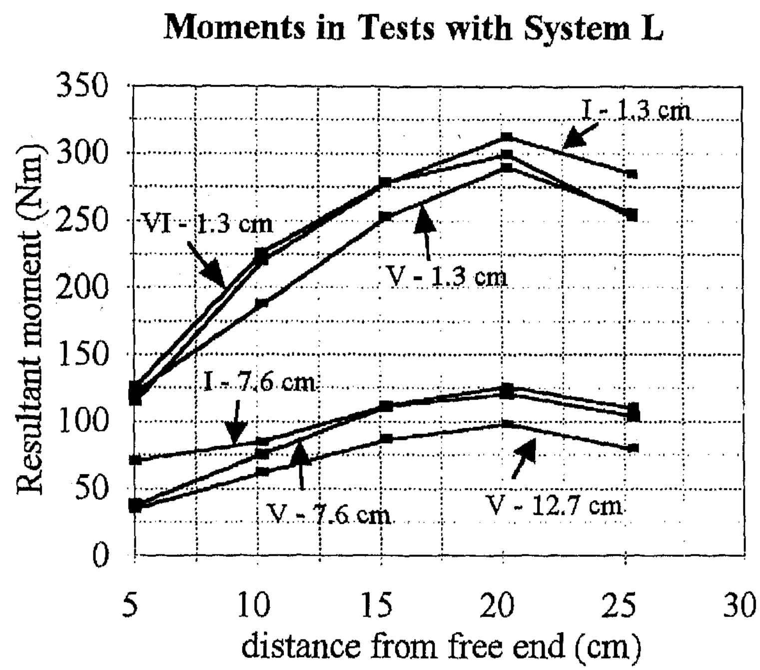


Figure 15: Peak moments for each station in tests with the System L air bag.

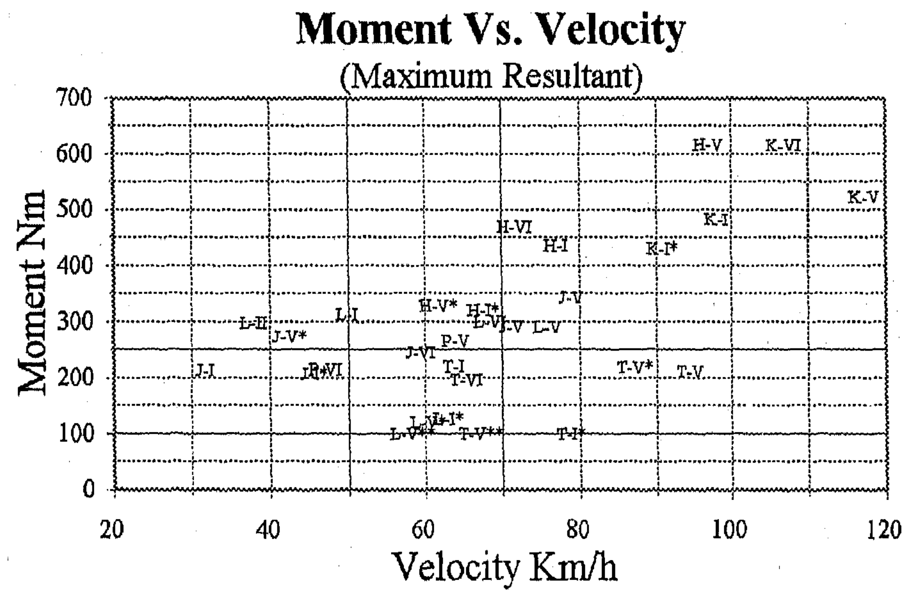


Figure 16: Maximum moment versus maximum velocity plots. K- System K, H-System H, L-System L, J-System J, P - System P, and T - System T. I, V, VI, indicates the test configuration. * =7.6 cm and ** =12.7 cm initial distance of RAID from steering wheel rim.

970400

Assessing Arm Injury Potential From Deploying Air Bags

Kristie L. Johnston, Kathleen DeSantis Klinich, and Daniel A. Rhule
Transportation Research Center, Inc.

Roger A. Saul
National Highway Traffic Safety Administration

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A study of the National Accident Sampling System (NASS) found an increase in upper extremity injuries when drivers were restrained by a seat belt and air bag as opposed to a seat belt alone. These injuries were attributed to forces from the air bag deploying or the air bag projecting the arm into vehicle components or the upper body of the driver. Two evaluation methods were used to assess the extent of injury and aggressiveness of different driver side air bags. The RAID, developed by Conrad Technology, and the Hybrid III instrumented arm, tested at the Vehicle Research and Test Center, were used in static testing to evaluate the effect of air bags on the arm.

The positions of the RAID and the Hybrid III arm simulated the arm in four different turning positions with the forearm across the center of the wheel. Both devices recorded arm moments and accelerations. Film analysis determined the cause of the peak resultant moment for each bag in the four configurations. The Hybrid III was able to attribute half of the moments that surpassed the Injury Assessment Reference Values (IARV) to forces from the air bag when deployed in close proximity to the forearm, while the other half of the moments that surpassed the IARV were from contact with the head, neck and chest of the dummy. In contrast, the RAID only recorded accelerations and moments due to the air bag but was unable to account for contact between the arm and the occupant or vehicle's interior. The Hybrid III instrumented arm and the RAID gave similar rankings of the severity of the different air bags.

INTRODUCTION

The main goal of the initial phase of the arm/air bag program was to determine if the RAID and Hybrid III instrumented arm led to similar conclusions regarding the arm injury potential of different air bag systems. To do this, testing was performed with the Hybrid III instrumented arm at the Vehicle Research and Test Center (VRTC) using the same testing configurations and air bags as those used with the RAID testing device.

BACKGROUND

ARM INJURY/AIR BAG INTERACTION STUDIES - A recent NHTSA study (conducted by Conrad Technology) found a higher incidence of arm injury in National Accident Sampling System (NASS) cases in accidents involving air bags. When drivers were restrained by only a three-point belt, about 1% experienced upper extremity injuries. Of those drivers restrained by both three-point belts and air bags, 4% suffered arm injuries. A more detailed NASS study of 36 specific cases of upper extremity injury to drivers found that half occurred in low impact collisions ($\Delta V<24$ km/hr), half occurred while turning, and half resulted in injuries with an AIS>3. Seventy percent of the injuries with an AIS>3 were inflicted on the radius and ulna. The increase in upper extremity injuries was attributed to two possible mechanisms. The first are direct forces from the air bag deploying on the arm. The second mechanism involves the air bag throwing the arm with high velocity into the driver body or vehicle surroundings.[1]

Another study conducted by the University of Michigan, Transportation Research Institute (UMTRI), has shown that upper extremity injuries usually occur in the form of abrasions, contusions, sprains and fractures in the forearm, wrist and fingers. It was found that these minor injuries occur in approximately 20%-25% of their sample data. These injuries were contributed to the forces afflicted by the air bag as it deploys (where the bag reaches velocities of 160-320 km/h), and contact with the occupant and vehicle surroundings. [3]

RAID

The RAID arm simulator was developed by Conrad Technologies Inc., to evaluate air bag severity relative to arm injury in a static laboratory setting. A drawing of the device appears in Figure 1. Its basic structure consists of an 46 cm aluminum tube and weighs 1.6 Kg (not including the weight of the hand mass of .5 Kg). The connecting end of the RAID consisted of a double pivoted joint which allowed rotations in two directions. The free hanging end of the RAID consists of a concentrated hand mass. The length of the RAID was extended

to 46 cm to protect the pivot attachments from the deploying air bag. This is longer than the Hybrid III forearm (approximately 28 cm), but is shorter than the entire length of the Hybrid III instrumented arm (approximately 58 cm).

Five strain gage "stations" permit moments in two directions along the length of the RAID, as seen in Figure 1. To conduct laboratory testing of different air bags, the steering wheel is mounted vertically to a test fixture. The RAID pivot is connected to a crossbar located above the steering wheel, so it hangs vertically and represents an arm positioned across a steering wheel. The distance between the RAID and steering wheel can be controlled by moving the crossbar.

HYBRID III INSTRUMENTED ARM - The Hybrid III instrumented arm [2] was developed to allow evaluation of air bag severity with a dummy in a vehicle setting. As illustrated in Figure 2, the structure of the device is the standard Hybrid III dummy forearm and hand which weighs approximately 1.5 Kg. The instrumented arm was equipped with four sets of strain gages and two accelerometers (as seen in Figure 2).

RAID/HYBRID III INSTRUMENTED ARM STRUCTURAL COMPARISON - Because of structural differences between the RAID and Hybrid III instrumented arm, the kinematics and measured responses of the two systems are expected to be different. The Hybrid III arm can bend at both the elbow and the shoulder, while the RAID testing device has only one joint at a location relative to a human arm about halfway between the elbow and shoulder. If we view each device as a beam in bending, the differences in length, diameter, material, and end conditions will lead to different responses to the same input.

Figure 3 shows the RAID and Hybrid III instrumented arm in a scaled comparison. The devices are lined up relative to how they are positioned to the steering wheel rim during testing. For the RAID, the hand mass is centered over the rim, while the junction of the Hybrid III hand and thumb is the approximate rim location of the Hybrid III instrumented arm. The side-by-side comparison indicates that none of the strain gage stations on the RAID line up with those of the Hybrid III instrumented arm.

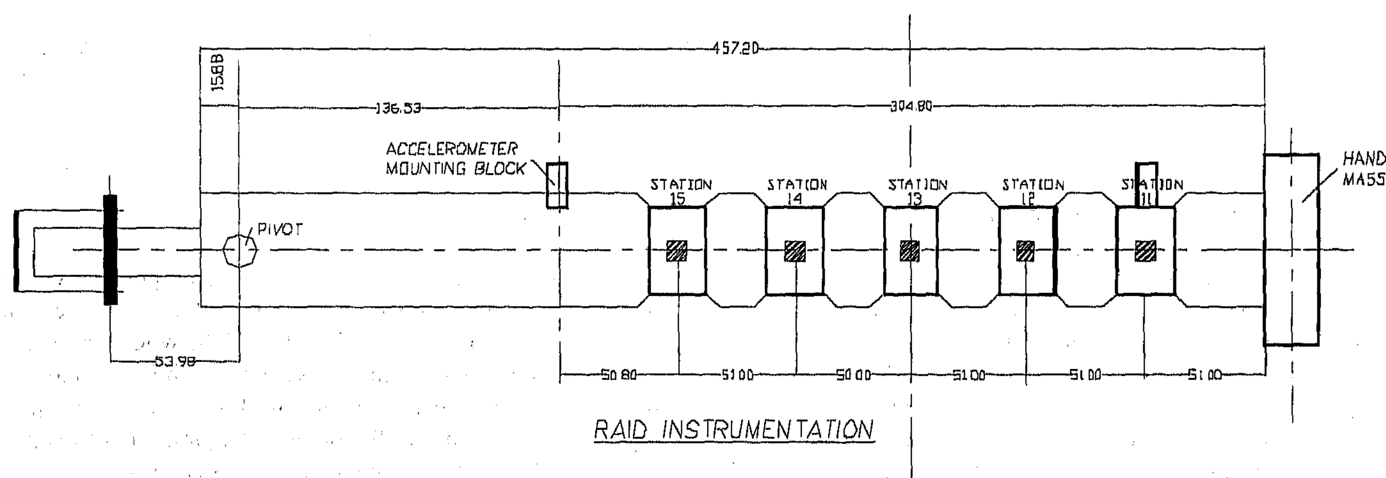


Figure 1. RAID testing device (dimensions in mm)

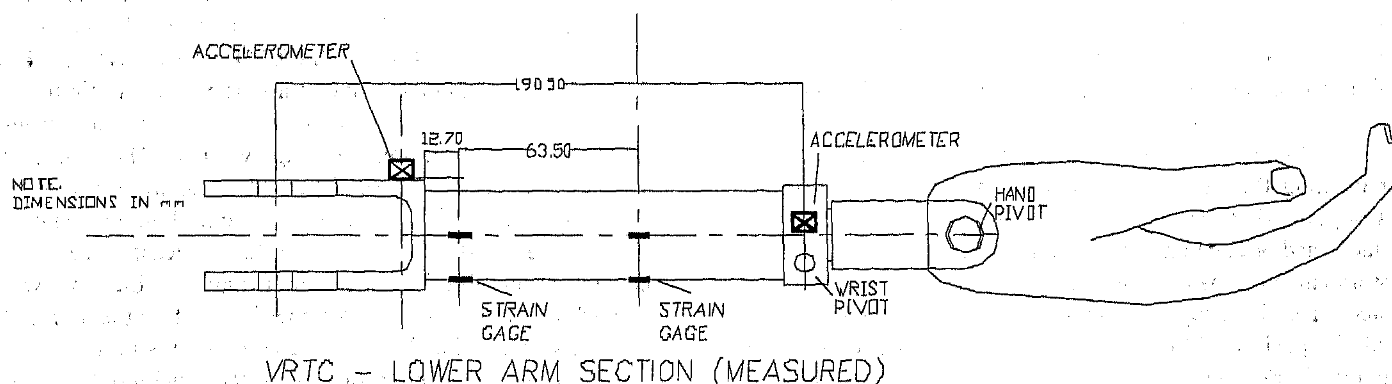


Figure 2--Hybrid III instrumented arm (dimensions in mm)

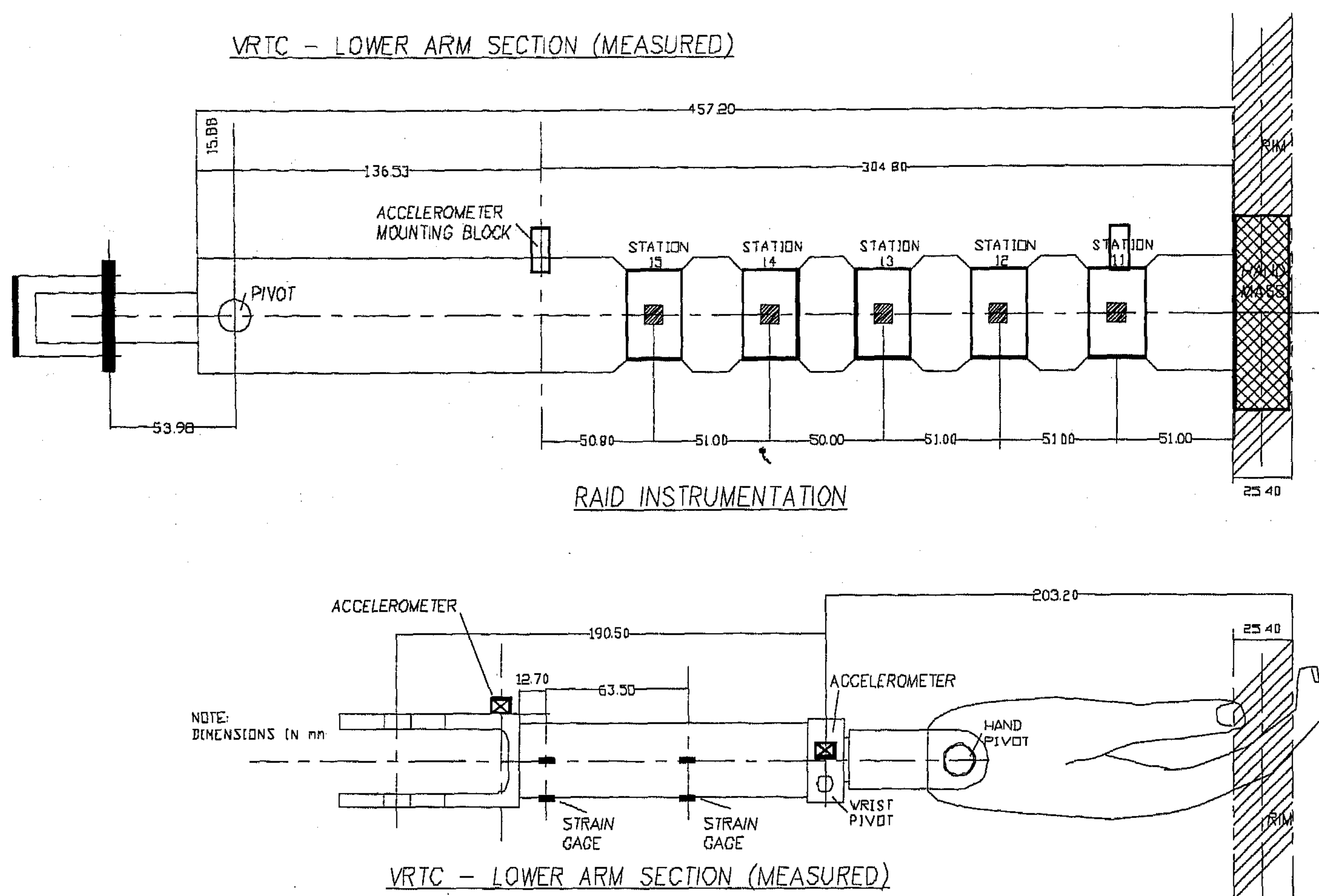


Figure 3. Comparison of RAID and Hybrid III instrumented arm (dimensions in mm)

ARM INJURY CRITERIA - An IARV was estimated in both the RAID and Hybrid III instrumented arm initial testing programs. The total dynamic bending moment for the arm was estimated to be 115 N*m for the RAID device.[1] A similar estimate was completed at VRTC to find a range for the peak load for males and for females. For the VRTC estimate, the female bending moment tolerance was 70-90 N*m, while the male tolerance was 120-150 N*m.[2]

PROCEDURE

TEST CONFIGURATIONS - The RAID was tested in a variety of different configurations that varied the orientation of the arm device relative to the steering wheel, and the distance between the arm and the air bag module. Four of the RAID configurations were selected for comparison testing with the Hybrid III instrumented arm, and were designated V-1, V-2, V-3, and V-4 as described in Table 1.

As mentioned before, the RAID testing protocol called for hanging the device from a pivot in front of the steering wheel. The wheel was turned and shifted to achieve the desired distance and orientation relative to the arm. Since the Hybrid III instrumented arm was being evaluated on the dummy seated in a sled buck, the setup was limited by how close the upper arm could be positioned to the dummy's torso without twisting the dummy. For example, the RAID hand/elbow was placed in the 12 and six o'clock positions in configurations V-2 & 3. With the dummy sitting centered in front of the wheel, it was not possible to place the elbow at the six o'clock position without moving the dummy's torso. Instead, the elbow was placed at a feasible angle (elbow at four o'clock) and the wheel was rotated to position the air bag seam relative to the arm in the same way it was positioned relative to the RAID. The same type of adjustment was done in the other configurations to keep the forearm and air bag seam orientations consistent with the RAID test conditions.

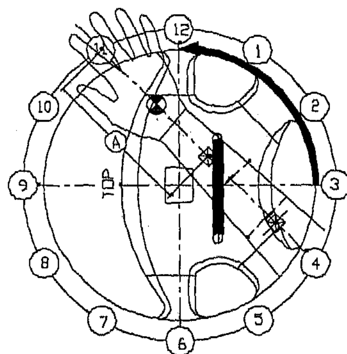
Figures 4, 5, and 6 illustrate specific examples of Hybrid III instrumented arm and RAID testing configurations. The arm positions are shown together with the amount of wheel rotation required. In all figures, the Hybrid III instrumented arm position is on the left and the RAID configuration on the right. The details regarding each setup are also included in Table 1, as are the configuration designations used by Conrad Technologies.

HYBRID III INSTRUMENTED ARM TEST MATRIX- In addition to using four of the same configurations from the RAID testing, the Hybrid III instrumented arm program also used four of the same air bag systems. The four air bag systems have been arbitrarily labeled H, J, K and L followed by the model year. For example, a 1995 model J would be noted as J-95. The RAID system was tested at least once in each configuration with four different air bags. Table 2 shows the Hybrid III instrumented arm test matrix. Each of the four air bag systems was tested in the V-1, V-2 and V-3 positions. Only the K-93 and the J-95 were tested in the V-4 position, because the RAID data indicated lower measured moments for this configuration

Table 1. Test Configurations						
VRTC Configuration	**Conrad Configuration**	**Nominal Distance from Wheel**	**Hand/Elbow "Clock" Position**		**Wheel Rotation**	
			H3 Inst Arm	**RAID**	**H3 Inst Arm**	**RAID**
V-1	I-1.3 cm	1.3 cm	11/4	11/4	90° CCW	90° CCW
V-2	V-7.6 cm	7.6 cm	10/4	12/6	60° CCW	0°
V-3	V-1.3 cm	1.3 cm	10/4	12/6	60° CCW	0°
V-4	VI-1.3 cm	1.3 cm	10/4	~12/6	150° CCW	90°

V-1

Hyb III INSTRUMENTED ARM
Arm @ 11-4 o'clock
Wheel rotated 90° CCW



RAID
Arm @ 11-4 o'clock
Wheel rotated 90° CCW

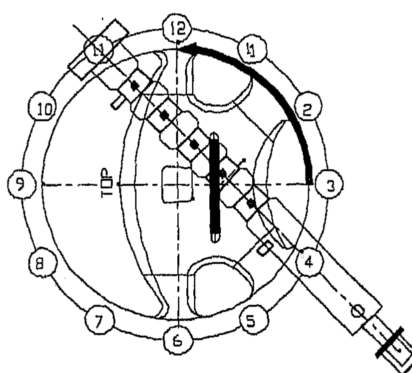
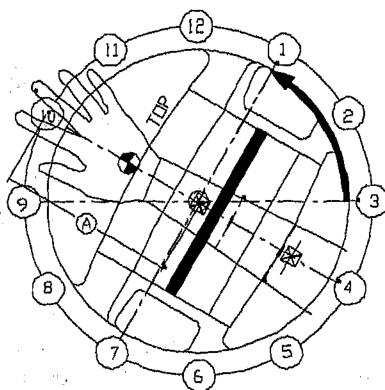


Figure 4. Hybrid III instrumented arm and RAID positions for configuration V-1

V-2, V-3

Hyb III INSTRUMENTED ARM
Arm @ 10-4 o'clock
Wheel rotated 60° CCW



RAID
Arm @ 12-6 o'clock
Wheel unrotated

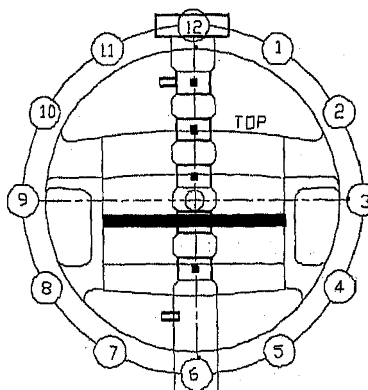


Figure 5. Hybrid III instrumented arm and RAID positions for configurations V-2 and V-3

V-4

Hyb III INSTRUMENTED ARM
Arm @ 10–4 o'clock
Wheel rotated 150° CCW

RAID
Arm centered over flaps
Wheel rotated 90° CCW

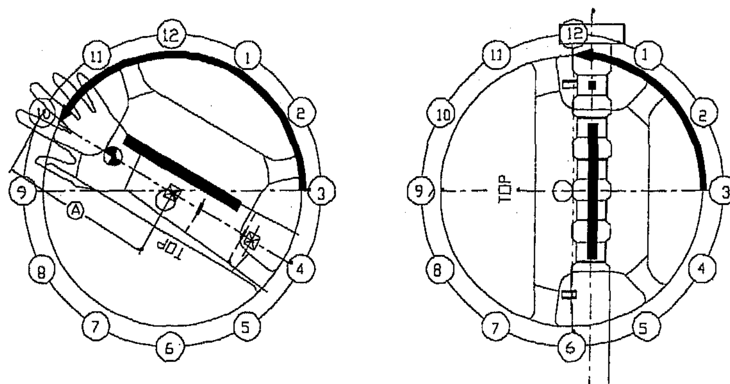


Figure 6. Hybrid III instrumented arm and RAID positions in configuration V-4

Table 2. Hybrid III Instrumented Arm Test Matrix				
VRTC Designation	**K-93**	**J-95**	**L-92**	**H-91**
V-1	✔	✔ ✔✔✔	✔	✔
V-2	✔	✔	✔	✔
V-3	✔	✔	✔	✔
V-4	✔	✔	Not Tested	

The J-95 air bag was tested with the Hybrid III instrumented arm four times at in the V-1 position to determine repeatability. In addition, some questions were raised regarding the Hybrid III thumb position on the steering wheel. Therefore, the thumb was placed under the rim for the first three tests and over the rim on the fourth test.

Table 3 summarizes additional tests performed in the V-1 and V-3 positions. Each bag was tested once in each configuration. The results generated by the 1994 air bags were compared with the corresponding 1996 models' results to assess any improvement in the later year model.

TEST PROTOCOL - The Hybrid III instrumented arm tests were conducted statically using a sled buck as the test fixture. The setup is shown in Figure 7. The dummy was positioned on a generic metal seat which could be shifted toward or away from the steering wheel. The steering wheel was attached to a column that allowed adjustment of the column and steering wheel angles. For the initial part of this program, the same steering wheel angle and column angle were used since the difference in vehicle measurements for the four air bags tested was minimal. The steering wheel and the column angles were set at 25°. Two ventilators were placed in front of the buck to contain the post-test air bag exhaust.

Prior to each test, a series of measurements were taken to ensure consistent placement of the dummy. Before each test, the dummy's face was painted with chalk to indicate the points of contact between the dummy's face, arm, air bag and the test buck. The event was captured on high speed film and high speed video. Still pictures also were taken before and after each test.

Table 4 reflects the channels recorded in each test and the channel filter class at which the data was computed.

RESULTS

This section summarizes test results of the Hybrid III instrumented arm and the RAID comparison in the V-1 through V-4 configurations, the fleet air bag results in the V-1 and V-3 configurations and the J-95 repeatability test summary.

Table 3. Fleet Vehicle Test Matrix		
Vehicle	**Configuration V-1 (Conrad I @ 1.3 cm)**	**Configuration V-3 (Conrad V @ 1.3 cm)**
A-94	✔	✔
A-96	✔	✔
G-96	✔	✔
H-96	✔	✔
F-94	✔	✔
F-96	✔	✔
D-96	✔	✔
I-96	✔	✔
B-96	✔	✔
E-94	✔	✔
E-96	✔	✔



Figure 7. Setup

Table 4. Channel Classes	
Abbreviations & Channel Descriptions	**Channel Filter Class**
Mid-forearm x and y moment	600
Proximal forearm x and y moment	600
Distal forearm x, y, z accelerations	1000
Proximal forearm x, y, z accelerations	1000
Distal upper arm x, y, z accelerations	1000
Proximal upper x, y, z accelerations	1000
Head x, y, z accelerations	1000
Chest x, y, z accelerations	1000
Chest deflection	180
Neck forces in the x, y and z directions	1000
Neck moments in the x and y directions	600

There are two main events that occurred in the following tests which contributed to moments surpassing the IARV. The first was caused by direct contact between the air bag and the Hybrid III instrumented arm at the time the air bag was deployed. This event usually occurred within the first 15 msec of each test (referred to as an 'air bag injury'). The second event that contributed to moments that surpassed the IARV was caused by contact between the Hybrid III arm and body (referred to as a 'contact injury'). This usually occurred at times greater than 20 msec. Table 5 contains a summary of the RAID and Hybrid III instrumented arm data, followed by the times at which the data was recorded.

HYBRID III INSTRUMENTED ARM/RAID COMPARISON

CONFIGURATION V-1 - Films reviewed of configuration V-1 showed very little contact between the dummy's arm, face or chest for all air bag systems. In all tests, the forearm seemed to rotate away from the wheel while the upper arm stayed still until later in the event. As seen in Table 6, the injury summary shows moments exceeding the IARV occurred from air bag contact, while one isolated case actually occurred when the arm was thrown to the side causing and excessive inertial load from the elbow locking. All of the peak wrist accelerations occurred in the first 10 msec and were therefore results of the air bag deployment.

CONFIGURATION V-2 - Contact injuries were evident in two of the four tests in the V-2 position. In the J-95 and H-91 air bags, the arm contacted the chest and continued upward into the chin and neck. These moments caused by contact were not severe enough to surpass the IARV, but the forces afflicted by the air bag at deployment exceeded the proposed female IARV (as indicated in the first two columns of Table 6). The L-92 and K-93 air bags on the other hand, caused the arm to travel in an upward direction directly contacting the chin. These moments surpassed the proposed male and female IARV.

CONFIGURATION V-3 - Configuration V-3 had the same hand, elbow and steering wheel position as configuration V-2. However, in configuration V-2 the forearm was placed 7.6 cm from the module and in configuration V-3, the forearm was placed at a distance of 1.3 cm from the module. This decrease in distance subjected the arm to some higher forces when the air bag was deployed.

Unlike configuration V-2, where the Hybrid III arm contacted the chest of the dummy in each case, chest contact was only evident in one of the vehicles (K-93). The other three vehicles (J-95, L-92 and H-91) experienced direct contact between the Hybrid III arm and the head and neck (without contacting the chest). This seemed to produce higher moments than the cases in which the arm slid up the chest.

The resultant moments for the V-3 position were a little more severe than the results from the V-2 position. The proximal placement to the point of air bag deployment increased the magnitude of the moments. As seen in Table 6, configuration V-2 registered two moments that exceeded the proposed female injury criteria. On the other hand, configuration V-3 registered one moment that exceeded the proposed male IARV and two that exceeded both the male and female proposed IARV.

CONFIGURATION V-4 - The K-93 and the J-95 air bags were the only bags tested in the V-4 position due to nonessential findings at Conrad Technology compared to the other configurations. Contact between the arm and the dummy was evident in each test, although the highest injury potential moment was caused by the air bag (as indicated by the second and third columns of Table 6). The K-93 forced arm contact high on the chest, then upward to the face. The J-95 caused direct contact between the hand and the face of the dummy resulting in a moment which exceeded the proposed male IARV.

The resultant moments registered portray the more mild characteristics of this configuration compared to the other three positions. The K-93 arm test obtained the most force from the air bag and not from contact in this position. The J-95 results were a little more severe than they were in the V-1 and V-2 positions simply because there was contact with the J-95 bag in this configuration and not in the others.

HYBRID III INSTRUMENTED ARM/RAID DISCUSSION

Table 6 reflects the peak resultant moments (N*m) relative to the proposed IARV, and the events that caused the peak moment. Values are from the ARM/RAID comparison testing only. A ✔ represents a moment that would be equivalent to the IARV for a fifth percentile female (70-90 N*m), an ✗ represents a moments that would be equivalent to the 50th male IARV (120-150 N*m) and a ✚ represents any moment that exceeds 150 N*m. The first column represents air bag injuries and the second column represents contact injuries. The third column represents moments created when the arm is projected horizontally to the side of the dummy and the elbow is in a locked position. The location of the maximum moment as recorded at the proximal or distal moment strain gage locations (Figure 2) is also noted.

The Hybrid III instrumented arm/RAID peak resultants for configuration V-1 through V-4 are listed in Table 5. For comparative purposes, only the first 20 msec of the Hybrid III arm data is considered here. This eliminated responses occurring due to dummy contact and thus provided a better comparison with the RAID.

For each of the following configurations some trends between the Hybrid III instrumented arm and the RAID data are evident. The peak resultant moments for the RAID configurations measured much higher and surpassed the IARV for both men and women in every case. On the other hand, the Hybrid III moment measurements surpassed the female IARV in each case but less than half the time for male IARV. Since women seem to be the subject of arm injury more often then men, the Hybrid III injury assessment may be more representational before any scaling factors are applied to either device.

The difference in wrist velocities can be accounted for by analyzing the differences in the two testing devices. The RAID velocities were recorded at values almost twice as high as Hybrid III instrumented arm values. This is most likely due to the difference in the weight, kinematics and set up of the two different test devices. As the air bag deployed, the Hybrid III forearm generally moved while the upper arm remained stationary. This allows us to contribute higher accelerations in the Hybrid III testing to the mass of the lower arm being less

Table 5. Air Bag Assessment

		ARM				RAID			
		K-93	L-92	J-95	H-91	K-93	L-92	J-95	H-91
Configuration V-1 Nominal 1.3 cm distance Wheel rotated 90° CCW (seam usually vertical) Arm from 11 to 4 o'clock	Peak resultant wrist acceleration (g)(ms)	512 @5.2	184 @6.4	200 @6.0	544 @6.6	135	71	55	116
	Peak Resultant moment (Nm) (ms)	135.6 @5.7	78.1 @8.6	83.5 @8.2	106 @15.1	483	313	213	436
	Wrist Velocity at 15 ms (km/h)	76.6	21.9	30.4	77.2	98	50	32	77
	Air bag ranking	4	1	2	3	4	2	1	3
Configuration V-2 Nominal 7.6 cm distance Wheel rotated 30° CCW (arm perpendicular to seam) Arm from 10 to 4 o'clock	Peak resultant wrist acceleration (g)(ms)	660 @8.2	138 @8.6	301 @6.8	NO DATA		39	72	106
	Peak resultant moment (Nm) (ms)	181.3 @8.5	65.1 @9.4	80.5 @8.1	107.7 @9.6	370	121	273	328
	Wrist Velocity at 15 ms (km/h)	44.7	11.2	26.4	NO DATA	109	60	43	62
	Air bag ranking	4	1	2	3	4	1	2	3
Configuration V-3 Nominal 1.3 cm distance Wheel rotated 30° CCW (arm perpendicular to seam) Arm from 10 to 4 o'clock	Peak resultant wrist acceleration (g)(ms)	518 @8.3	330.3 @5.8	235.3 @7.4	497 @8.2	137	57	64	183
	Peak Resultant moment (Nm) (ms)	238.7 @8.9	117 @6.4	144.3 @9.1	233.5 @8.6	522	290	343	617
	Wrist Velocity at 15 ms (km/h)	61.7	29.8	34.4	49.8	111	76	79	97
	Air bag ranking	4	1	2	3	4	1	2	3
Configuration V-4 Nominal 1.3 cm distance Wheel rotated 120° CCW (arm parallel to seam) Arm from 10 to 4 o'clock	Peak resultant wrist acceleration (g)(ms)	769 @5.0		278 @7.5		133	55	63	183
	Peak resultant moment (Nm) (ms)	197.6 @6.5		103.3 @9.4		525	280	340	620
	Wrist Velocity at 15 ms (km/h)	83.6		44.8		118	77	79	98
	Air bag ranking	2		1		3	1	2	4

*Injury criteria for peak moments are 120-150 N*m for men and 70-90 N*m for women.

Preliminary Results from Arm/RAID Comparison Testing

Table 6. Hybrid III Injury Summary						
Air Bag & Configuration	Injury From Air Bag		Injury From Contact		Injury From Elbow Locking	
	Proximal	Distal	Proximal	Distal	Proximal	Distal
K-93 V1					✚	✗
L-92 V1	✔					
H-91 V1		✔	✔			
J-95 V1	✔					
K-93 V2			✚			
L-92 V2			✚			
H-91 V2	✔					
J-95 V2	✔					
K-93 V3	✚					
L-92 V3			✗	✔		
H-91 V3	✚			✔		
J-95 V3	✗					
K-93 V4	✚					
J-95 V4			✗			
J-95 V11	✔					
J-95 V12			✗			
J-95 V13	✔					
J-95 V14	✔					

than that of the RAID. This difference in weight would produce the different accelerations experienced by each of the test devices. Another source of discrepancy may result from the time at which the velocity values were recorded. The RAID velocities were taken from the maximum velocity reached by the device after losing contact with the air bag (generally at 30 to 40 msec). On the other hand, the Hybrid III maximum velocities were recorded at an earlier time because of the contact that occurred between the arm and the dummy's head and body (at 15 msec).

The difference in results can also be attributed to different kinematics between the two systems. The Hybrid III arm cannot move freely in any direction because its motion is dependant on the movement of the rest of the dummy. If the arm is moving, the elbow and upper arm and body will restrict some of its motion while also moving backwards to absorb some of the initial force from the air bag. On the other hand the RAID is attached to an immobile support so the arm reacts to all of the force applied to it.

Despite the difference in kinematics some parallel conclusions do exist in the air bag assessment rating between the Hybrid III instrumented arm and the RAID. The RAID and the Hybrid III arm device ranked configurations V-2 and V-3 exactly the same while V-4 was able to differentiate between the most and least aggressive bags. For configuration V-1 the air bag ratings were very similar. The L-92 and the J-95 bags were found to be the least aggressive bags while the K-93 and the H-91 were found to be the most aggressive bags. The evaluation methods used to assess the air bags used with the Hybrid III arm were not stringent. The Hybrid III arm tests were ranked by the magnitude of the resultant moments, accelerations and velocities. For example, the bag with the highest resultant moment was

ranked #4, or most aggressive. The bag with the lowest moments and accelerations was ranked #1, or least aggressive. There is some ambiguity to the ranking system, and other parameters could slightly sway the ranking of the bags. Nevertheless, the two systems were able to accurately differentiate between the most and least aggressive systems.

FLEET AIR BAG RESULTS

The fleet air bag testing was conducted on 1994 and 1996 systems. For three of the systems, tests were conduced on both 1994 and 1996 models, while the remaining units were tested only for the 1996 model. The 1994 and 1996 model year air bag module performances were compared to assess improvements. In the charts which follow, the injury reference ranges for the male (Male IC, M-Max IC) and female (Female IC, F-Max IC) IARV are indicated by the straight horizontal lines.

FLEET VEHICLES IN POSITION V-1 - The charts that follow show the peak resultant moments for the fleet vehicles in the V-1 position. Figure 8 depicts the improvement between the A, F and E, 1994 and 1996 air bags. Each vehicle's aggressiveness improved in the more recent model year. The 1994 models all surpassed the peak male IARV while the 1996 model year was decreased. The E vehicle's moments had the most dramatic reduction form 1994 to 1996, but still exceeded the female IARV.

The other vehicle performances ranged from very aggressive (exceeding the peak male criteria) to non-aggressive (not exceeding the lowest female IARV). As seen in Figure 9, the G-96 and the H-96 both surpassed the IARV substantially while the I-96 and the D-96 fell within the maximum male IARV. The B-96 moments created during the test did not exceed either the male or female IARV.

FLEET VEHICLES IN THE V-3 POSITION - The following results show the Fleet air bags in the V-3 position. As illustrated by the bar graphs (Figures 10 & 11), this is a much more severe position than the V-1 position. Figure 10 indicates a significant improvement for the 1994 model year air bags. The 1994 model year air bags all exceed the peak male IARV. The 1996 model year air bags for these three models all registered slightly within the male injury range or midway between the maximum and minimum female IARV. The remaining 1996 air bag systems all exceeded both the male and female IARV as shown in Figure 11.

REPEATABILITY

CONFIGURATION V-1 - The following tests were done to demonstrate consistency in the test results for a given test configuration. Four tests were run using the same air bags, configuration and set up. The dummy was placed in the V-1 position, with its arm in the 11 o'clock and four o'clock position. The steering wheel was rotated 90° CCW. Table 7 describes the events seen on film.

Table 8 shows the magnitude of the resultant accelerations for the four J-95 repeatability tests. The first three tests were performed with the thumb placed under the steering wheel. In order to check the effect of the thumb position on the wheel, test #4 was conducted with the dummy's thumb placed above the wheel. There was no significant difference in the data to indicate that the thumb affected the test results. Test #2 registered higher in most of the accelerations and the lower arm elbow resultant moment because the dummy's arm contacted the dummy's face during the test. In test #1, 3 and 4, the arm was thrown from the bag and the hand was very close to experiencing contact with the dummy's face and head. Therefore, any slight variation in the air bag or test set up could have caused the contact and was not considered in the repeatability aspect of the testing. For comparison, the point of contact was filtered out of the data in order to obtain a better comparison of the data. The following table reflects the moments with the contact point filtered out for Test #2, but the graphs of the moments do not. The point of contact was left in to show the magnitude and time at which it occurred.

The following graphs (Figure 12) show that the tests were very consistent with the exception of the moment created by contact in the second test. When the hand brushed the chin creating a moment at about 30 msec, an increase in the velocity can be noted on the graph. If this event were omitted from the analysis, the graphs of the distal and proximal moments and the velocity curves track each other very well. The distal moments all peak just before 10 msec and are all going back to zero by 20 msec. The proximal moments also peak before ten msec and return to zero at the same times.

DISCUSSION & CONCLUSIONS

The Hybrid III instrumented arm and the RAID were fairly consistent in differentiating between the severity of the different air bag systems. The RAID and the Hybrid III moments and velocities registered at different magnitudes, but each measured the same bags as having the highest moments at given velocities relative to the other systems. Both devices recorded the K-93 and the H-91 as the bags with the highest peak moments and the J-95 and the L-92 with the lowest peak moments. A moment vs. velocity scatter graph was used to portray the severity of each air bag system as measured by the Hybrid III (Figure 13) and RAID (Figure 14) devices. Each air bag model and year labels are preceded by its configuration designation (for example, J-95 in the V1 position is designated V1-J-95). The K-93 and H-91 were found to be the most severe air bags in the V-3 position, since they measured both the highest moments and velocities. In addition, the graphs also show that each device was able to correlate the severity of the moment with the velocity. Both graphs show a somewhat linear relationship between an increase in velocity and an increase in the moment recorded.

Both the Hybrid III and RAID devices recorded moments and accelerations resulting from deployment of the air bag systems. The Hybrid III was also able to detect moments created when the arm contacted the dummy's head, neck or chest. The RAID could assess severity of the bag based solely on the moments and accelerations applied by the air bag and could not account for any injury potential resulting from contact.

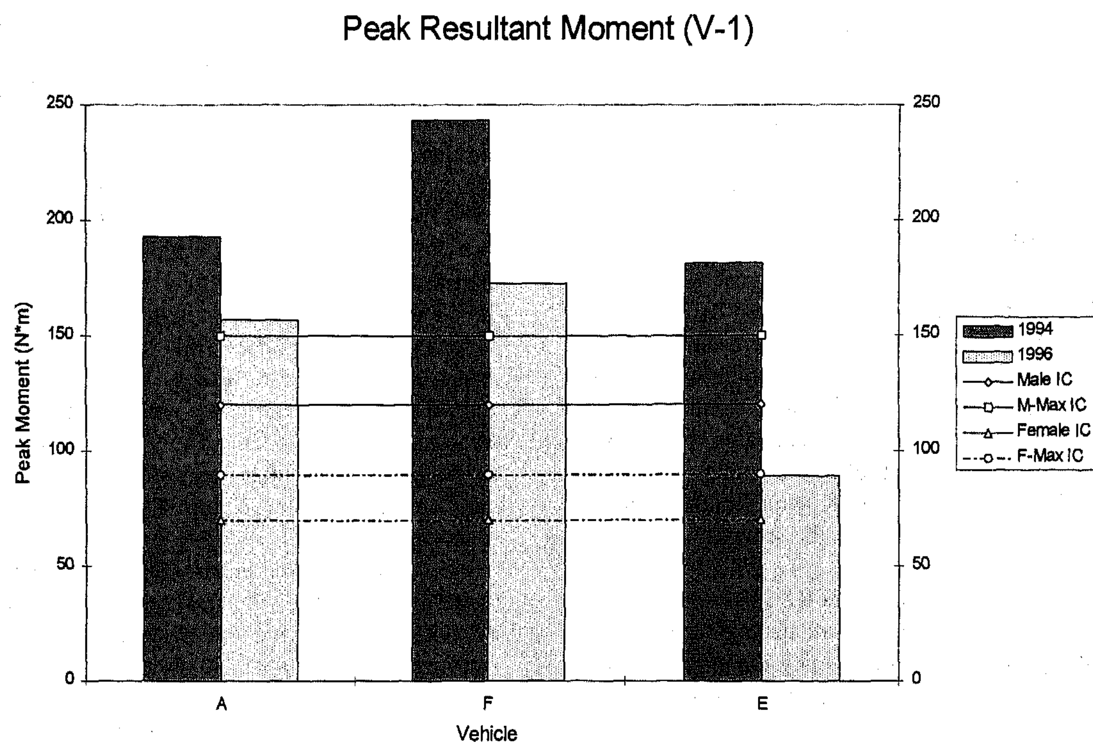


Figure 8. 1994, 1996 fleet air bag comparison in the V-1 configuration

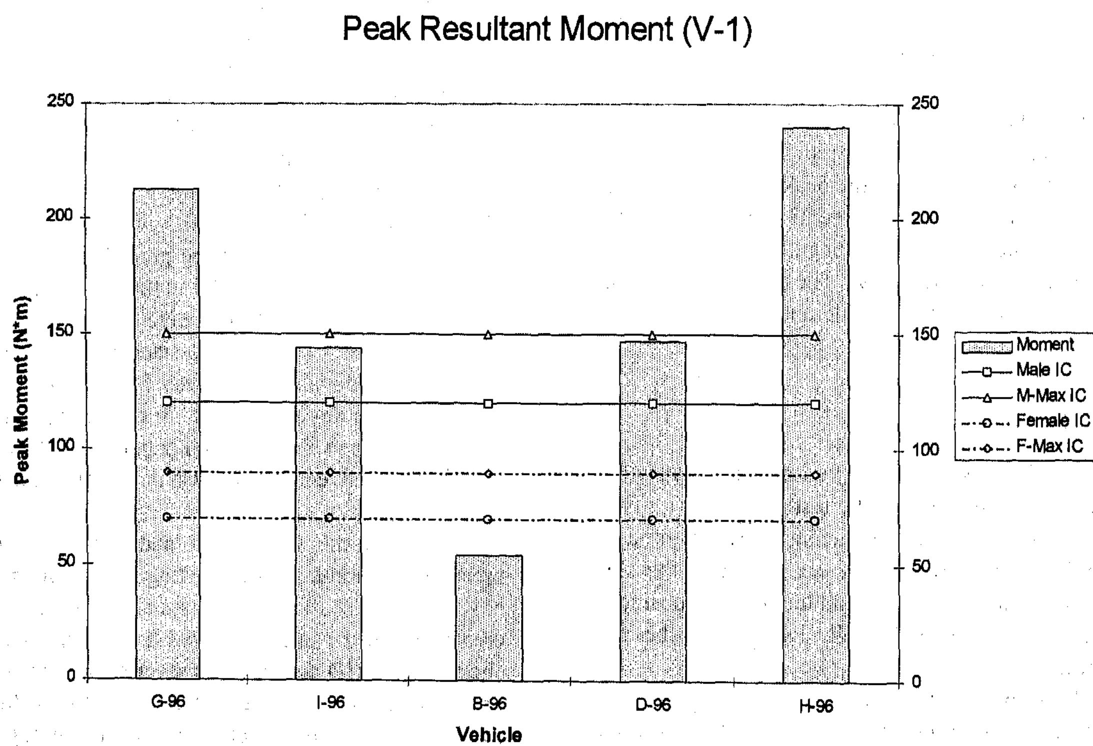


Figure 9. 1996 fleet air bag in the V-1 configuration

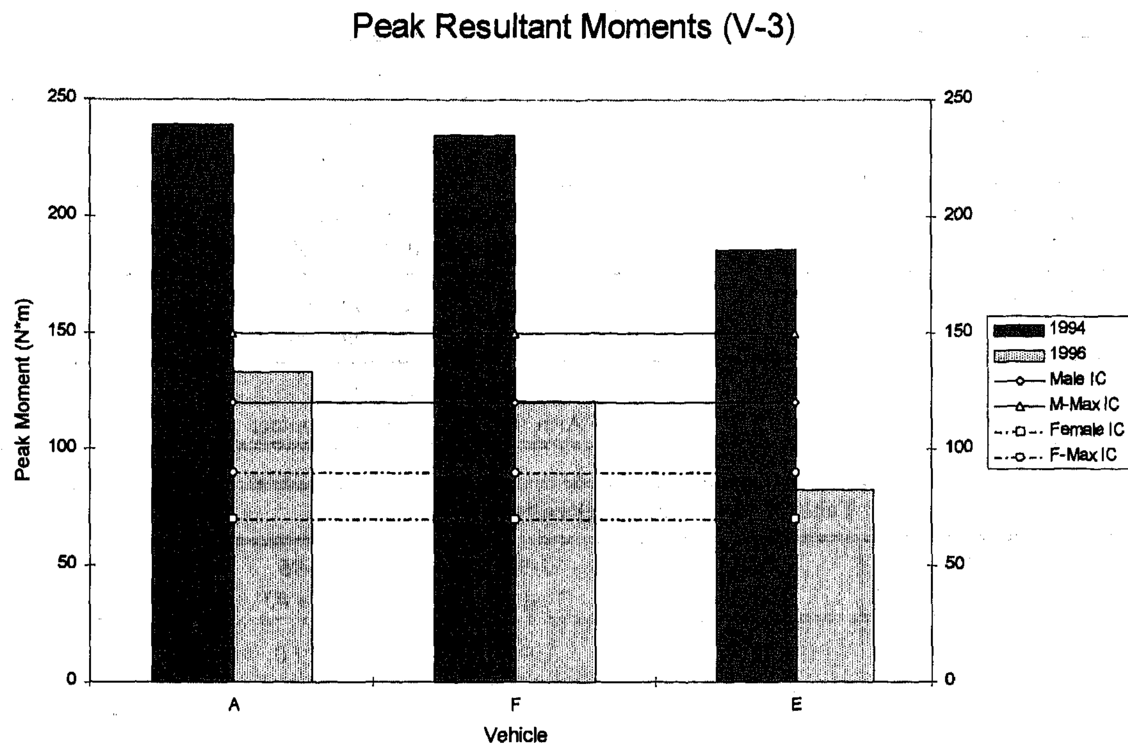


Figure 10. 1994, 1996 fleet air bag comparison in the V-3 configuration

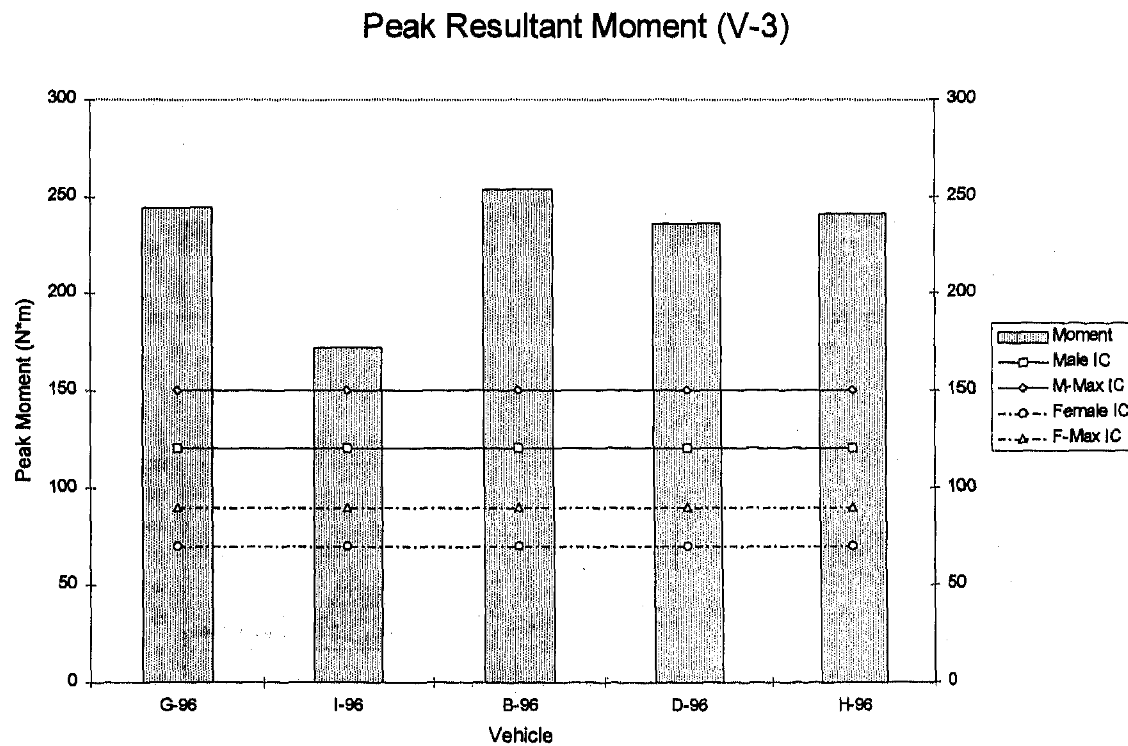


Figure 11. 1996 fleet air bag in the V-3 configuration

Table 7. Film Review (Repeatability)	
Vehicle	**Film Comments**
J-95 Test #1	Thumb under wheel Arm thrown to side, no contact
J-95 Test #2	Thumb under wheel Wrist brushes chin then goes off to side
J-95 Test #3	Thumb under wheel Arm thrown to side, no contact
J-95 Test #4*	Thumb over wheel Arm thrown to side, no contact

Table 8. Resultant Accelerations (Repeatability)				
Channel/units/ description	**J-95 Test #1**	**J-95 Test #2**	**J-95 Test #3**	**J-95 Test #4***
ARLRGL (g, msec) Wrist resultant	200.1 6.0	219.5 5.8	222.0 5.3	251.1 7.3
ARRML (Nm, msec) Lower arm midshaft resultant moment	23.2 8.3	21.9 10.3	33.0 8.5	21.2 9.3
ARRMU (Nm, msec) Lower arm elbow resultant moment	83.5 8.2	107.5 10.3	103.6 8.6	74.9 9.3
VELOCITY @ 15 ms (m/msec)	27.1	35.7	27.4	27.7

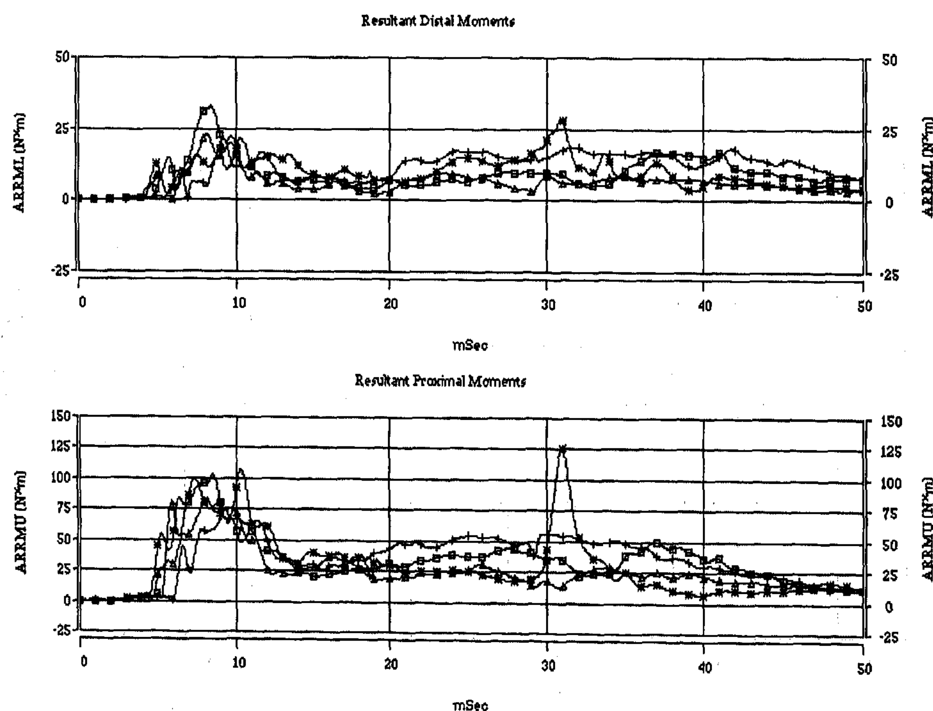


Figure 12. Proximal and distal moment curves for J-96 repeatability

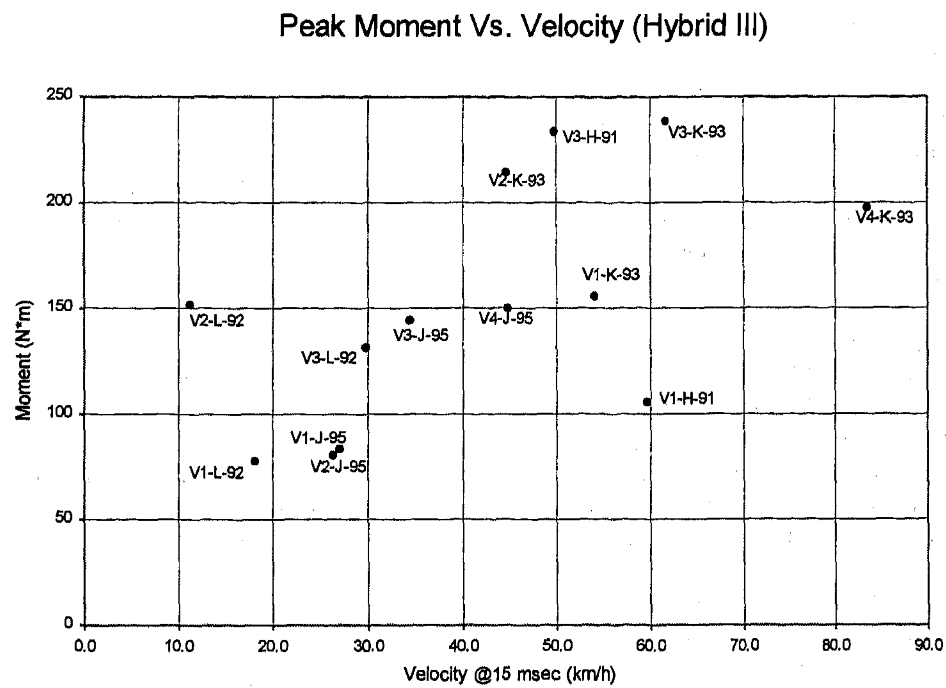


Figure 13. Hybrid III Moment Summary

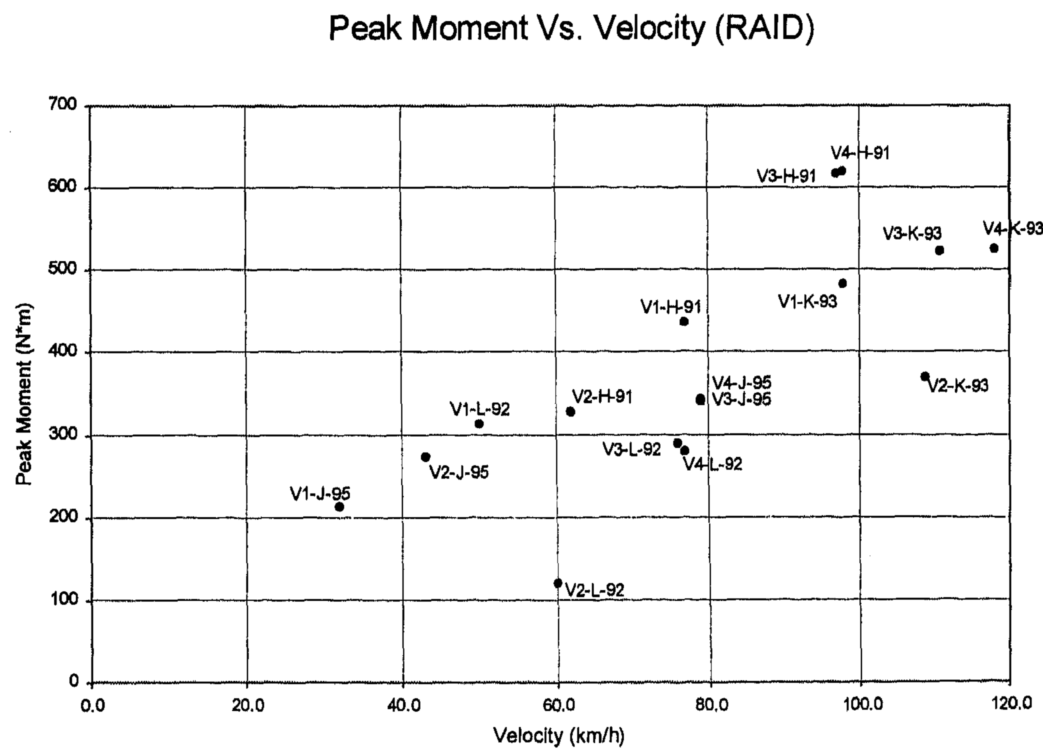


Figure 14. RAID moment summary

The RAID and the Hybrid III did not share the same kinematics, therefore, the moment and velocity magnitudes differed from each other. The RAID was an aluminum structure which made it somewhat lighter (2.1 Kg) than the Hybrid III arm (4.5 Kg), and connected at a joint acting as the elbow. The Hybrid III was heavier and was able to move at the elbow and the shoulder. The Hybrid III movement was also governed by the dummy which played a role in how the arm moved and inevitably created different moments and velocities than the RAID.

The Hybrid III moment measurements were the same order of magnitude as the proposed IARV. The Hybrid III testing mostly exceeded the female injury IARV of 70-90 N*m, but not always the male IARV of 120-150 N*m. The RAID moments registered rather high, and would have implied a high likelihood of forearm fracture in all systems for both male and females, suggesting that the arm IARV's may not be applicable to the RAID.

Based upon these results, the Hybrid III instrumented arm was selected for testing of fleet air bag systems. While both the RAID and Hybrid III devices appeared capable of distinguishing air bag deployment severities, the Hybrid III device offered two advantages over the RAID. First, the Hybrid III device provided the potential for assessment of arm interaction with the occupant and vehicle interior which was more realistic than the RAID. Secondly, the moment measured with the Hybrid III was of generally the same magnitude as the proposed arm IARV. This was felt to be more consistent with the available accident data.

ACKNOWLEDGMENTS

Our recognition goes to Ken Welty and Rob Jacobs for test set up, Herman Jooss for photography, and Gerda England for AutoCAD drawings. Our thanks also extends to Conrad Technologies Inc., for their cooperation and communication with us throughout the initial stage of this program.

REFERENCES

1 Conrad Technologies Inc., Personal Communications and Internal Correspondences, 1996.

2. Saul, R. A., et al., "Hybrid III Dummy Instrumentation and Assessment of Arm Injuries During Air Bag Deployment," SAE Paper No. 962417, Fortieth Stapp Car Crash Conference, 1996

3. Huelke, D. F., et al.,"Upper Extremity Injuries Related to Airbag Deployments," The Journal of Trauma: Injury Infection, and Critical Care, Vol. 38, No. 4, Pages 482-488, Williams & Wilkins, 1995

4. Kuppa, S. M., et al., "RAID-An Investigative Tool to Study Air Bag/ Upper Extremity Interactions," To be presented at the SAE International Congress and Exposition, February 24-27, 1997

970404

Transient Aerodynamic Simulation in Crosswind and Passing an Automobile

Kenji Okumura and Toshihiko Kuriyama
Daihatsu Motor Co., Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The one-box type automobile's stability on the highway is often influenced by encountering crosswinds or when passing large automobiles such as trucks and buses. This causes the automobile to behave unexpectedly. Many experiments for improving this situation have been carried out. In this respect, the analysis of transient aerodynamic characteristics is important for automobile safety and stability on the highway. Conventional transient aerodynamic simulations require a supercomputer and about million grid points. Also there were few case studies that dealt with situations such as plunging into crosswind and passing an automobile.

In this paper, a transient aerodynamic simulation by using a sliding mesh of discontinuous interface and the Arbitrary Lagrangian-Eulerian (ALE) method is presented. This method uses the k-ε turbulence model and the third-order upwind scheme, which is introduced to the convective term of the Navier-Stokes equation to improve the flow field and pressure distribution. It takes 100 hours to get the total transient results by using an engineering work station (EWS). Calculated yaw characteristics in a crosswind are in good agreement with experimental results.

Furthermore, we have simulated the effect of an active spoiler in a crosswind, and the aerodynamic characteristic of passing a one-box automobile are carried out.

INTRODUCTION

The RV (Recreational Vehicle), which was created to meet contemporary user needs, is now establishing itself as a new category of automobile. The wagon-shaped one-box automobile, for example, is one type of RV. This type of automobile, however, is said to be aerodynamically sensitive when hit by a gust of wind upon emerging from a tunnel at high speed, when overtaking a large automobile or when passing a large on-coming automobile. One of the reasons for this sensitivity, according to many previously published experiment reports[(1)-(5)] is the yawing movement acting on the one-box automobile. In our previous paper[(6)], we carried out a crosswind analysis method within a steady flow and reported on a wind-tunnel experiment using a crosswind (active) spoiler, the verification of the experiment and the spoiler's effect in the crosswind. This steady flow analysis, however, is not capable of grasping the actual phenomenon and is therefore insufficient for the analysis of the behavior and dynamics of an automobile driven on the road.

In this paper we will discuss a new method that we have developed for predicting the transient aerodynamic characteristics of an automobile, in which transient aerodynamic phenomena are calculated using a discontinuous interface grid and moving boundaries. Our simulations of yawing moment variations, which influence the high-speed driving stability of an automobile, were in good tendency with the results of our experiments. We have also verified the effect of a crosswind spoiler. In this paper we will also report on the simulations of one automobile passing another and the analysis of the influence of the transient aerodynamic characteristics of two automobiles.

COMPUTATIONAL METHOD

For computation, we used a commercial fluid analysis solver, SCRYU V1.7,and in order to calculate the k-ε turbulence model, separation and reattachment with high accuracy, we used the third-order upwind scheme, MUSCL, for the convection term of the discretization method of the Navier-Stokes equation. This solver applied the first-order Euler-backward scheme for the temporal term to get implicit solution. All variables are colocated. Solution algorithm is SIMPLE or SIMPLEC. In our previous crosswind steady analysis method[(6)] we sought to compute the steady aerodynamic characteristics of a constant direction flow such as that inside a wind tunnel, but the real phenomena encountered during driving, such as a gust of wind or unexpected passing by another automobile, are unsteady.

Therefore, in order to predict transient aerodynamic phenomena, we have developed a computational method using a

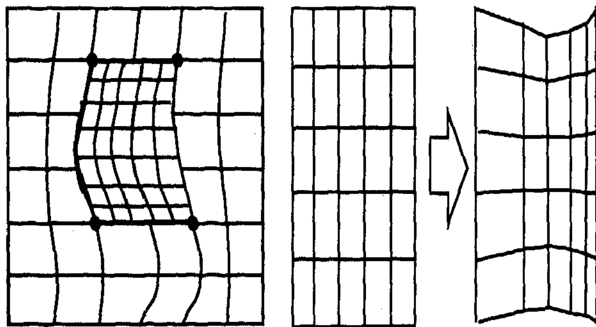


Fig. 1. Concept of Dis-continuous Interface Grid

Fig. 2. Concept of ALE Method

$$\frac{\partial (u_i - v_i)}{\partial x_i} = 0$$

Equation 1. Mass-Conservation Equation

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial (u_j - v_j)\rho u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\mu\frac{\partial u_i}{\partial x_j} - \rho g_i \beta (T - T_0)$$

Equation 2. Momentum-Conservation Equation

discontinuous interface grid and the ALE (Arbitrary Lagrangian-Eulerian) method[7], which is capable of expressing moving boundaries.

Fig. 1 shows the concept of a discontinuous interface grid. In this method, as long as the edge points of a discontinuous surface match, the mass and momentum of the discontinuous surface are conserved by interpolation. The overlaped or separational ratio of this method allows in a half of element size. By establishing groups of elements that are movable relative to each other across this discontinuous surface, moving meshes can be rendered.

Equation 1 is a mass-conservation equation and equation 2 is a momentum-conservation equation according to the ALE method, which is capable of expressing moving boundaries. These equations are different from those used in the static coordinate system in that moving velocity components Vi,j are introduced. By treating the energy-conservation equation and the k-ε equation in the same manner, the movement and deformation of the meshes can be handled.

This concept is shown in Fig. 2. The trouble with this method is that if the areas of the interfaces at a moving boundary differ greatly, the resolution of the larger surface influences the computational accuracy, resulting in low interpolation accuracy. In order to avoid this problem, in our computation we concentrated a computational mesh around the automobile, in the positions where a crosswind hits and the positions most influenced by passing automobiles, so as to minimize area variation of the interfaces during movement.

Using a computational method based on the above concept, we defined meshes that expand and shrink as they move on a discontinuous surface, thereby expressing the movement of the automobile.

COMPUTATIONAL CONDITIONS

The computational conditions used in the present analysis are shown in Table. 1. We used a 1:1 scale one-box model for verification. To simulate the automobile's behavior into a crosswind, we set the automobile velocity in the driving direction at 27.8 m/s (100 km/h) and the crosswind velocity at 13 m/sat right angle as in the case of the crosswind evaluation. For the Reynolds number we employed a value of 2 x 10^6 using the automobile height as the representative length, and for the initial computational condition we assigned a crosswind of 13 m/s over the entire area from the crosswind generator region to the outlet. We set the computation time unit ΔT as 5 x 10^{-4}. For the kinetic energy k we assigned a value of 1 x 10^{-4} m^2/s^2, and for the dissipation rate ε we assigned a value of 1 x 10^{-4} m^2/s^3 over the entire area for the stablity of computation. For the boundary conditions, we set an inlet velocity regulation of 13 m/s in the crosswind generator region as the inlet condition, and defined the entire area of the downstream side of the wind generator region as the outlet and provided a pressure regulation. For the automobile surface and the road surface we used the logarithmic law for the wall conditions, and to make the proper wall conditions, we generated an initial grid around the surface so that Y^+ of the boundary layer is between 100 and 500. All other boundary surfaces were made symmetrical and assigned a free sliding condition. In this computation, tire rotation is not taken into account. The third order upwind scheme, MUSCL, used in our computation features a stable computation performance in that when the computation overshoots and tends to diverge, it automatically switches to a first order scheme. The computation time when the number of elements was 250,000 was 400 seconds per step using SGI's INDIGO 2 SOLID IMPACT R10000 engineering workstation. It took 900 steps from the automobile's plunge into the crosswind to its recovery to a steady running condition, making the overall computation time approximately 100 hours.

In the case of passing, the running velocity of the preceding automobile and the passing automobile were set as 22.2 m/s (80 km/h) and 33.3 m/s (120km/h) respectively. For the boundary conditions, 22.2 m/s of inlet velocity regulation was assigned as the inlet condition at the upstream side, and a pres-

Table 1. Computational Conditions

	Crosswind	Passing
Turbulence model	Standard k- ε model	
Convective Term	3rd order upwind MUSCL	
Coordinate System	Discontinuous and ALE Mesh based on BFC (SCRYU)	
Calculational Method	Unsteady	
Computer	SGI Indigo2 SOLID IMPACT (R10000)	
Wall Function	Logarithmic Law	
Outlet Conditions	Pressure Boundary (zero dp/dx)	
Inlet Conditions	Automobile Velocity =27.8 m/s Crosswind = 13 m/s k, ε= $1x10^{-4}$	Preceding Automobile Velocity = 22.2 m/s Passing Automobile Velocity = 33.3 m/s
Element Number	$2.5x10^5$	$2.7x10^5$
CPU Time	100 hours	220 hours

sure regulation was assigned to the downstream side, which was defined as the outlet. To provide the initial conditions, we conducted a steady flow analysis at the starting position (2 times the automobile length behind the preceding automobile), and performed a transient moving computation after obtaining the steady flow field. For the moving conditions, an automobile velocity difference of 11.1 m/s was assigned to the passing automobile. Other conditions were the same as those in the case of a plunge into a crosswind.

With approximately 270,000 elements, it took about 220 hours to calculate a simulation of the passing automobile from the position two automobile lengths behind to the position four automobile lengths ahead of the overtaken automobile on the above-mentioned engineering workstation.

COMPUTATIONAL MODEL

In order to reduce the mesh generating time, a boundary fitting coordinate (BFC) is used for each block constituting the mesh. For the automobile body mesh, the half model computational mesh used in the steady flow analysis has been converted to a full model and partially used, thereby reducing the computation time. Fig. 3 shows the computational mesh for a crosswind plunge in this analysis, and Fig. 4 shows the concept of a computational mesh for passing. The mesh for the crosswind plunge consists of four basic blocks: acrosswind generator block and a running automobile block, where the mesh forms are maintained, and an expandable block and a shrinkable block, where the meshes are deformed using the ALE method. These blocks are discontinuously connected along the interface of the gap between the crosswind generator and the automobile, and express the movement of the automobile as the blocks corresponding to the automobile itself and its front and rear slide while shrinking and expanding. Similarly, the computational mesh for passing also comprises four basic blocks, in which the preceding automobile's meshes are maintained as they are and the overtaking automobile's meshes deform as they move.

We assigned a space of 53 m x 26 m x 5.5 m (Length x Width x Height) for the analysis of the crosswind plunge. As the automobile moves ahead a wake is generated in the rear, and a separation vortex is generated on the leeward portion of the automobile's sides in response to this wake and the crosswind. The above dimensions are to ensure a sufficient space for capturing the separation vortex. The dimensions of the space with reference to the automobile dimensions are 3 times the automobile length for the spatial area from the automobile front to the crosswind generator region, 9 times the automobile length for the crosswind generator region itself, 3 times the automobile length for the space behind the automobile, 1.8 times the automobile width for the distance between the automobile side to the crosswind generator, 15 times the automobile width for the distance between the leeward portion of the automobile's sides and the outlet downstream, and 3.5 times the automobile height for the height of the space. In order to ensure accuracy, the computational meshes were concentrated on the entrance to the crosswind generator region and around the automobile itself. Fig. 5 shows the mesh around the automobile used in this analysis. The number of elements was approximately 250,000. Using the commercial pre-processor, PROSNER, for generating and correcting meshes while applying the mesh data for aerodynamic analysis of the

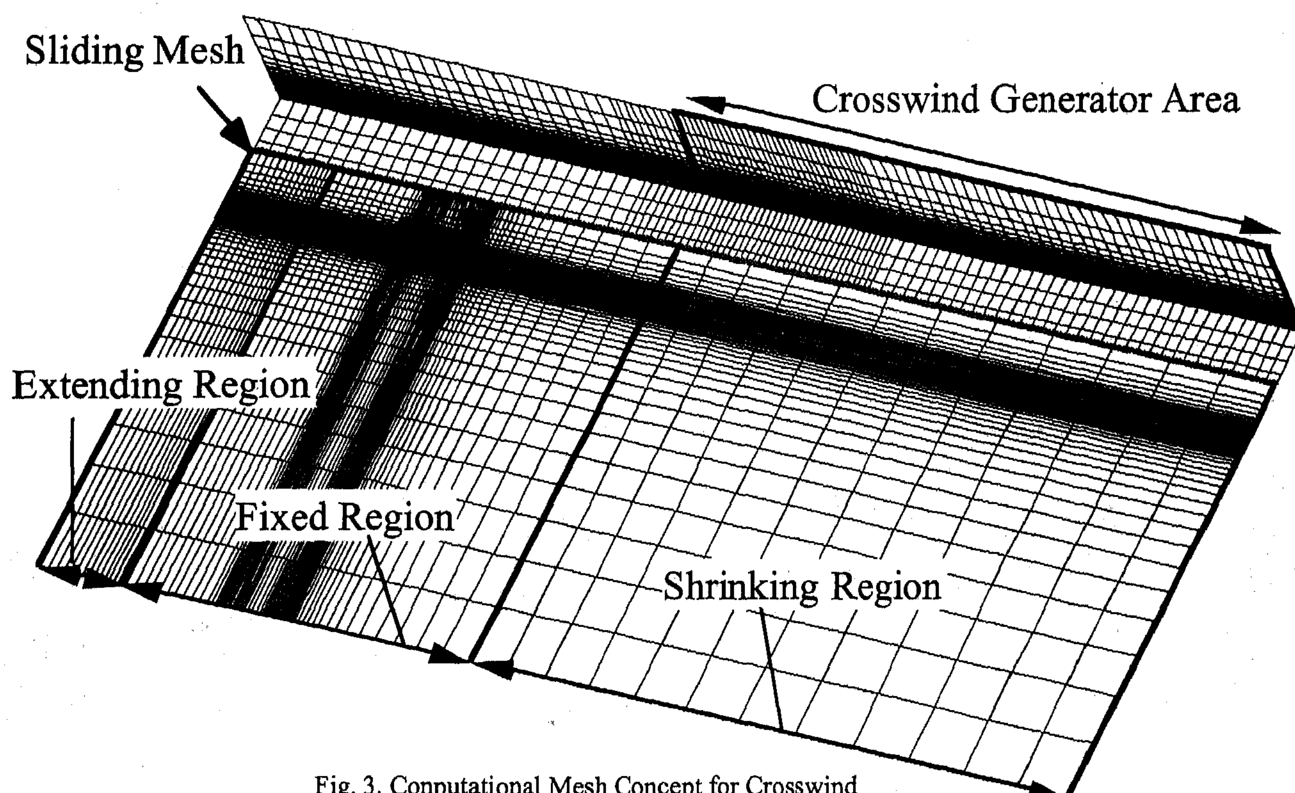


Fig. 3. Conputational Mesh Concept for Crosswind

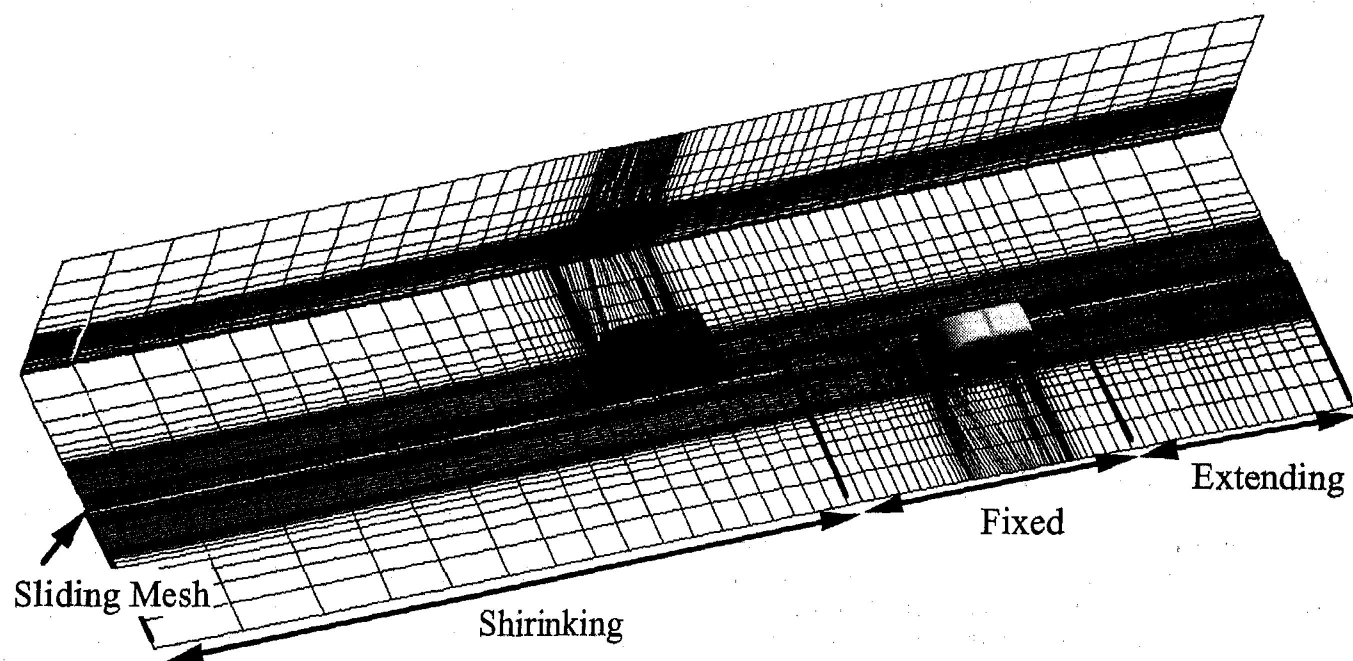


Fig. 4. Conputational Mesh Concept for Passing

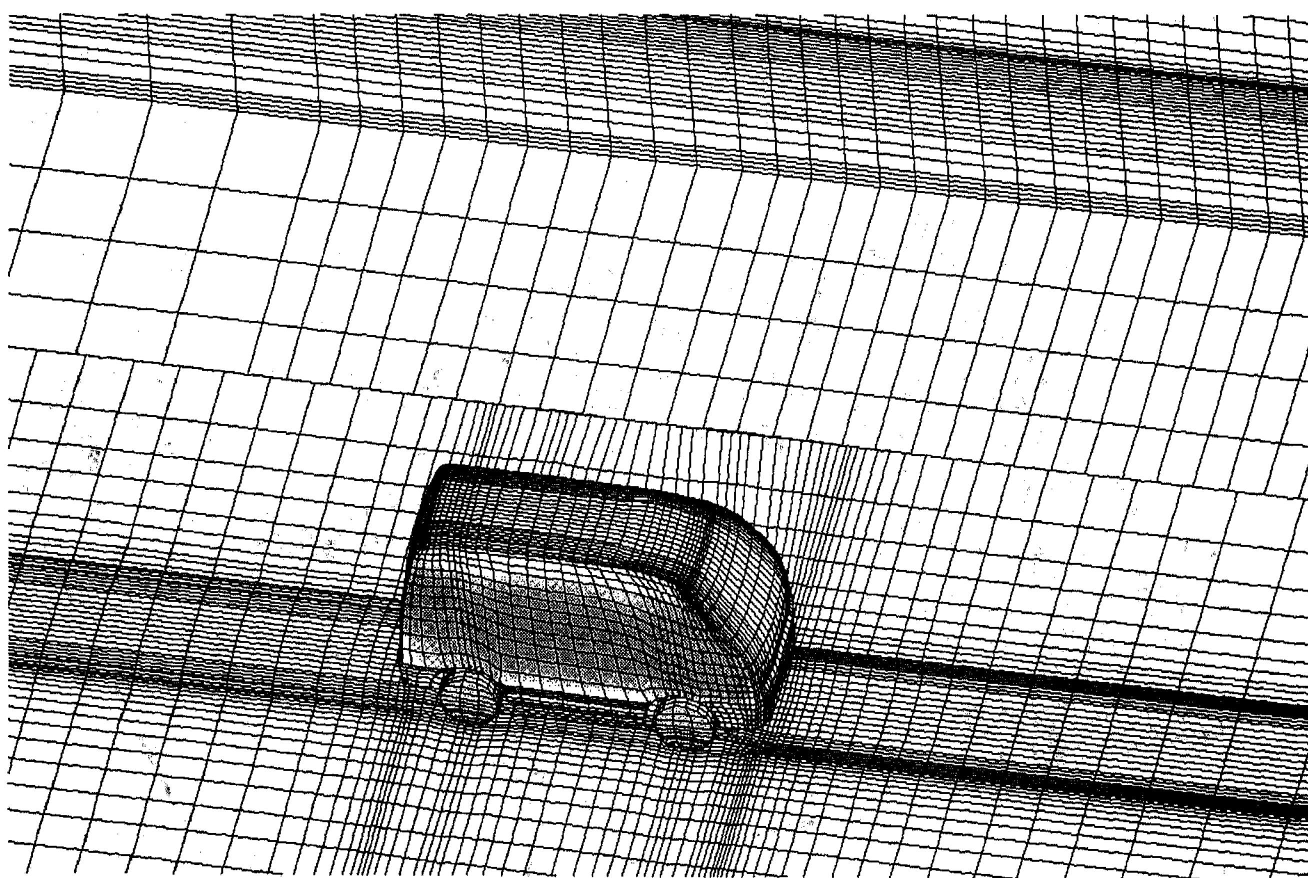


Fig. 5. Mesh around Body for Crosswind

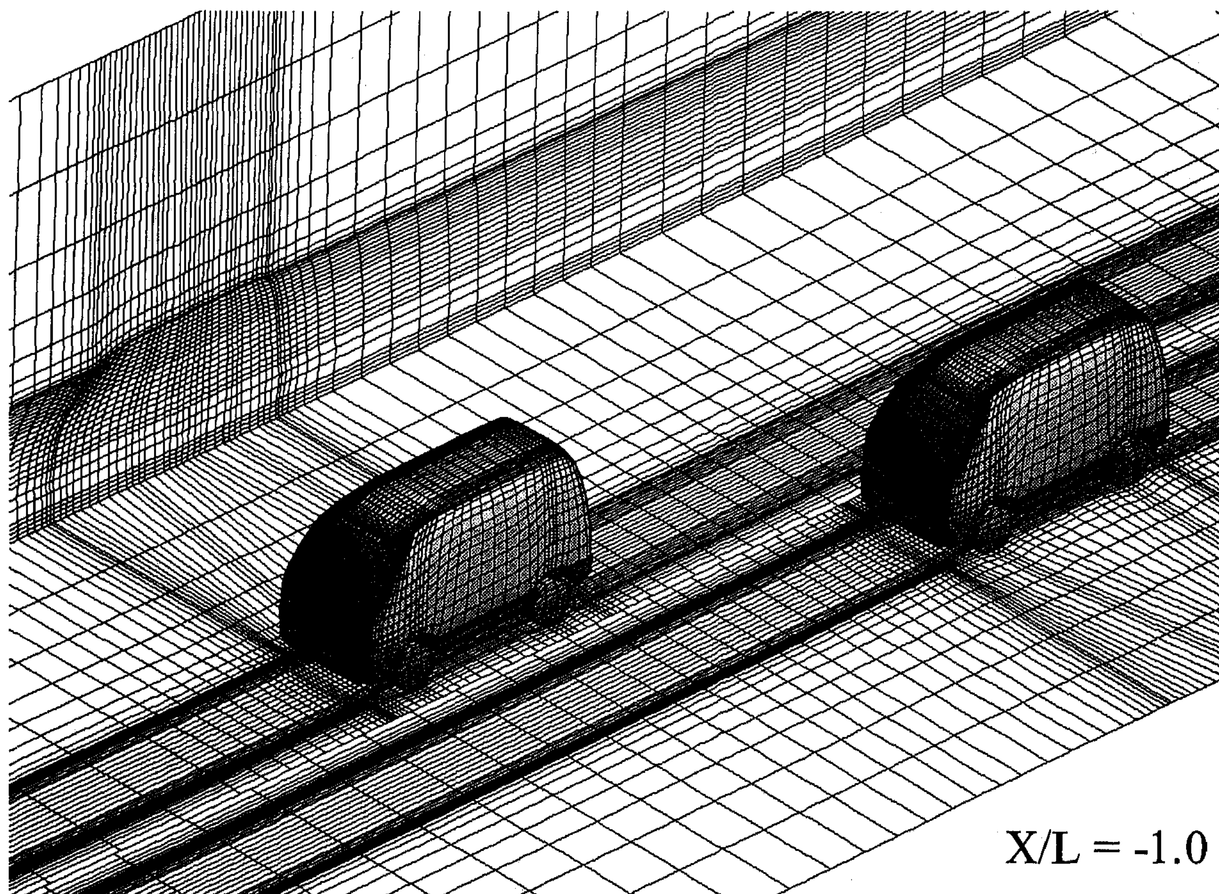


Fig. 6. Mesh around Body for Passing

steady flow, it took approximately five hours to complete the computation. Computation transferred from CAD shape data would have taken approximately 25 hours.

The analytical space for the computational mesh for passing measures 40 m x10 m x 5.5 m (Length x Width x Height). The automobiles are presumed to be of the same one-box type with a distance of 1/2 width between them. The dimensions of the space with reference to the automobile dimensions are: 5 times theautomobile length for the spatial area from the preceding automobile to the inlet, twice the automobile length for the automobile-to-automobile distance, 3 times the automobile length for the space between the passing automobile and the outlet, 2.5 times the automobile width for the exterior space on either side of the automobile, and 3.5 times the automobile height for the height of the space. Fig. 6 shows the mesh around the automobile. The number of elements is about 270,000. It took about the same time to generate the mesh for the plunge into the crosswind.

UNSTEADY COMPUTATION RESULTS FOR THE PLUNGE INTO A CROSSWIND

The yawing moment that acts on the one-box automobile when it enters into a crosswind reaches its maximum value immediately after it enters the crosswind and then transiently changes to a steady value as shown in Fig. 7. When the relative angle of the composite wind is 30 degrees, the yawing moment reaches its maximum value when the automobile has moved 1.3 times the automobile length into the crosswind, regardless of the automobile shape, according to the experimental study by Yamada, et al[1].

In our study, we conducted a simulation of a plunge into a crosswind in an attempt to discover the mechanism of yawing-moment generation through visualization. We used a commercial fluid post-processor, FIELDVIEW in order to shorten the post-processing time.

YAWING MOMENT COEFFICIENT (C_{YM}) - As shown in Fig. 8, the automobile is divided into four blocks that border each other at the center of theautomobile. The results of obtaining the yawing moment generated in each block are shown in Fig. 9. In order to average out the transient data, an ensemble average is taken from ten serial steps.

Because of the diffusion of the flow from the wind generator region, the automobile begins to come under the influence of the crosswind when it has moved approximately one automobile length from the entrance to the wind generator region. When this position is set as X/L=0 in figures, the yawing moment reaches its maximum value when the automobile has moved approximately 1.4 times the automobile length into the

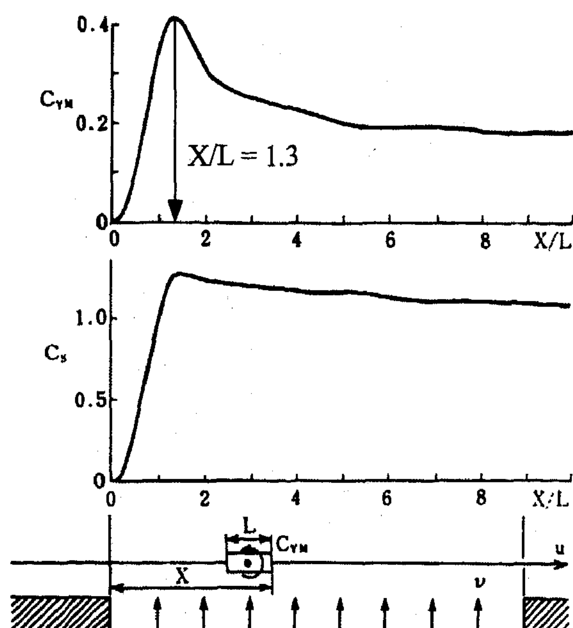


Fig. 7. Variations of Experimental CYM and Cs plunge into a Crosswind (quated from Kobayashi's SAE paper 881878)

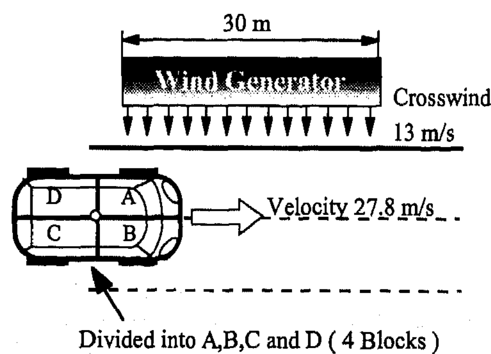


Fig. 8. Automobile devided into 4 Blocks

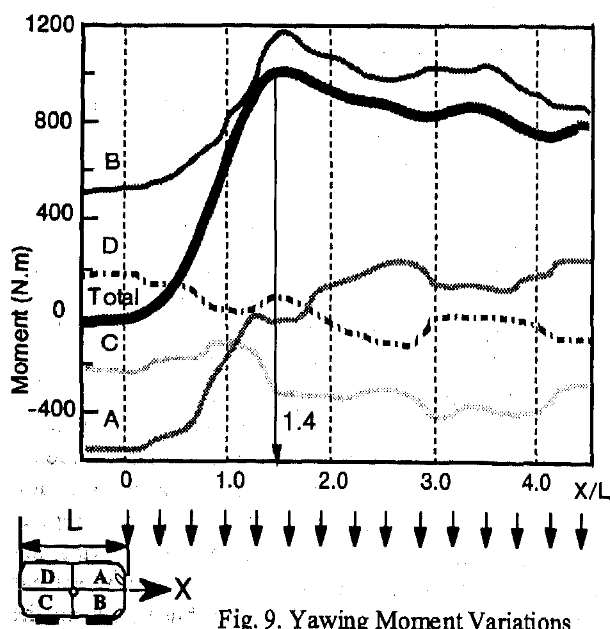


Fig. 9. Yawing Moment Variations

crosswind (X/L=1.4). The composite angle between the automobile velocity (27.8 m/s) and the crosswind (13 m/s) is approximately 28 degrees, and this finding agrees with Fig. 8.

From the changes in each block it has been found that the increase is most evident on the leeward front portion B, of the automobile, then in the windward front portion A of the automobile. In other words, the leeward portion B and the windward portion A of the front part of the automobile are the blocks that contribute to the generation of the yawing moment. This tendency agrees with our steady flow analyses(6),(8). We will now discuss the drag force, side force, lift force and pressure distribution.

DRAG FORCE (CD) - Fig. 10 shows the variation in the drag forces. We have found that the total drag force slightly decreases at first and then increases after the automobile moves one length into the crosswind. Drag force variations in each block show that the drag force on the windward front portion A of the automobile suddenly increases right after the automobile's plunge into the crosswind, but the drag force on the leeward front portion B of the automobile suddenly decreases, so that the two forces cancel each other. Whereas the drag forces on the rear parts of the automobile, C, D, decrease at first and then increase gradually, showing the same tendency as the total drag force.

These results indicate that the total drag force is dependent on thestructure of the wake that forms toward the rear of the automobile.

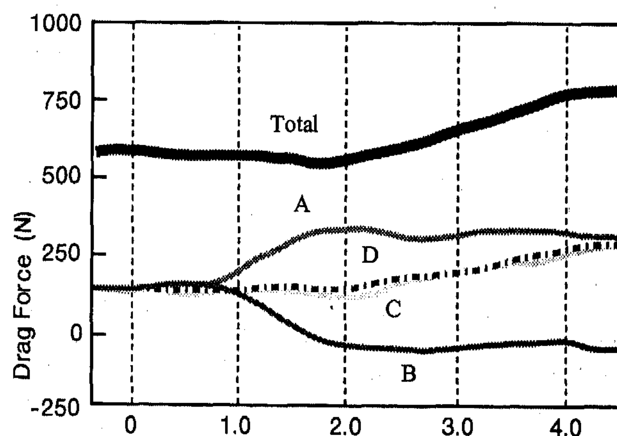


Fig. 10. Drag Force Variations

SIDE FORCE (Cs) - Fig. 11 shows the variation in the side forces. We have found that the total side force increases at the moment when the automobile moves into the crosswind, and there is no noticeable peak. The variations in each block indicate that the leeward portion B and the windward portion A of the front of the automobile influence the side forces greatly. The negative pressure regions formed on both sides of the front part of the automobile as it proceeds are transformed into a positive pressure in the upstream side and anegative pressure in the downstream side. Like the yawing moment, it can be concluded that the front parts A and B of the automobile play an important role in the formation of the side force.

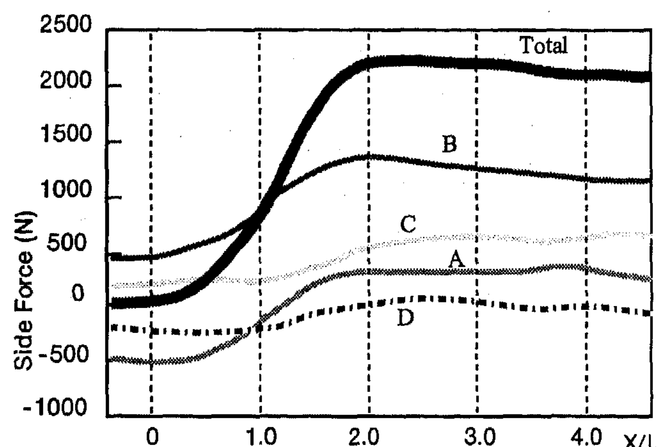


Fig. 11. Side Force Variations

LIFT FORCE (CL) - Fig. 12 shows the variation in the lift forces. We have found that the total lift force turns into a down force when the automobilemoves into the crosswind region, but suddenly increases when the automobile proceeds further into the crosswind. The lift-force variations in each block indicate that while a down force is generated in the windward front portion A of the automobile, a lift force suddenly increases in the leeward front portion B of the automobile. The lift force tends to increase in the leeward rear portion C of the automobile too. The reason for the generation of the down force is that in the windward front portion A of the automobile, stagnation points are created by the composite flow of the crosswind and front wind, and a strong negative pressure due to separation is generated under the floor of section A. The reason for the increased lift force is that a separated flow reattaches to the underfloor on the leeward sides B and C after plunging into the crosswind, so the pressure recovers and increases.

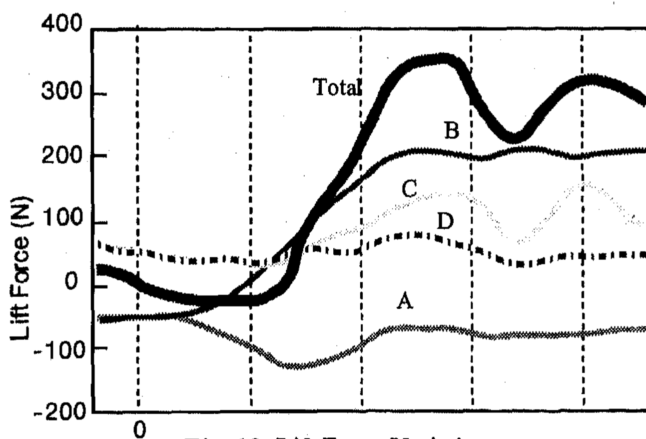


Fig. 12. Lift Force Variations

THE MECHANISM OF YAWING MOMENT GENERATION - Fig. 13 shows the variation in the automobile surface pressure distribution and in the 3D negative pressure isoplethic surface (Cp = -1.5) and positive pressure isoplethic surface (Cp =+0.8) immediately before the automobile's plunge into the crosswind (X/L = 0, X: moving distance, L: automobile length), at the automobile's plunge into the crosswind by 1.4 times the automobile length (X/L = 1.4), and at the automobile's plunge into the crosswind by 4 times the automobile length (X/L = 4.0).

At X/L = 0, it is evident that positive pressure is concentrated in the center of the automobile front and that stagnation point exist at the tip of the front bumper. Negative pressure areas generate symmetrically on both sides of the automobile front end, cancelling out the yawing moment.

At X/L = 1.4, the stagnation points are concentrated in the corner of the windward front portion A of the automobile. In the leeward front portion B of the automobile the negative pressure area expands greatly. These phenomenon contribute to the sudden increasement of yawing moment of block A and C as shown in Fig. 9. This is the reason why the yawing moment reaches its maximum value.

At X/L = 4.0, the negative pressure in the leeward front portion B of the automobile recovers, and a negative pressure is also generated in the leeward rear portion C of the automobile. In addition a crosswind hits the leeward rear portion D of the automobile, cancelling the yawing moment.

From the above observation, we have found that the yawing moment reaches its maximum value immediately after the automobile's plunge into the crosswind.

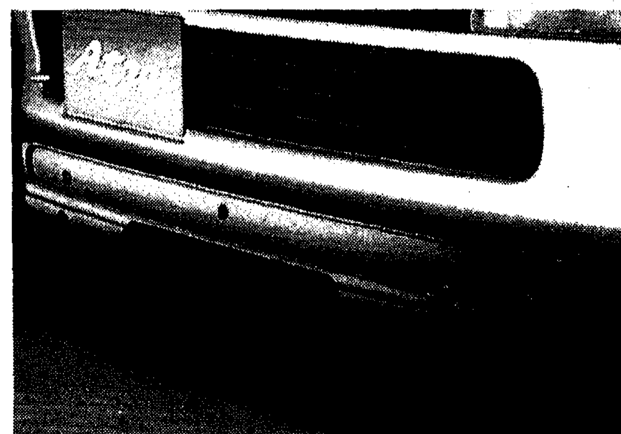
THE EFFECT OF A CROSSWIND SPOILER FOR THE PLUNGE INTO A CROSSWIND

Fig. 14 shows the view of a crooswind spoiler. Fig. 15 shows the pressure isoplethic surface of an automobile equipped with a crosswind spoiler. The stagnation point pressure increases slightly at X/L = 0, but at X/L = 1.4 the strong negative pressure on the leeward front potion B substantially decreases. At X/L = 4.0 the same tendency as that of a standard automobile is exhibited.

The computation results of the yawing moment with the crosswind spoiler are shown in Fig. 16.

These data indicate that a crosswind spoiler can reduce a dynamic yawing moment by approximately 20%, because of the reduction effect of the leeward front portion B.

The yaw rate variations in the crosswind generator region


Fig. 14. Close View of Crosswind Spoiler

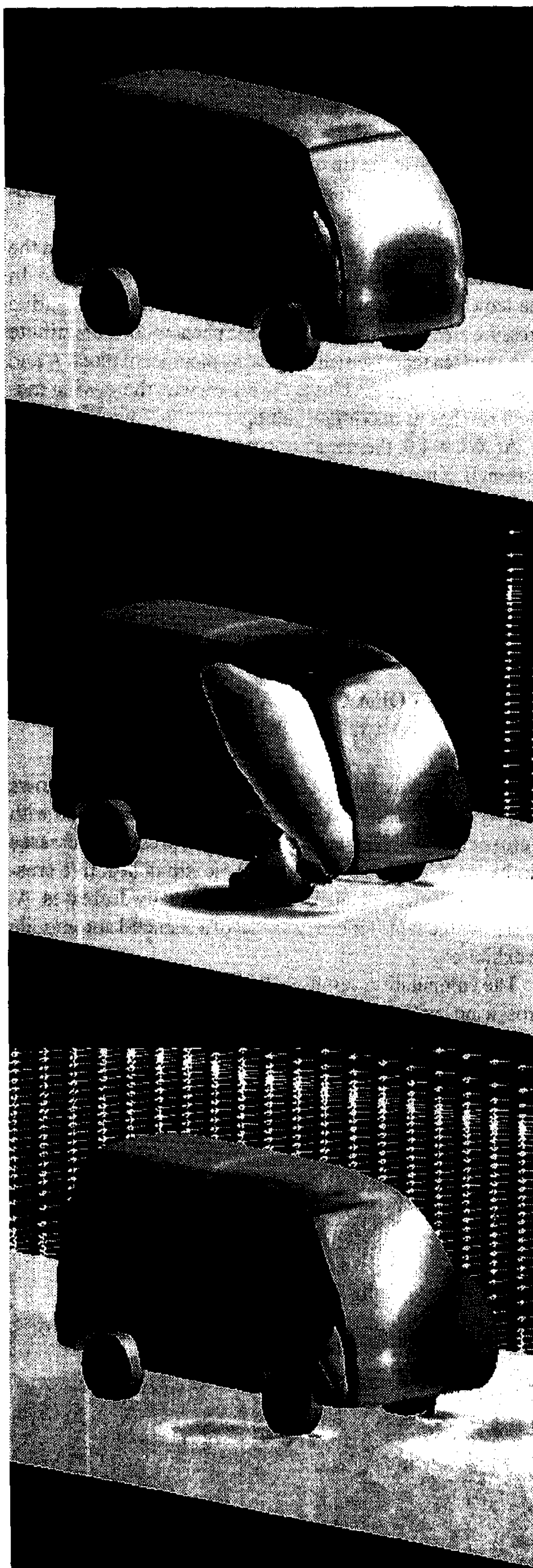


Fig. 13. Pressure Iso-Surfaces of Standard Automobile



Fig. 15. Pressure Iso-Surfaces with Crosswind Spoiler

during an actual evaluation are shown in Fig. 17. The dynamic yaw rate was found to be consistently reduced by the crosswind spoiler by more than 20% within the wind generator region.

Fig. 18 shows a comparison of the total lift forces acting on automobiles with and without a spoiler. In the case of an automobile with a crosswind spoiler a down force is obtained even after the automobile enters into the wind generator region. As a result no sudden increase in the lift force is observed, as opposed to the case of a standard automobile without a spoiler. This is because the spoiler cuts the underfloor flow, creating a stable and strong negative pressure area under the floor. This reduced lift force and the above-mentioned yawing moment reduction effect contribute greatly to the driving stability under actual running conditions.

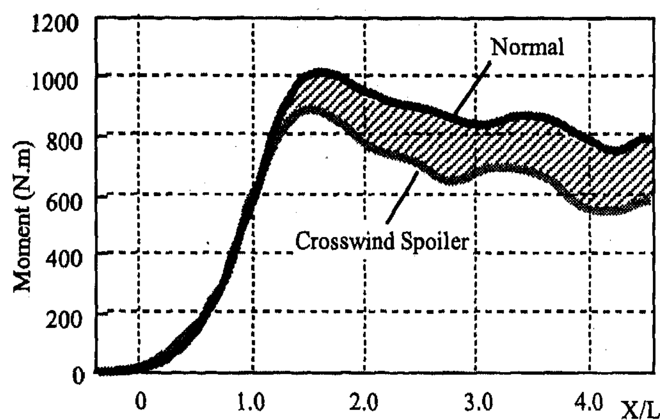


Fig. 16. Comparison of Total Yawing Moment

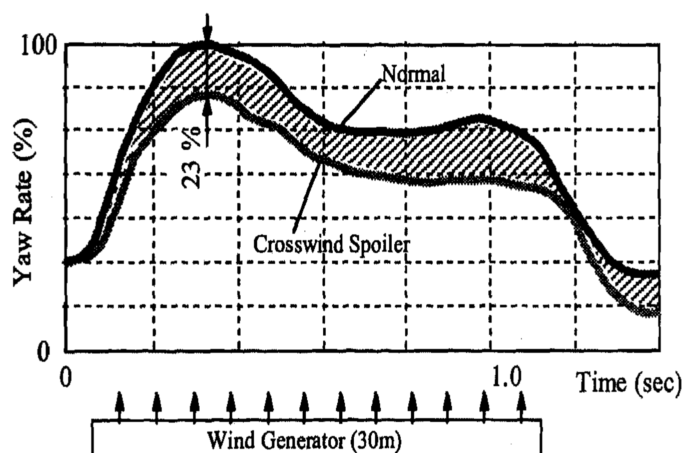


Fig. 17. Actual Evaluation of Yaw Rate

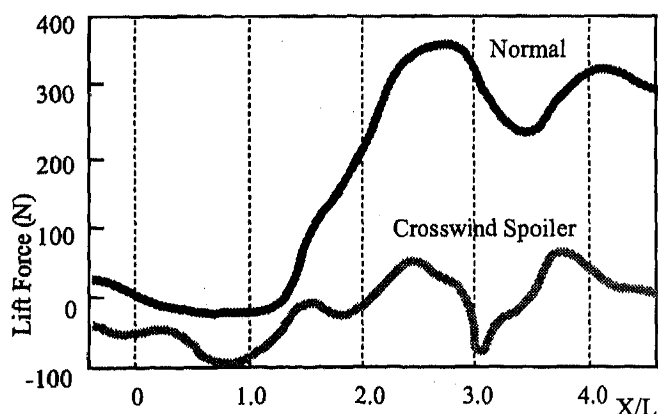


Fig. 18. Comparison of Total Lift Force

UNSTEADY COMPUTATION RESULTS IN PASSING PHENOMENON SIMULATION

It has been reported that the maximum values of the aerodynamic forces (drag force, side force and yawing moment) that apply on the automobile being passed depend on the position of the passing automobile, and that among these forces the yawing moment has the greatest effect on the automobile's behavior[(4),(5),(9)].

In this section we will report on the simulation of a passing phenomenon, in which we visualized a generation mechanism for transient aerodynamic forces.

Yamamoto et al report that when the difference between the velocity of the passing automobile and the passed automobile is relatively small (5.6 m/s or less), static aerodynamic characteristics can explain the passing phenomenon[(4)]. In our study, we attempted computations of a steady flow with two automobiles running at 22.2 m/s with a 0 m/s velocity difference between them and also of a situation in which there is a velocity difference of 11.1 m/s (22.2 m/s and 33.3 m/s respectively) at which a dynamic effect begins to appear, and we identified the difference between these two situations.

YAWING MOMENT (CYM) - Fig. 19 shows a comparison between experiment (the one-box automobile passed by a large bus, which is quated from Yamamoto's report[(4)]) and computation (the one-box automobile passed by one-box automobile) of preceding automobile in a static condition. The computional results are good tendency with experiment. But those are smaller than experimental data. Because the pressure around bus is stronger than one-box automobile and its force is larger.

Fig. 20 shows a comparison between a static computation and a dynamic computation. In order to average out transient data, an ensemble average is taken from ten serial steps. At X/L = 0 (X: moving distance, L: automobile length), the passing automobile catches up with the preceding automobile. It is known from a static passing experiment[(4),(9)] involving a large bus and a one-box automobile in a wind tunnel that the drag force, side force and yawing moment have qualitative maximum values regardless of the shape of the passing automobile. The maximum values for the drag force, side force and yawing moment are seen between X/L = 0 and X/L = 0.5, with that of the drag force and side force approaching closer to X/L = 0, and that of the yawing moment approaching closer to X/L = 0.5.

According to the static computation result of Fig. 20, the waveforms representing the yawing moments are of the same shape on either side of X/L = 1.0 at which the two automo-

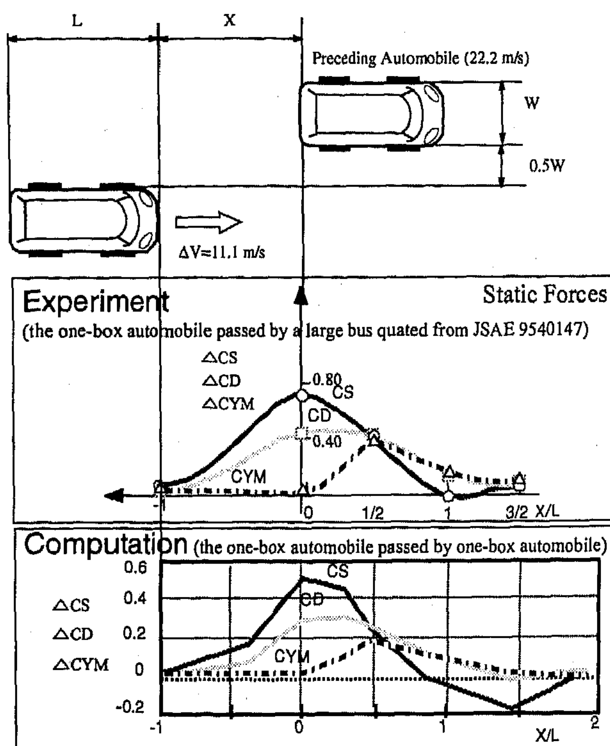


Fig. 19. Comparison of Static Aerodynamic Static Forces of Preceding Automobile

biles run abreast. Whereas in the dynamic computation result, in which there is a difference of 11.1m/s between the velocity of the two automobiles, the passed automobile receives a yawing moment that is almost twice as large as in the case of the static computation. Accordingly the waveforms received by the preceding automobile and the passing automobile are asymmetrical. We believe the reason is that the pressure wave of the passing automobile influences the preceding automobile, and that the greater the velocity difference between the two, the greater the difference between the yawing moments received by the two.

From the above findings it can be concluded that when the velocity difference between the two automobiles is substantial the preceding automobile receives the greater moment and its waveform and that of the passing automobile are also asymmetrical to each other, so it is difficult to evaluate the aerodynamic forces dynamically acting on an automobile using a steady flow analysis.

The waveform of the preceding automobile shows a strong peak when X/L = 0.5 regardless of static or dynamic analysis, and this agrees qualitatively with the above-mentioned experiment reports[(4),(5),(9)]. Fig. 21 is one example of the measured dynamic yawing moment of station wagon passed by a bus (quated from SAE 700037). This result shows a strong peak about X/L = 0.5 too.

Fig. 22 shows dynamic computation results of 3D pressure isoplethic surfaces at X/L = -1.0, 0.25, 0.5, 2.4. Fig. 23 shows 3D isoplethic velocity magnitude surface at 10 m/s.

From X/L = 0 to 0.25, the positive pressure wave from the passing automobile pushes the interference rear side of the

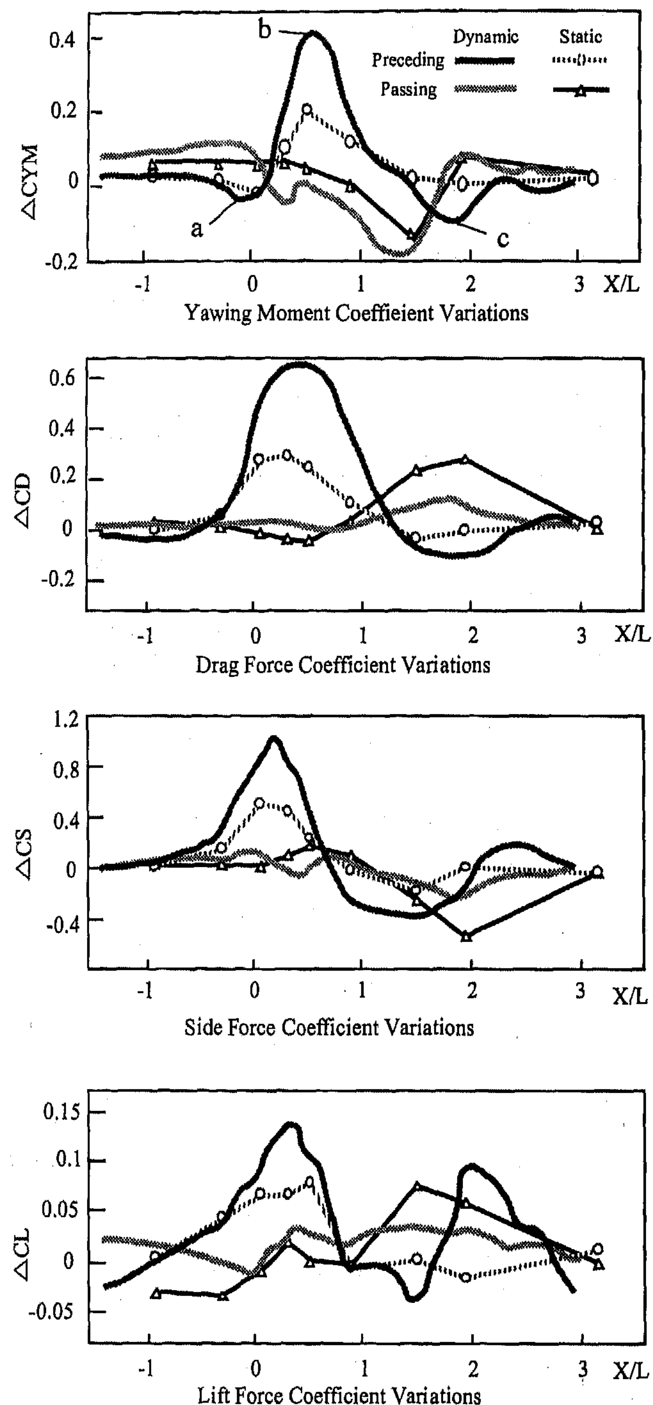


Fig. 20. Computed Static and Dynamic Aerodynamic Forces

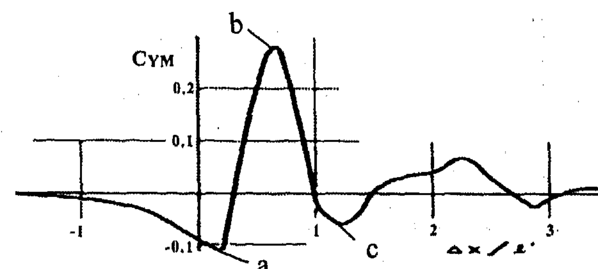


Fig. 21. Experimental Dynamic Yawing Moment of station wagon passed by a bus (quated from SAE 700037)

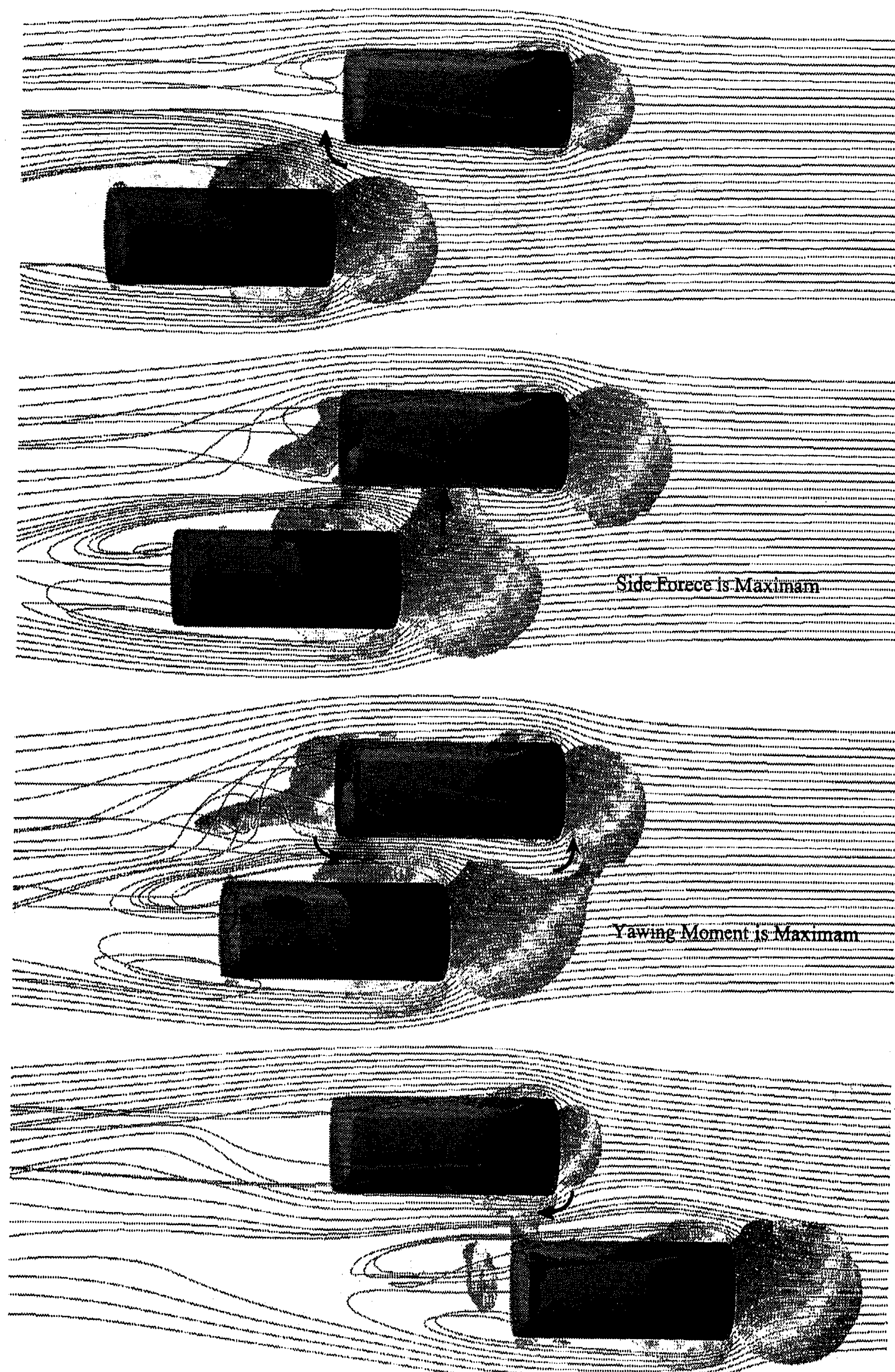
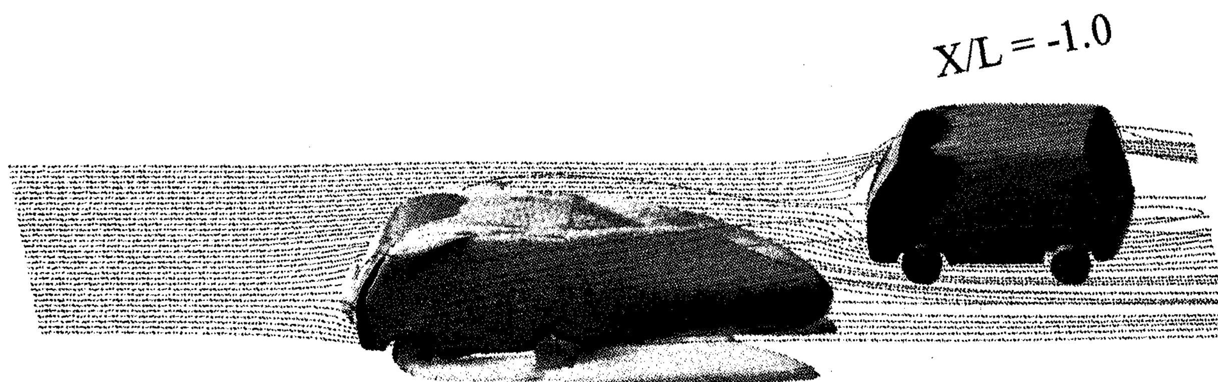
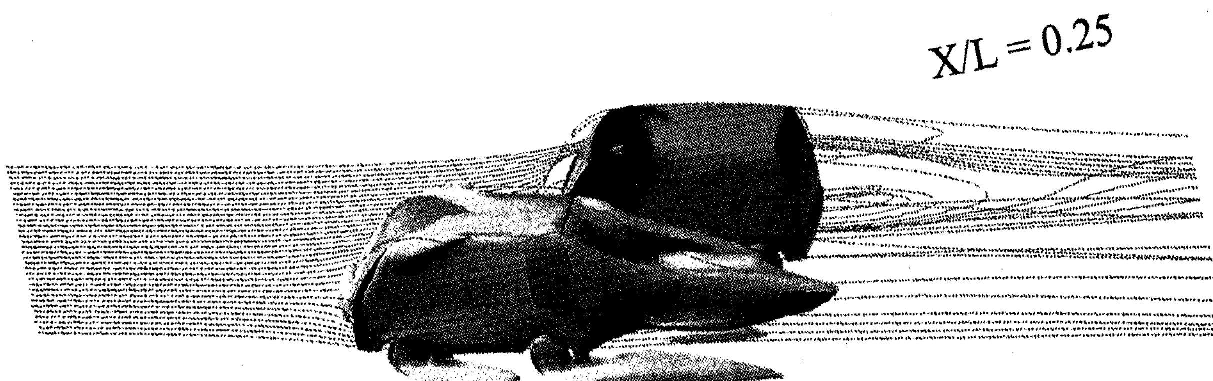
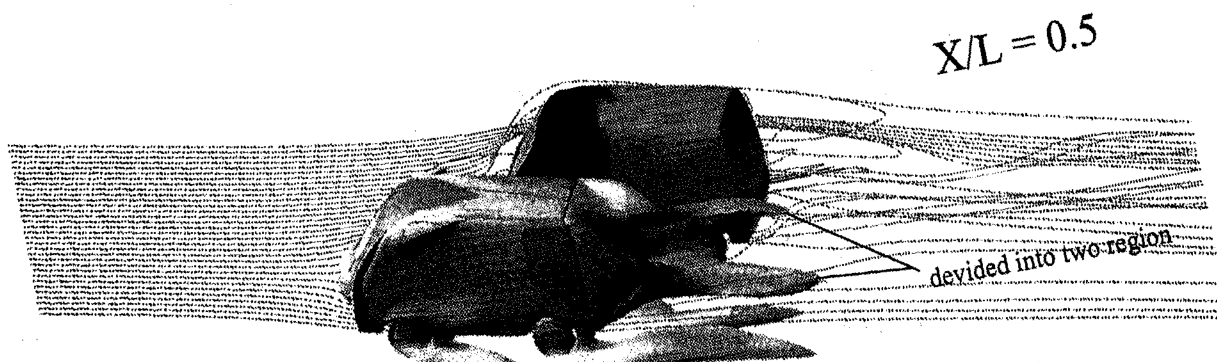


Fig. 22. Pressure Iso-Surfaces from Top View







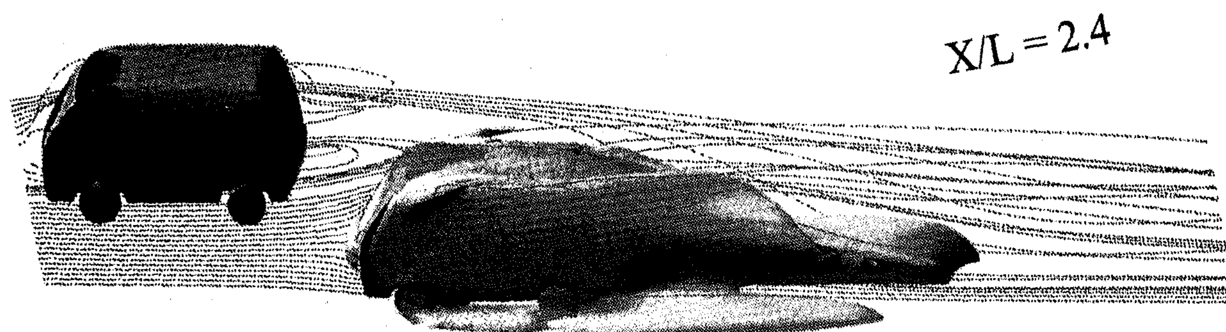


Fig. 23. Velocity Magnitude Iso-Surface of Preceding Automobile at 10 m/s

preceding automobile. This force generates a small clockwise yawing moment as shown "a" in Fig. 20 ,21.

Near X/L = 0.5, the interference side surface of the preceding automobile receives a positive pressure wave and a negative pressure wave from the passing automobile simultaneously, generating a strong counterclockwise moments in the same direction. For this reason the total yawing moment has the maximum value at this positiont as shown "b" in Fig. 20 ,2

From X/L = 1.5 to 2.0, the negative pressure wave from the passing automobile pulls the interference rear side of the preceding automobile. This force generates a small clockwise yawing moment as shown "c" in Fig. 20 ,21.

The preceding and passing automobiles are divided into four blocks that border at the center of each automobile and the results of obtaining the dynamic moments generated in each block are shown in Fig. 24.

At X/L = 0, at which the preceding automobile is caught up by the passing automobile, the moment, which was inclined to decrease at the preceding automobile's rear interference side D due to the positive pressure wave of the passing automobile, suddenly increases as the preceding automobile enters the negative pressure region of the passing automobile.

At X/L = 0.5 the moment at D saturates, but the moment at the automobile front interference side A suddenly decreases.

Consequently, we have found that the yawing moment of the preceding automobile is influenced by the interference sides A and D,depending on the front positive pressure wave and the side negative pressure wave of the passing automobile.

In the case of the passing automobile, on the other hand, a moment acting in the opposite direction from that acting on the preceding automobile increases, and its variation is smaller than that of the preceding automobile because of its higher velocity.

DRAG FORCE (CD) - Fig. 25 shows the variation of the total and each block drag force. In a steady flow the drag forces are symmetrical at X/L = 1.0, but in a dynamic state, in which the velocities of the two automobiles differ, the total drag force of the preceding automobile suddenly increases whereas the

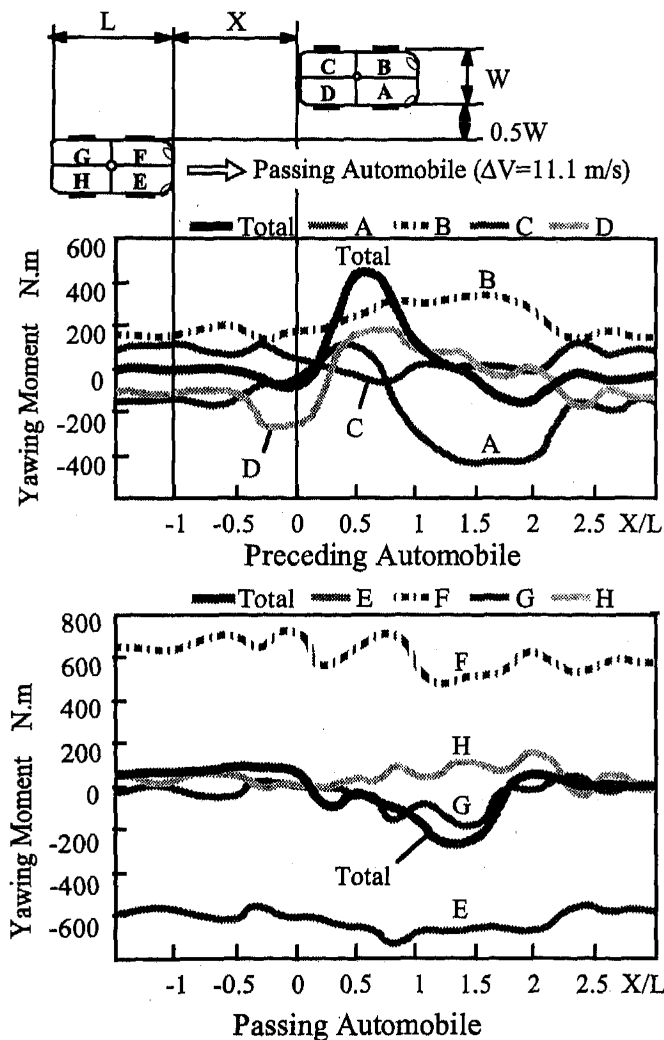


Fig. 24. Daynamic Yawing Moment Variations

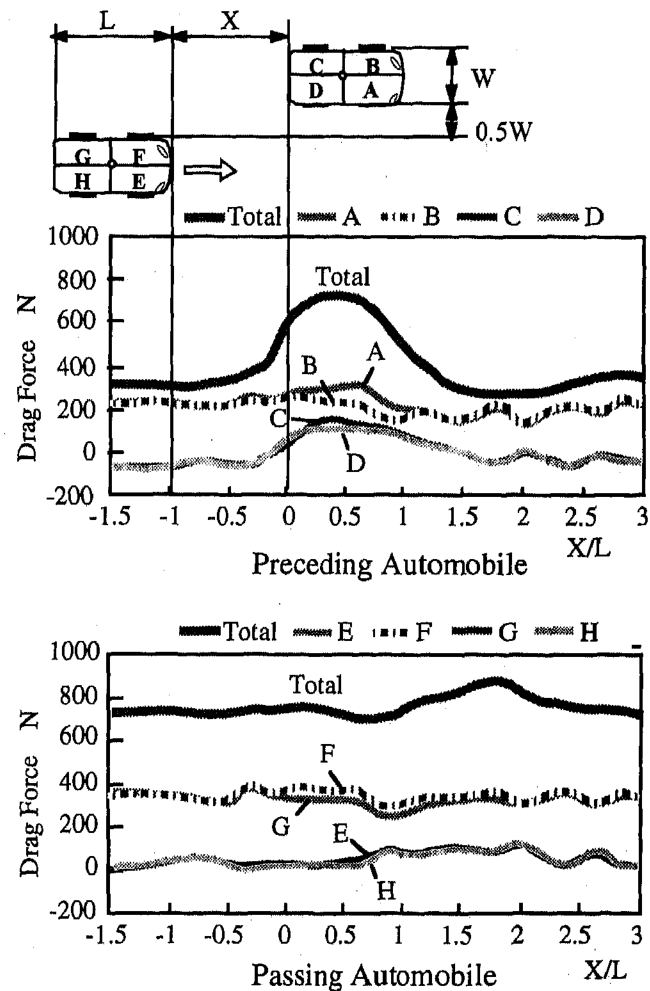


Fig. 25. Dynamic Drag Force Variations

drag force of the passing automobile changes only slightly. The forces applied in the static and dynamic states are apparently different. In either case, however, the maximum values of the drag forces are around X/L = 0 to 0.5, and this agrees qualitatively with the above-mentioned experimental results.

As for each block, the drag forces at C and D in the rear of the preceding automobile increase as the passing automobile approaches, and converge after the passing automobile has passed.

Next, let us focus on the drag forces acting on the preceding automobile. We will examine these forces, which vary remarkably, with reference to the pressure distribution variations (Fig. 22).

Until X/L = 0.5, the positive pressure wave destroys the balance of the wake of the preceding automobile, so the drag force increases gradually.

Near X/L = 0.5, the positive pressure wave recovers a negative pressure on the sides of the preceding automobile. At the same time the negative pressure wave attracts the wake of the preceding automobile and divides into two (Fig.23 at X/L =0.5), so its negative pressure region is expanded into a long, narrow shape. This phenomenon causes a large imbalance of vortices around the automobile, so that the drag force reaches its maximum value.

After X/L = 1.0, the preceding automobile is behind the passing automobile, so the drag force becomes smaller. We found that it converges at X/L = 3.0 or larger.

From the above observations, we have found that the drag force of the preceding automobile depends on its wake shape, which is attracted by the negative pressure wave of the passing automobile.

SIDE FORCE (Cs) - Fig. 26 shows the variation of the total and each block side force. Like the drag forces, the side forces are statically symmetric. Dynamically, while the side force acting on the preceding automobile increases rapidly, the side force acting on the passing automobile changes only slightly. As in the case of moments and drag forces, side forces applied statically and dynamically are different. The maximum values, however, are exhibited at X/L = 0 to 0.25, and this agrees qualitatively with the experimental results.

As for each block, we will focus on the side forces acting on the preceding automobile. We will examine these forces, which vary remarkably, with reference to the pressure distribution variations (Fig. 22).

When X/L approaches 0, first the automobile rear interference side D and then the automobile front interference side A receives a side force.

Near X/L = 0.25 the side force reaches its maximum by the composite force A and B, and when X/L exceeds 0.25 the positive pressure wave of the passing automobile deviates from the center of the preceding automobile, and the flow velocity between the preceding automobile and the passing automobile increases, generating negative pressure. As a result, the side forces at A and D suddenly decrease due to this negative pressure.

Consequently the maximum value of the total side force is shown near X/L= 0.25 and the waveform is peaked.

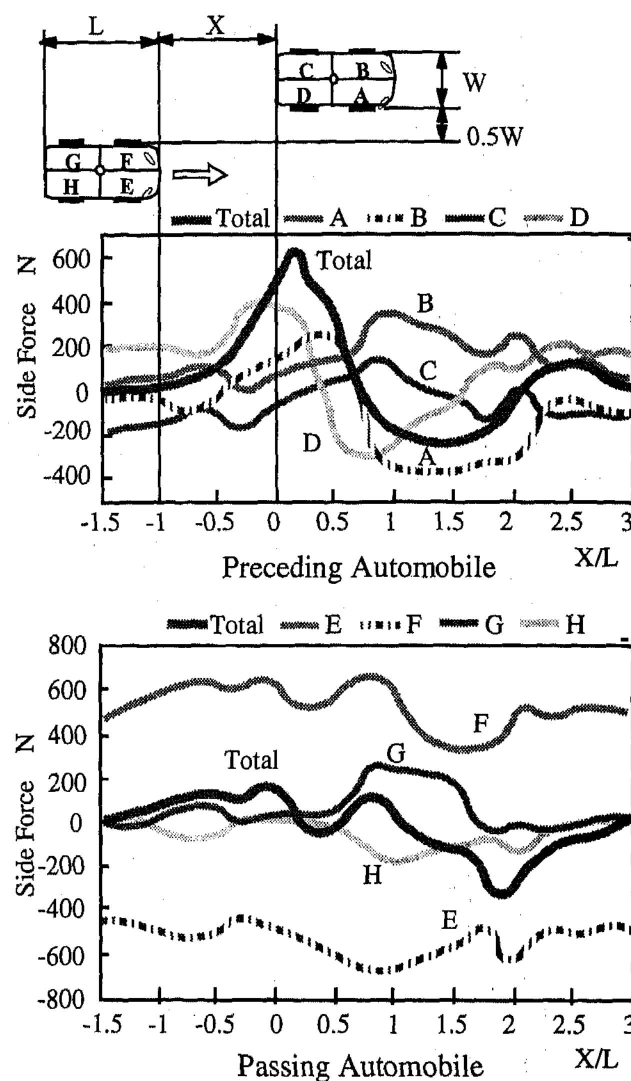


Fig. 26. Dynamic Side Force Variations

Beyond X/L = 0.75 the pressure at D recovers, then the pressure at A also recovers. Beyond X/L = 2.5 the effect of the wake of the passing automobile becomes small, and the pressures converge.

Likewise we have also found that a small counter side force is generated on the passing automobile near X/L = 1.75, which is a symmetrical position of both automobiles.

From the above observations, we have found that the side force of the preceding automobile depends on the pressure variation of the interference side A and D, and is greatly increased by the positive pressure wave of the passing automobile and decreased by the negative pressure generated between the two automobiles.

LIFT FORCE (CL) - Fig. 27 shows the variation of the total and each block lift force. Like all the aerodynamic characteristics so far explained, the lift forces become symmetrical statically at X/L = 1.0, but dynamically the lift forces acting on the preceding automobile vary remarkably while the variations

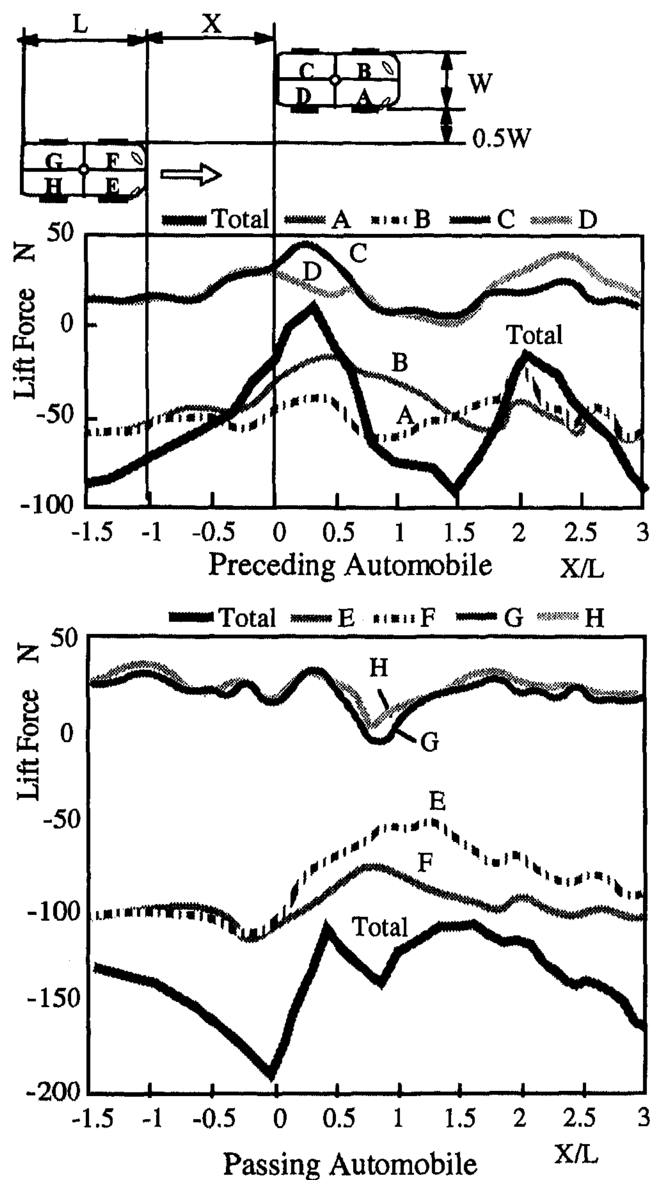


Fig. 27. Dynamic Lift Force Variations

on the passing automobile are small. Different forces are applied depending on whether the relationship between the two automobiles is static or dynamic.

We will focus on the preceding automobile, for which the dynamic forces generated in each block vary remarkably.

Near X/L = 0, only the lift forces acting on the automobile rear interference side D tend to decrease and the lift forces acting on the other areas increase. The increase is the smallest at A, which is the most distant point from the passing automobile, so the positive pressure wave of the passing automobile influences the pressure distribution of the lower part of the automobile. There are two remarkable peaks in the total lift force of the preceding automobile near X/L = 0.25 and 1.9.

Between X/L = 0.75 and 1.75, the negative pressure generated between the preceding automobile and the passing automobile causes the total lift force to decrease.

Near X/L = 2.0, a flow blows up the underfloor, attracted by the negative pressure of the wake of the passing automobile. As the passing automobile becomes distant the values come close to the single driving value.

Consequently, we have found that the lift force depends on the positive pressure wave variation of the passing automobile and the negative pressure variation between two automobiles and the passing automobile's wake.

From these observations of the aerodynamic forces involved in a passing phenomenon, we have found that when the velocity difference between the preceding automobile and the passing automobile is substantial, the preceding automobile receives a larger aerodynamic force than in the static case, and the waveforms are asymmetrical with reference to the moment of the static passing automobile. Furthermore, the force depends on the pressure variation of the interference automobile side.

In other words, static analysis such as wind-tunnel experiments and steady flow simulations are not sufficient to correctly evaluate the aerodynamic forces dynamically acting on an automobile.

CONCLUSION

1) We have developed a computational method using a discontinuous interface grid and moving boundaries which is capable of analyzing transient aerodynamic phenomena (plunge into a crosswind, passing, etc.).

2) The aerodynamic characteristics (yawing moment variations) obtained from our computation results were in good tendency with the experimental results.

3) The yawing moment reaches its maximum value when the automobile plunges into a crosswind by approximately 1.4 times the automobile length, due to the negative pressure on the leeward front portion of the automobile.

4) The preceding automobile receives a larger aerodynamic force than during steady driving when the difference between the passing and preceding automobile's speeds is large and its waveform is asymmetric.

5) The aerodynamic force of a passing phenomenon depends on the pressure variation on the interference side of the preceding automobile, which is influenced by the front positive pressure wave and the side negative pressure wave of the passing automobile.

We believe our development of a dynamic real phenomenon analysis method provides a new possibility for furthering understanding of transient aerodynamic phenomena.

REFERENCES

1. N.Kobayashi, M.Yamada:Stability of a One Box Type Vehicle in a Cross-wind An Analiysis of Transient Aerodynamic Forces and Momenta, SAE Paper 881878

2. K.Sumitani, M.Yamada : Development of "Aero Slit" Improvement of Aerodynamic Yaw Characteristics for Commercial Vehicles, SAE Paper 890372

3. K.Sakakibara, T.Tsutsui : Aerodynamic Yaw Characteristics of Commercial Vehicle Influenced by Front Side Vortices, JSAE Review, Vol13, No.3

4. S.Yamamoto, K.Nakagawa : Aerodynamic Influence of Passing Vehicle on the Stability of Other Vehicles (In Japanese with English summary), JSAE Pre-Paper 956, 9540147,P117-120, 1995-9

5. F.N.Beauvais : Aerodynamic Inputs to a Parked Vehicle by a Passing Bus, SAE Paper 700037

6. K.Okumura, T.Kuriyama et al. : Development of Cross-wind Spoiler, JSAE Review, Vol17, No.3

7. Analysis Moving Boundary Flows (In Japanese), Computational Fluid Dynamic SeriesÇS,University Tokyo Press.

8. K.Okumura, T.Kuriyama : Practical Aerodynamic Simulations (CD,CL,CYM) Using a Turbulence Model and 3rd-Order Upwind Scheme, SAE Paper 950629

9. M.Yokouchi,Y.Kato : A Study of the Aerodynamic Interferences between Vehicles(In Japanese with English summary), JSAE Pre-Paper 952, 9433489, P157-160, 1994-5

970408

Coast Down Method in Time-Distance Variables

V. A. Petrushov
The Central Automobile & Automobile Engine Scientific Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The differential equation of motion of a freely coasting vehicle is analysed and its integral solution in the form of an explicit speed-time function and then a distance-time one is derived. The latter form eliminates the traditional speed term from the equation. By making speed or deceleration measurements unnecessary, a whole set of potential error sources is eliminated, thus improving the sensivity and accuracy of the coast down method. By replacing the distance measurements with precision test track markings, the coasting data requirements are reduced to accurate measurements of time only. This simplifies both the method and the test facility, and improves the reliability of the test data. Because of the efficiency of this a significant volume of test data has been gathered, and included in this report, which characterizes vehicle aerodynamic drag and rolling resistance. In addition to smaller vehicles, these data also include trucks, buses, and articulated vehicles in production and under development which are too large and too heavy to be tested in wind tunnels or on tire dynamometers. For the smaller vehicles, road-to-tunnel and road-to-dynamometer correlations are also presented.

NOTATION

a, b, c coefficients for zero, first, and second power of speed, respectively, in vehicle dimensionless total resistance function

B characteristic relationship between parameters of coast down differential equation

c_x aerodynamic drag coefficient of full-size vehicle in road conditions

c_w aerodynamic drag coefficient of full-size vehicle in automotive wind tunnel

$\bar{c}_w$ aerodynamic drag coefficient of full-size vehicle in the aircraft TsAGI wind tunnel

c_D aerodynamic drag coefficient of vehicle scale model in wind tunnel

c_z aerodynamic lift coefficient at zero yaw

d dimensionless drag force of freely coasting driveline with drive wheels

d_0 initial value of d (at V approaching to zero)

F frontal area

f rolling resistance coefficient

f_0 initial value of rolling resistance coefficient (at V approaching to zero)

G vehicle weight

g gravitational acceleration

$h = gb/(2\delta)$

$h_0 = gb/\delta_0$

I_d rotation inertia of the driveline, excluding the wheels of the drive axles

I_w rotation inertia of one isolated wheel

k_a coefficient of speed influence on dimensionless aerodynamic drag χ

k_f coefficient of speed influence on rolling resistance

m number of vehicle drive wheels

n total number of vehicle wheels

p_t tire inflation pressure

r tire rolling radius

s current distance

S residual coasting distance

S_1 total coasting distance

S_Δ distance between adjacent reflective stakes

ΔS_c distance of vehicle coasting conforming to T_c

t real time

t_a ambient temperature

T residual coasting time

T_1 total coasting time

T_c	time of continuous coasting between test segments under time record interruption
V	vehicle speed
V_1	initial coast down speed
z	number of vehicle wheel revolutions
z_1	total number of vehicle wheel revolutions during coasting
β	symbol for speed-dependent term of traditional form time-speed relation for coast down motion
β_1	value of β for initial coast down speed
β_0	value of β for zero coast down speed
χ	dimensionless aerodynamic drag
δ	rotating mass coefficient of the vehicle
δ_0	rotating mass coefficient of driveline with drive wheels
δc_x	significant error of c_x at 0.95 confidence level
$\delta \bar{c}_w$	significant error of $\bar{c}_w$ at 0.95 confidence level
Π	criterion equal to k_a when it is unaffected by vehicle weight changes
ρ	air mass density
σ	length ratio for two successive road test segments

INTRODUCTION

As known, a triangle of correlation links, which is traditional for applied science (Fig.1), has vertices such as physical modelling, mathematical simulation, and full-scale testing in service conditions. But in spite of intensive development of modern aerodynamics, this triangle still remains incompletely closed at the latter vertex.

The reason is that the measurement of external forces for vehicles and railway transport on the real road surface or on the rails, respectively, is a severe problem.

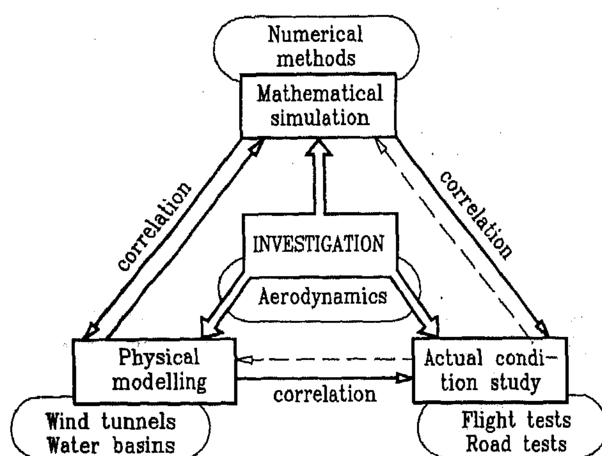


Figure 1: Triangle of applied investigations

So, refinement attempts of the coast down test method for examining vehicle resistance in real conditions are continuing from 1930's up to the present time [1]-[27]*.

The paper that is most often considered to be a pioneer one in coast down test method development is [1]. However, even in 1936 B.S.Falkevitch, in his book [2] published in Russia, described the method being "successfully used in USA" he had called as "Iowa College Method". The main relationship is:

$$S_1 = \frac{\delta G}{2gk_a} \ln\left(1 + \frac{k_a}{Gf} V_1^2\right)$$

(where the notation of the present paper is used).

Taking into account the time of exchanging scientific experience and issuing of that book it may be supposed that the USA method is one of the earliest.

At the present time the significance of this research lead increases due to the impossibility for the full-scale vehicles with frontal area more than 3-4 m^2 to be tested in wind tunnels on account of the stream blockage.

As research practice indicates, the coast down test method also appears most promising for investigation of small values of vehicle resistance in road conditions and even more of its small increments [6][23][24][25], etc.

The main common feature of known varieties of coast down test method is that all of them are based on the measurements of the vehicle speed or deceleration, the latter being extremely small during the coasting. Thus an investigator, independently of his theoretical conceptions, is forced to use one of the following types of speed meters or its combinations: 5-th wheel; Correvit type of device (a noncontact version of 5-th wheel); various light-barrier type devices, or their electro-magnetic analogs; different air-velocity probes, and propeller-vane anemometers. Except for Pitot-static tube and hot wire probes, all the others are operating with the use of division of small measured distance segment by small measured time interval as an algorithm. Meantime the error of measuring the speed and its derivative would always be higher than those for distance and time measurements, being equal to the sum of these correspondent errors.

The replacement of speed or acceleration measurements by distance and time records was not a success until now. Integration of the coasting equation with the function of motion resistance being polynomial, p(V), of the 2^{nd} or higher power yields an implicit function, t=f(V). Consequently, integration of this function instead of explicit one,

* Numbers in parentheses designate references at end of paper

Table 1: Initial relationships for examining motion resistance

Forces reduced to drive wheel circumference	Dimensionless values (forces divided by vehicle weight)		Coefficients of influence of speed V and deceleration $\frac{dV}{dt}$	Formulae of speed influence and motion equations
	Nomenclature	Values at V=0		
Rolling resistance	f	f_0	$k_f\ [s^{-2} \cdot m^{-2}]$	$f = f_0 + k_f V^2$ (1)
Driveline loss	d	d_0	$b\ [s \cdot m^{-1}]$	$d = d_0 + bV$ (2)
Air drag	χ	0	$k_a = c_x \frac{\rho}{2} \frac{F}{G} \left[s^2 \cdot m^{-2}\right]$	$\chi = k_a V^2$ (3)
Total drag	$d + f + \chi$	$f_0 + d_0 = a$	$k_f + k_a = c$	$a + bV + cV^2$ (4)
Inertia of drive wheels and driveline	$-\frac{\delta_0}{g}\frac{dV}{dt}$	d_0	$\delta_0 = \frac{(mI_w + I_d)g}{Gr^2}$	$-\frac{\delta_0}{g}\frac{dV}{dt} = d_0 + bV$ (5)
Vehicle inertia	$-\frac{\delta}{g}\frac{dV}{dt}$	a	$\delta = 1 + \frac{(nI_w + I_d)g}{Gr^2}$	$-\frac{\delta}{g}\frac{dV}{dt} = a + bV + cV^2$ (6)

s=f(t), gives s=f(V,t) where distance is both time and speed dependent. From the mathematical point of view, searching for the problem solution seemed unpromising until now.

INTEGRATION OF THE COAST DOWN EQUATION

An equation of vehicle motion can be formed by means of expressions (1)-(3) as the terms of resistance function (4) represented in Table 1. The resulting equation (6) is most widely used for description of vehicle coasting. Usual consideration of the rolling resistance-speed dependence to be linear at first sight appears like the simplest way of drag component separation, because for such a representation a term proportional to V^2 is an aerodynamic drag. However, relating to the test data obtained by such a simplification no successful repeatability of results was achieved.

The reason is a significant distortion of tire performance because of nonlinear dependence of tire rolling resistance upon speed, as confirmed by [7][8][24] and others. In the paper presented the function f of rolling resistance coefficient is a 2nd power binomial of speed (1) that is fairly adaptable for tire loss description.

The following notation and expressions are used when integrating equation (6) in Table 1.

$$\beta = \arctan\frac{1}{B}\left(\frac{g}{\delta}cV + h\right), \quad (7)$$

$$\beta_1 = \arctan\frac{1}{B}\left(\frac{g}{\delta}cV_1 + h\right), \quad (8)$$

$$\beta_0 = \arctan\frac{h}{B}, \quad (9)$$

where

$$B = \frac{g}{2\delta}\sqrt{4ac - b^2}, \quad (10)$$

$$h = \frac{gb}{2\delta} \quad (11)$$

Then, the traditional integral of equation (6) may be written in the following brief form:

$$t = \frac{\beta_1 - \beta}{B} \quad (12)$$

Defining total coasting time as T_1 (Fig.2) and basing on (12), one can write:

$$T_1 = \frac{\beta_1 - \beta_0}{B} \quad (13)$$

Solving equations (12) and (13) by eliminating β_1 yields

$$\beta = B(T_1 - t) + \beta_0 \quad (14)$$

Introducing a new variable, T, that may be called the residual vehicle coasting time one can write, according to Fig.2

$$T = T_1 - t$$

Then, expression (14) may be written as

$$\beta = BT + \beta_0$$

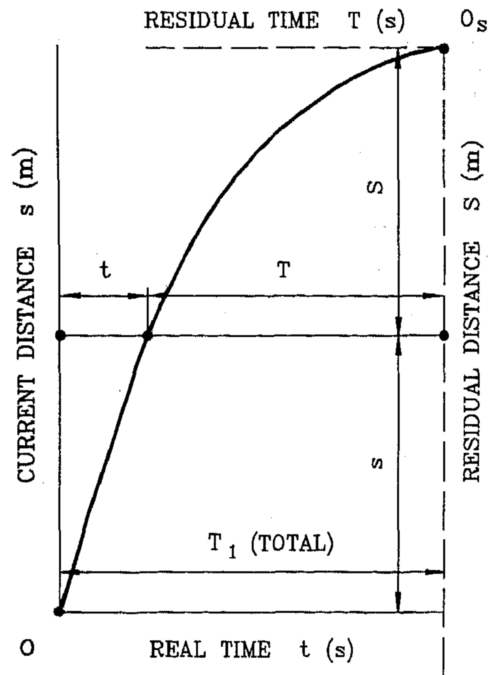


Figure 2: Coasting distance-time plane in coordinate systems of real and residual time

From equality of the angular arguments it follows that its tangents are also equal to each other:

$\tan\beta = \tan(BT + \beta_0)$

Solving this equality and (7) jointly one obtains the following explicit function, V(t)

$$V = \frac{dS}{dT} = \frac{\delta}{gc}\left[B\tan\left(BT + \beta_0\right) - h\right], \qquad (15)$$

where, in accordance with Fig.2, S is the residual coasting distance corresponding to the residual coasting time T. In other words, for the curve S=f(T) the origin of coordinates is shifted from 0 to the stop point 0_S of the vehicle under study.

By integrating expression (15) and by finding the integration constant from S=0 at T=0, the following explicit speed-independent distance-time relationship is deduced

$$\boxed{S = \frac{\delta}{gc}\left(\ln\frac{\cos\beta_0}{\cos\left(BT + \beta_0\right)} - hT\right)} \qquad (16)$$

Formulae (15) and (16) were published in paper [9] for the first time. Similarly, integrating the equation of motion (see Table 1) for the raised vehicle drive wheels and driveline (equation 5) yields:

$$V = \frac{d_0}{b}\left(\exp h_0 T - 1\right), \qquad (17)$$

where

$$h_0 = \frac{g}{\delta_0} b, \text{ and} \qquad (18)$$

$$\boxed{S = \frac{d_0}{b}\left[\frac{e^{h_0 T} - 1}{h_0} - T\right]} \qquad (19)$$

Now it is possible to determine parameters a, b, c for vehicle resistance polynomial from (16), (19) with no coasting speed measurements. In this way a whole set of measurement error sources, as well as the need of correspondent test data processing, is excluded. This suggests that several alternate solutions of the problem under consideration are available.

TOTAL DRAG DETERMINATION METHOD

The most rational way of measurement is the coasting time record starting from four fixed points of the track being marked out and by which two reference distance sections ΔS_{12} and ΔS_{34} are formed (Figs.3a, 3b). Expressing the section length ratio for the general case by σ one can write

$$\frac{\Delta S_{12}}{\Delta S_{34}} = \frac{S_1 - S_2}{S_3 - S_4} = \sigma \qquad (20)$$

where S_i (i=1,2,3,4) - residual coasting distances from the corresponding track points to the vehicle stop point.

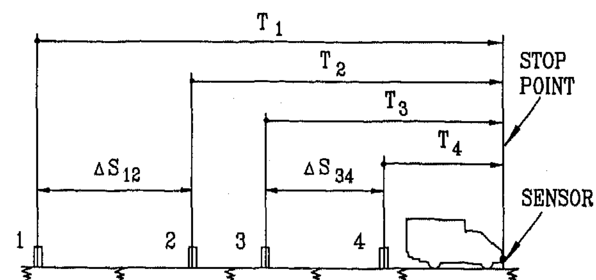


Figure 3a: 4-point time measurement scheme (general case: $\Delta S_{12} \neq \Delta S_{34}$)

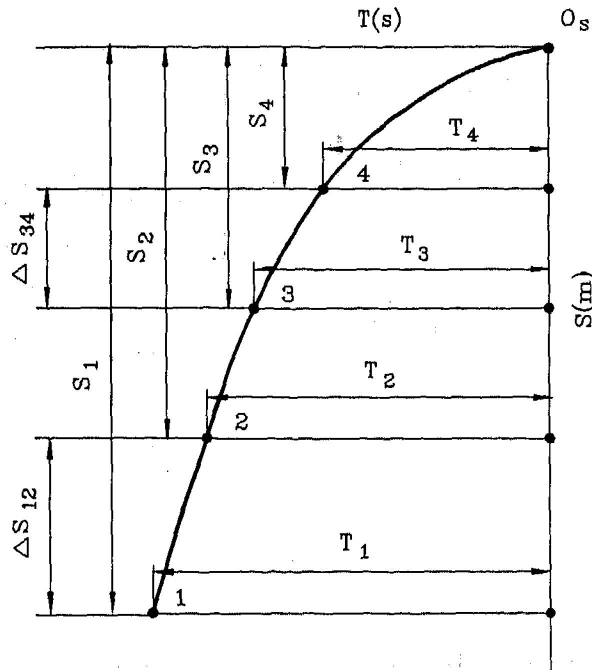


Figure 3b: Distance-time plane for measurement scheme shown in Fig.3a

Having measured the values T_1... T_4 of coasting time and applying (16), one can write two expressions for ΔS_{12} and ΔS_{34} of the form

$$\Delta S_{i,i+1} = \frac{\delta}{gc}\left[\ln\frac{\cos(BT_{i+1}+\beta_0)}{\cos(BT_i+\beta_0)} - h(T_i - T_{i+1})\right] \quad (21)$$

By substituting these expressions into (20) and after simple transformation one obtains

$$\frac{\cos\alpha_2}{\cos\alpha_1}\left(\frac{\cos\alpha_3}{\cos\alpha_4}\right)^{\sigma} = e^{h[T_1-T_2-\sigma(T_3-T_4)]}, \quad (22)$$

where

$$\alpha_i = BT_i + \arctan\frac{h}{B} \quad (i=1,2,3,4) \quad (23)$$

A transcendental equation (22) with (23) having one unknown B is an expression of general form for the method developed. It can be solved by iteration technique.

After calculating B, c can be found from the expression obtained with the help of term-by-term summation of two expressions of the form (21) for ΔS_{12} and ΔS_{34},

$$c = \frac{\sigma\delta}{(\sigma+1)g\Delta S_{12}} \times \left[\ln\frac{\cos\alpha_2}{\cos\alpha_1}\frac{\cos\alpha_4}{\cos\alpha_3} - h(T_1 - T_2 + T_3 - T_4)\right] \quad (24)$$

where α_i is expressed by (23).

With known B and c, parameter a can be found, using (10) and (11):

$$a = \frac{\delta^2}{g^2c}(B^2 + h^2) \quad (25)$$

To apply the expression (25) with (11), the value of b related to the vehicle driveline is needed. On the whole, employment of formulae (22), (24), (25) reduces the parameter measurements to the coasting time only, due to replacement of distance measuring by the track marking being undertaken once per several years.

RATIONAL MEASUREMENT SCHEMES

The simplest scheme of test track marking is that having two adjacent sections of equal length S_Δ (Figs. 4a, 4b), i.e. with three retroreflective stakes in each running direction. For this kind of marking the desirable meanings of S_Δ should lie within the range $(1/3)S_1<S_\Delta<(1/2)S_1$, while its specific values are presented in Table 2.

Overwhelming majority of vehicles were tested by employment of this scheme as applied to which equations (20)-(24) are simplified, because in this case σ=1, T_3=T_2, and T_4 is substituted by T_3.

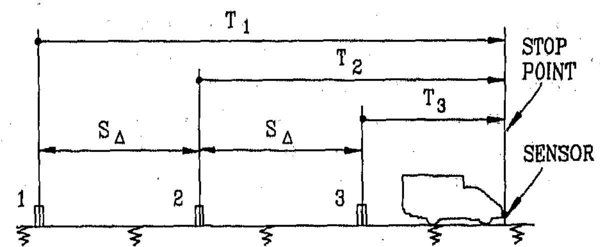


Figure 4a: 3-point time measurement scheme for 2 equal measurement sections

Table 2: Distance S_Δ between reflective marks according to total coasting distance S_1 (a good practice)

S_1 (m) or $z_1 \cdot 10$ (rev.)	850-1100	1100-1300	1300-1500	1500-1700	1700-1900	1900-2100
S_Δ (m) or $\Delta z \cdot 10$ (rev.)	400	500	600	700	800	900

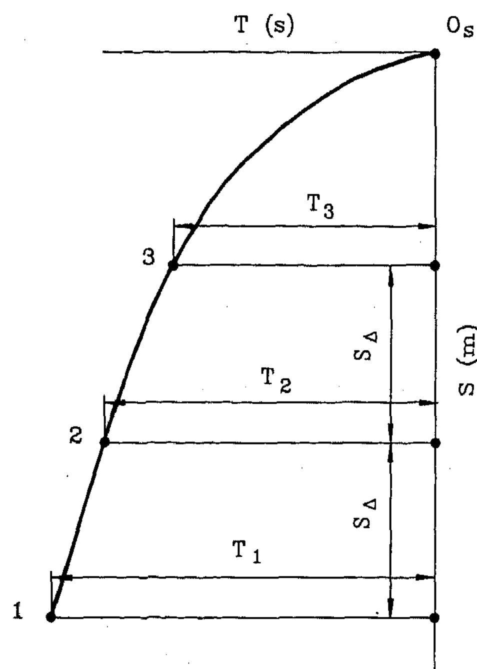


Figure 4b: Distance-time plane for measurement scheme shown in Fig.4a

The scheme with two equal but not adjacent test sections allows the equations (22)-(24) to be simplified, since σ=1.

The scheme with interruption of coasting record is suited for the cases when coasting distance of vehicle or trailer-train exceeds the length of the straight-line horizontal measuring section. It is illustrated by the track having two test sections 1-4 and 4'-7 each of them being marked out into three equal pieces (Fig.5a).

An initial coasting speed on the section 4'-7 is chosen to provide vehicle stopping within section 6-7. An involute of this measurement scheme shown in Fig.5b has an additional unknown time value of record interruption, T_c, in comparison with Fig.3a. Naturally, value of T_c is not a duration of test vehicle U-turn, but it is equal to a magnitude matching both high and low speed segments of curve S=f(T) so that they fall on the common curve. For this case two equations of the form (22) may be related with two unknowns, B and T_c. The

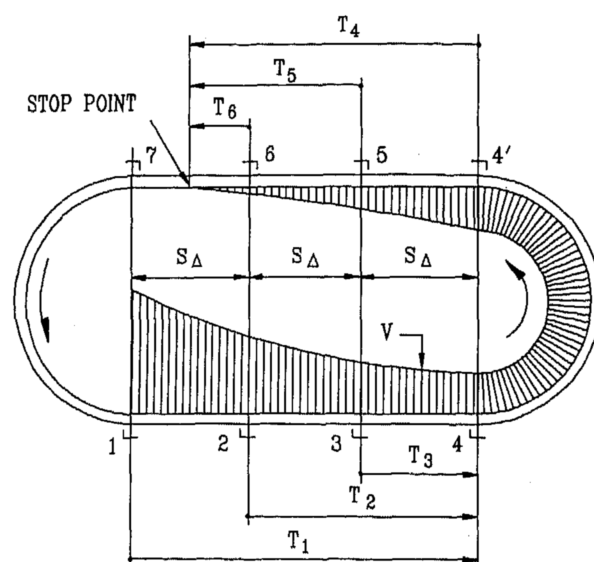


Figure 5a: Time measurement scheme with coasting interruption on U-turns of the annular track

first equation should be composed for two not adjacent but equal track segments 1-3 and 4'-6, when σ=1, and the second one should be written for two unequal segments 2-4 and 4'-5, when σ=2.

By this manner the following set of equations with unknowns B and T_c is obtained, where the coordinates of all six points of the curve S=f(T) are rationally applied:

$$\frac{\cos A_3}{\cos A_1} \cdot \frac{\cos \alpha_4}{\cos \alpha_6} = e^{h(T_1 - T_3 - T_4 + T_6)} \quad (26)$$

$$\frac{\cos A_0}{\cos A_2} \cdot \left(\frac{\cos \alpha_4}{\cos \alpha_5}\right)^2 = e^{h[T_2 - 2(T_4 - T_5)]} \quad (27)$$

where

$A_i = B(T_i + T_c + T_4) + \beta_0$,
(i = 0, 1, 2, 3 with T_0 = 0),
$\alpha_j = BT_j + \beta_0$,
(j = 4, 5, 6).

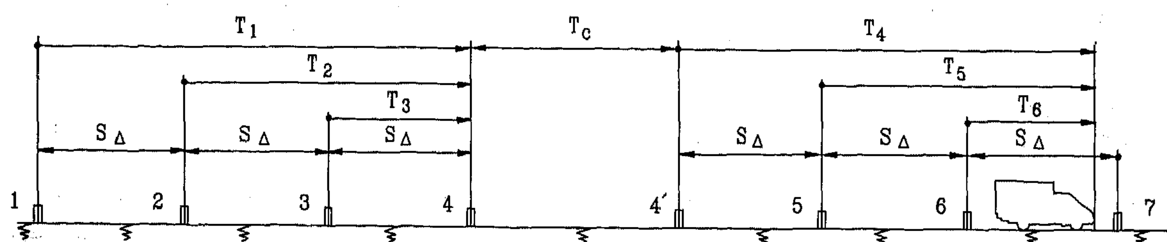


Figure 5b: Involute of measurement scheme with coasting interruption

Thus, each of equations (26) and (27) indirectly incorporates the length of segments of both high and low speed parts of test track. This is the main feature of the set which is solved by iteration method in exterior and interior cycles.

To determine parameter c by utilization of both measuring section data (Fig.5b), equation (24) should be written for extreme coordinates of the segments 1-4 and 5-7. For this case $\sigma=3/2$, $\Delta S_{14}=3S_\Delta$, hence

$$c = \frac{\delta}{5gS_\Delta}\left[\ln\frac{\cos A_0}{\cos A_1}\cdot\frac{\cos\alpha_6}{\cos\alpha_4}\right] - h(T_1 + T_4 - T_6) \quad (28)$$

When applying the measurement scheme shown in Fig.5a and the like (for example, with turn-around loops on the track terminals), the paired runs should be performed in both directions.

Besides that, basing on equations (16) and (21) the variant of determination of parameter a by low speed coast downs and of parameter c by high speed testing is under development now [26][27].

DETERMINATION OF DRIVELINE LOSSES

The simplest variant of the measurement scheme similar to that related to vehicle can be used for determination of neutral driveline losses (see Fig.4b) when raised vehicle' drive wheels and the rotating transmission parts. Using (19), one can write similarly to (21):

$$\Delta S_{i,i+1} = \frac{d_0}{b}\left[\frac{e^{h_0 T_i} - e^{h_0 T_{i+1}}}{h_0} - \left(T_i - T_{i+1}\right)\right] \quad (29)$$

Assuming $\Delta S_{12}=\Delta S_{23}$ and defining $e^{h_0} = x$, one obtains the following expression for determination of x:

$$x^{T_1} - 2x^{T_2} + x^{T_3} = \left(T_1 - 2T_2 + T_3\right)\ln x \quad (30)$$

After evaluation of its root x, parameter searched can be found by the following relation:

$$b = \frac{\delta_0}{g}\ln x \quad (31)$$

Parameter d_0 may be found by summation of two expressions of the form (29) as well as from the total coasting distance value, $S_l=2\pi r z_l$, where z_l is total revolution number of drive wheel, and r is its rolling radius. By summation of two expressions (19) it is easy to obtain

$$d_0 = \frac{2\delta_0}{g}\cdot\frac{S_\Delta \cdot h_0^2}{x^{T_1} - x^{T_3} - h_0(T_1 - T_3)} \quad (32)$$

Inasmuch drive wheel revolution number, z, during the driveline coasting is measured, the value of $S_\Delta=2\pi r\Delta z$ should be substituted into (32), where r is rolling radius of drive wheel, and Δz is revolution number interval to be chosen from Table 2.

SEPARATION OF VEHICLE DRAG SHARES

Through parameters a and d_0, found from (25), (29), the rolling resistance coefficient for zero speed can be obtained as

$$f_0 = a - d_0$$

Separation of aerodynamic drag and rolling resistance that are dependent upon speed to the second power, i.e. terms k_a and k_f, the sum of which is equal to c (see Table 1, column 4, lines 3 and 4), is accomplished by vehicle weight G varying under condition of its aerodynamic configuration invariability and, consequently, of c_x and F constancy. For the pair of tests with different vehicle weights, G_1 and G_2, the following set of equations can be formed.

$$c_1 = k_{f1} + c_x\frac{\rho_1}{2}\frac{F}{G_1}$$
$$c_2 = k_{f2} + c_x\frac{\rho_2}{2}\frac{F}{G_2} \quad (33)$$

As experiments had shown, the difference of 12-15% between values of weight G is sufficient.

If k_f remains unaffected with G changes, then assuming this value as a criterion Π one obtains from the set (33):

$$k_{f1} = k_{f2} = \Pi = \frac{c_2\theta - c_1\nu}{\theta_2 - \nu_2}$$

where $\theta = G_2 / G_1$ and $\nu = \rho_2/\rho_1$.

The searching tests had also demonstrated that more often even for insignificant changes of G parameter k_f is not constant depending mainly on the tire load and temperature. A potential field of these changes is described by the adaptable exponent dependence

$$k_{f2} = k_{f1}\cdot\theta^{\mu}, \quad (34)$$

where μ - exponent.

Solving the set of equations (33) by substitution of (34) obtains

$$k_{f1} = \frac{c_2\theta - c_1 v_2}{\theta^{\mu+1} - v_2}, \quad (35)$$

After that, k_{f2} can be found by (34). Having carried out a test series for three vehicle or trailer-train weight levels, one obtains a set of three equations of form (33) to evaluate μ.

From an experience being accumulated during the tests of significant number of objects a relationship is obtained that allows to calculate parameter μ by varying vehicle weight at two levels, G_1 and G_2:

$$\mu + 1 = 20 \cdot \Pi \cdot 10^4 \quad (36)$$

The latter relation was widely verified by the numerous test series when determining vehicle on-road and wind tunnel aerodynamic drag.

For the passenger cars with relatively small suspension rigidity an additional ballast is placed over the rear axle, and corrector of static deflection (Fig.6) is inserted to the suspension elastic elements.

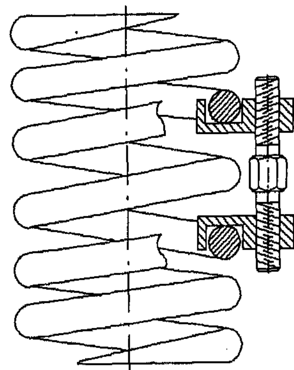


Figure 6: Corrector of suspension static deflection

This corrector readjusts an original body position relative to the road surface. As it is known, the natural frequency of suspended masses is in inverse proportion to the square root from suspension static deflection. Using the correctors like that in Fig.6 additional suspension rigidity is available and equal static deflection of elastic element can be selected for both load values on the rear wheels. Such correction of suspension static deflection is important to provide natural frequency and hysteresis constancy to avoid the additional variables in test series with two vehicle weights. Because of higher suspension rigidity of trucks and many buses, an insignificant change of body height and pitch over the road surface by 12-15% weight increasing may be ignored. For these cases an additional ballast should preferably be located into the centre of mass of such vehicles or close to it. Then, after the certain changes of body position accompanied by vehicle aerodynamic configuration changes, coefficient c_x assumes the meaning averaged for both configurations. For trailer trains it is reasonable to locate an additional ballast in the centre of trailer mass. For the buses with automatic floor level control the conditions of maintenance of invariable body position for two weight levels are the most favourable.

RESEARCH EQUIPMENT

Inasmuch as all variants of the scheme include the record of only one coasting parameter, namely time, the main equipment of the method is essentially simplified taking the form of time interval meter with photoelectric retroreflective sensors (Figs.7, 8).

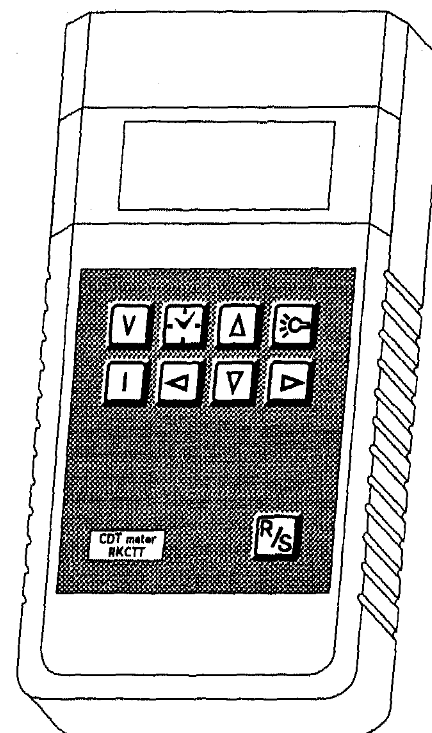


Figure 7: Coast down time meter

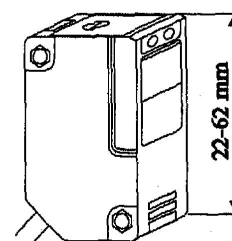


Figure 8: Retroreflective sensor for detection of track marks

As analogy of such time meter, a set of precision stop-watches that are switched on in the moment of vehicle passing by the correspondent distance mark can be conceived of. A photoelectric retroreflective sensor responds to the distance marks in the form of light-reflective stakes (Fig.9). All stop-watches are switched out in the moment of vehicle stopping. So, time data shown in Figs.3a, 4a, and 5a are recorded. For the driveline tests the time meter is equipped by the programmable counter of revolution number of vehicle wheels. At first, an approximate total revolution number, z_1, during the coasting is determined, and then interval revolution number, Δz, is preset providing the condition similar to that presented in the scheme in Fig.4b for the vehicle: $(1/3)z_1<\Delta z<(2/3)z_1$. Recommended Δz values are given in Table 2.

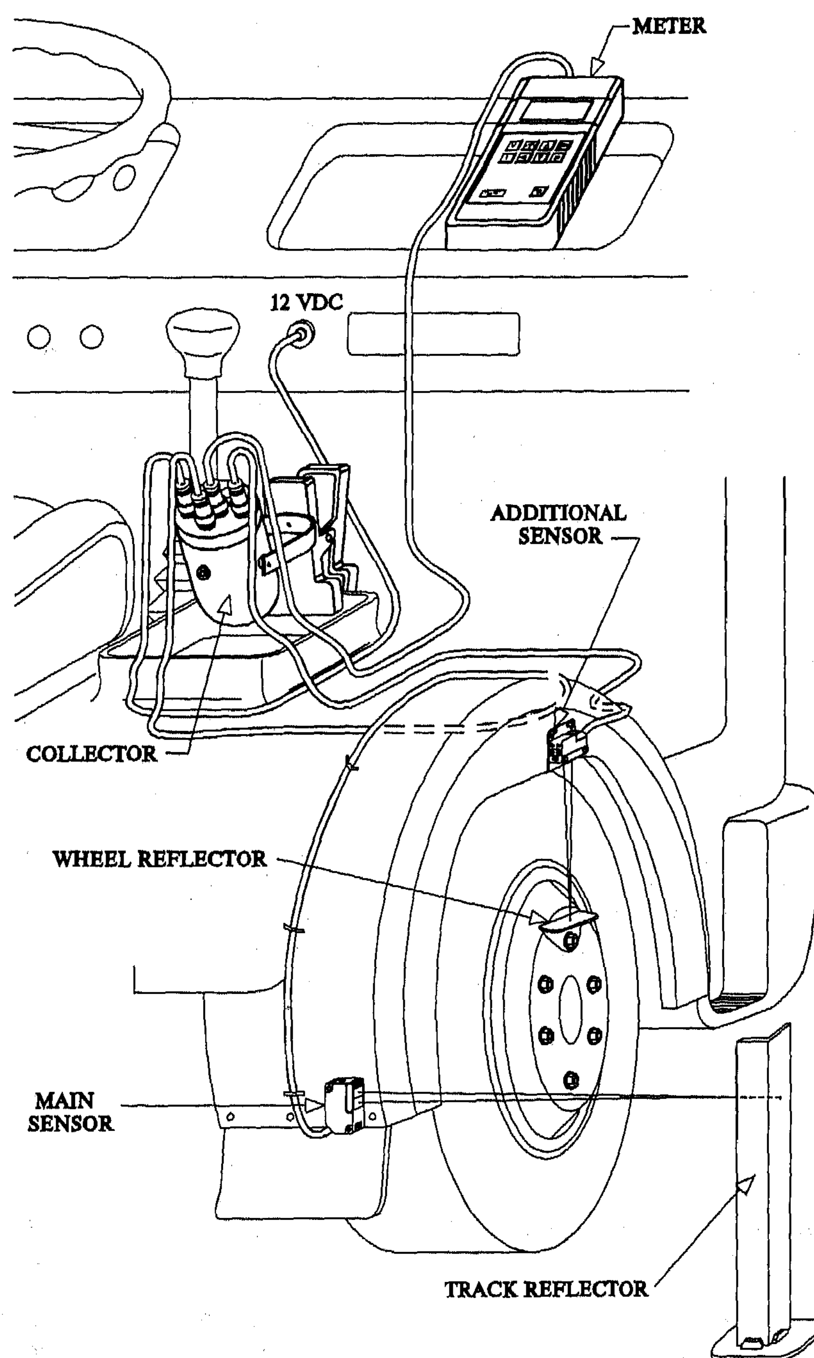


Figure 9: Arrangement of coast down time meter and sensors on the truck

The first specimens of time meter had no built-in memory. The modern time meter with built-in storage capacity 64 Kbyte shown in Fig.8 can transmit data to computer for its processing.

It is designed for simultaneous reception of signals from two retroreflective sensors, one of them being responsive to the track marks, and the second operating as a sensor of vehicle wheel revolution number (Fig.9). It makes possible to measure the wheel rolling radius and to examine carefully an influence of the various factors on it. It also permits to use vehicle tires of high inflation pressure for distance measurements.

An accuracy of coasting time measurements according to the schemes in Figs.3a, 4a is within ±50μs.

The experimental data that are transmitted from the measuring device to computer are processed by use of an adequate software, that was developed on the base of formulae (16)-(36).

In addition, a set of research equipment includes the instruments for measuring tire and final drive temperature as well as tire inflation pressure. It is important to conduct driveline tests at a temperature equal to that of the road tests for proper identification of transmission losses. A portable weather station for atmosphere parameters recording, and vehicle balance or lodometer are also incorporated into the set in use.

It is also essential to mark out the test track with an accuracy of 0.05% or better. For the long tracks, use of the laser geodetic meter that provides marking accuracy of ±1 sm at 1000 m, i.e. ±0.001%, is most efficient.

As a part of research program in Russia two test tracks were marked out and used. The first of them is disposed in the Central Automobile Proving ground (Dmitrov), having the straight-line measuring section of 4.4 km length, and the second 2.2 km length is situated in the GAZ Works Proving ground (Nizhny Novgorod). The marking step of the central part of both tracks is 100 m, while for the rest part it is 200 m. For each particular test series only those reflective stakes were positioned that were required for the measurement scheme selected for the given vehicle.

RESULTS AND DISCUSSION

WIND EFFECT MINIMIZING -Ambient wind is a main potential source of the track data uncertainties.

As known, there are three ways to decrease test errors due to wind action. The first of them is taking into account an influence of wind speed on aerodynamic drag analytically basing on the assumption that wind speed is described by constant vector over a length of the whole test track. Such supposition results in the inadequate model of process in question because uniform wind in the different points of the test track is extremely rare.

The second way which is presented in newer works incorporates into analysis additional simultaneous measurements of the relative airspeed and yaw angle histories by propeller-vane anemometer mounted ahead of the vehicle on the boom. An application of this method is associated at least with four following problems.

1. Aerodynamic interference of the vehicle on the air flow at the anemometer's location and contribution of anemometer to the vehicle drag.

2. Validity of assumption that instantaneous freestream acting on whole vehicle area is adequately modeled by instantaneous relative airspeed and yaw angle which are pointwise estimated by boom-mounted anemometer ahead of the vehicle.

3. The way of introducing a correction for propeller-vane anemometer related to distance constant as well as to vane damping ratio and delay distance.

4. Disregarding the fact that relationship between drag coefficients with yaw angle and at yaw angle equal zero is individual for each vehicle.

Recently, some reports were published with the information about solving the above-mentioned problems and very high drag measurement accuracy reached [23][24][25] that was associated with the need of carrying out up to 20 coastdowns for each vehicle configuration.

The third way of decreasing wind caused variabilities of track data is to limit wind level allowable for test performing as it is regulated by some standards relating to determination of vehicle road fuel consumption and of road resistance parameters for dynamometer simulation.

On the first stage of development it seems rational not to load the method presented by additional methodical constructions and to carry out all the tests at wind speed of no more than 3 m/s as a generalization of experience of standards in action.

As a result, measurement accuracy of parameters c and c_x estimated by the half of significant c_x error at 0.95 confidence level and under 6-fold runs is 2-3% and better for the wind speed up to 1.5 m/s. It may decrease to 5-6% for wind gust speed values of 2-3 m/s. As a rule, a half of significant f_0 error value is not more than 2-5% for the same conditions. The results of numerous road tests are presented below to illustrate these data.

For both above mentioned proving grounds being situated in the wooded middle region of Russia, such wind conditions accompanied by absence of sediments more often occur during the period from April to October and, as a rule,

approximately from 19.30 P.M. to 2.00 A.M. In each month of this period the number of favourable days is sufficient for the intensive experimental working, to be alternated with test preparation and data processing when rainy or windy.

ROAD - TUNNEL CORRELATION OF AERODYNAMIC DRAG - The method was initially applied for the operational development of the passenger cars concurrent to wind tunnel testing. Two types of wind tunnels were used. The first, an aircraft type, was the Central Aerohydrodynamic Institute (TsAGI) subsonic wind tunnel with the open working section and nozzle end cross section area of 38.5 m^2. A car was placed in the flow core on the flat fixed disk screen 6 m diameter and 0.2 m thickness. Supporting plates for the car wheels were connected to the aerodynamic balance by the long holders and positioned in the screen slots. The second, an automotive type, was the wind tunnel of the Central Proving ground in Dmitrov having 26.7 m^2 cross-sectional area and 16 m length working section with slotted walls.

The comparison data obtained by both methods for 4.8 m length sedan are represented in Table 3. By means of an additional drag reducing equipment installed to the vehicle coefficient c_x has been successively diminished. The changes were measured with comparable accuracy both by the aerodynamic balance in the wind tunnel and by the road method. Each vehicle aerodynamic configuration was evaluated by means of 6 runs on the road (three in each direction). An example of the similar comparative tests with the automotive wind tunnel being applied is represented in Table 4.

Table 3: Comparison between c_x obtained on the track and $\bar{c}_w$ in TsAGI wind tunnel for 4.7 m length car

Completeness (step-by-step element furnishing)	Track*			Wind tunnel**		
	$c_x(\pm\delta c_x\ \%)$	$\frac{c_x - c_{x1}}{c_{x1}}$ (%)	wind speed [m/s]	$\bar{c}_w(\pm\delta\bar{c}_w\ \%)$	$\frac{\bar{c}_w - \bar{c}_{w1}}{\bar{c}_{w1}}$ (%)	$\frac{c_x}{\bar{c}_w}$ (%)
1. Original	0.571(±2.1)	0.0	1.2	0.568(±2.6)	0.0	1.01
2. Linings on front body pillars	0.549(±3.5)	-3.9	1.8	0.565(±2.5)	-0.5	0.97
3. Flat wheel caps	0.545(±1.1)	-4.6	0.6	0.564(±2.5)***	-0.7	1.03
4. Front spoiler	0.519(±1.3)	-9.1	0.7	0.539(±2.6)***	-5.1	0.96
5. Rear spoiler	0.491(±1.4)	-13.4	0.8	0.500(±2.6)	-12.0	0.98

Notes: * 6 measurements for each configuration;
** 10 measurements for original configuration only;
*** Without linings on front body pillars.

Table 4: Comparison between c_x obtained on the track and c_w in automotive wind tunnel for 3.7 m length car

Completeness (step-by-step element furnishing)	Track*		Wind tunnel**	
	$c_x(\pm\delta c_x\ \%)$	wind speed [m/s]	c_w	$\frac{c_x}{c_w}$ (%)
1. Original	0.492(±6.5)	3.0	0.442	1.11
2. Shields of radiator	0.490(±5.1)	2.5	0.439	1.11
3. Fairing on rear-wheel suspension	0.495(±3.6)	1.8	0.436	1.14
4. Triangles under rear bumper	0.485(±3.9)	1.9	0.435	1.11
5. Rear fairing under engine	0.483(±1.9)	0.9	0.429	1.12
6. Enhanced negative body pitching by (12 mm body rear lifting)	0.501(±5.0)	2.5	-	-

Notes: * 6 measurements for each configuration;
** No error data.

The testing of additional elements of aerodynamic equipment has shown its small efficiency. An adequate correlation was obtained but, unlike previous example, with about 10% higher c_x for the road conditions.

By the regular utilizing the method developed a significant quantity of test data was accumulated during joint investigation with Russian automotive works. These data characterize correlation between c_x in road condition and the similar coefficients c_w in automotive wind tunnel and $\bar{c}_w$ in TsAGI wind tunnel (Fig.10). Results obtained having very high level of correlation are generalized by the following relations:

$$c_x = 1.12c_w \pm 6.5\% \quad (37)$$

$$c_x = \bar{c}_w \pm 2\% \quad (38)$$

This expression means that all data scatter excluding rough errors (as point 29 in Fig.10) is covered by represented maximum-minimum values.

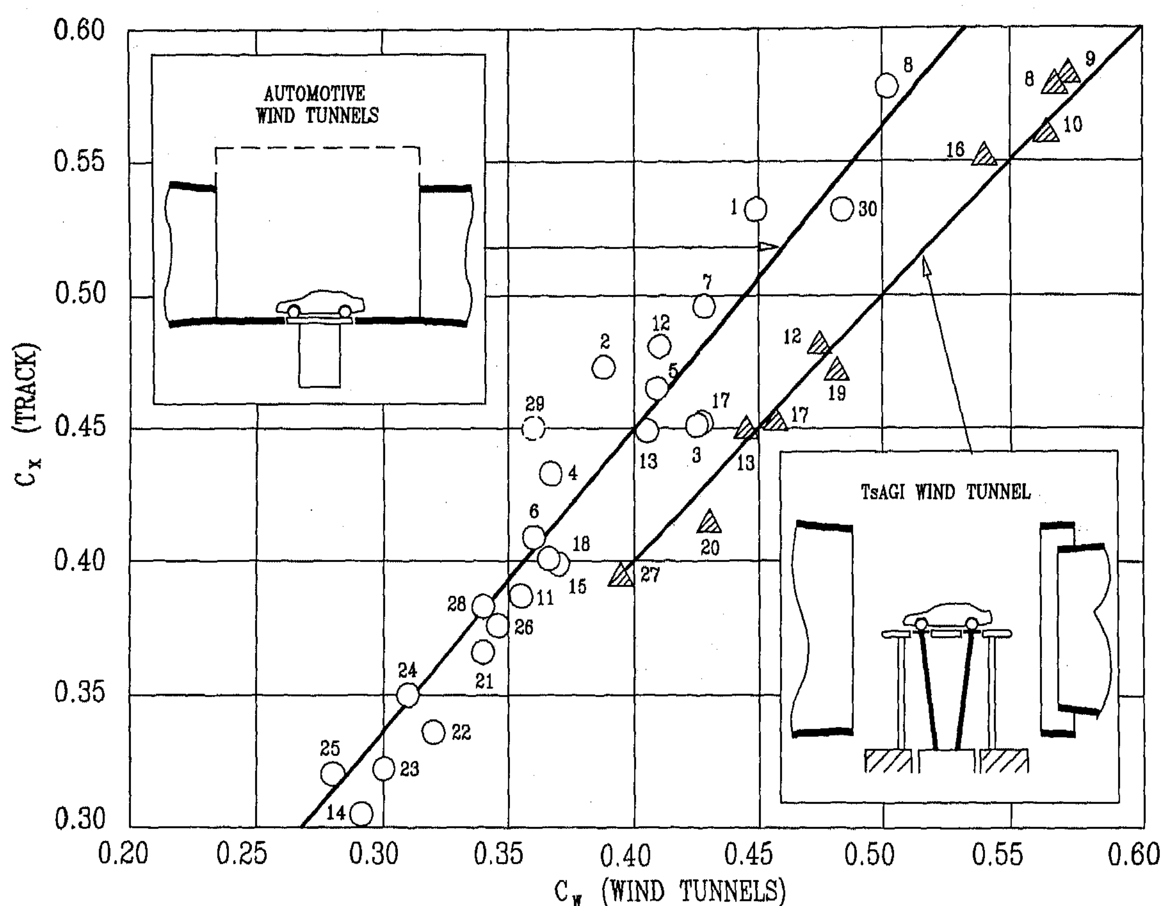


Figure 10: Correlation between on-road and wind tunnel aerodynamic drag coefficients of full-scale cars. Symbol numbers correspond to the following car makes with frontal area in m^2:

1	ZAZ-1102 series 1	1.61	11	Fiat Uno 55 S	1.83	21	Moskvitch-2141	1.89
2	ZAZ-1102	1.61	12	Renault-14	1.83	22	Ford Taurus	2.02
3	Austin Mini-Metro	1.66	13	Opel Kadett	1.84	23	Mercedes-260 E	2.06
4	Opel Corsa 1.25	1.71	14	Debute 1 (NAMI)	1.85	24	Renault-25 GTX	2.06
5	Peugeot-104	1.72	15	Renault-9	1.85	25	Peugeot-605 SR	2.08
6	Peugeot-205 GL	1.74	16	VAZ-2108 series 1	1.87	26	Ford Scorpio	2.08
7	Honda City E	1.74	17	VAZ-2108 series 2	1.87	27	GAZ-3105	2.09
8	VAZ-2101	1.74	18	VAZ-2108 series 3	1.87	28	BMW-750	2.13
9	VAZ-2105	1.76	19	VAZ-2108	1.87	29	Mercedes-560 SEL	2.19
10	Moskvitch-2140	1.78	20	Moskvitch-2141 series 1	1.89	30	UAZ-3160	3.03

Better correlation factor to TsAGI (38) in comparison to automotive wind tunnel (37) is explained by testing the same vehicle specimens that have been tested on the road. A large share of vehicle specimens tested in automotive wind tunnels is of the same make names and numbers, but has different body numbers than tested on the road. Besides that, the data were obtained in the different wind tunnels. Despite the fact that Buchheim's et al system of comparison [28][29] is indirectly applied for the data obtained in different wind tunnels, the value of correlation factor of the data field expressed by relationship (37) was affected by the difference of the tunnels themselves.

The reasons of higher c_x values for the same vehicle in the road condition in comparison with automotive wind tunnels have been discussed for a long time including wind tunnel blockage, stationary vehicle wheel, etc. By opinion of different authors, this phenomenon is mainly affected by the presence of the boundary layer on the stationary floor of the working section of the automotive wind tunnels that is notified by [1][3][5][25], and in the other works. This conclusion has been also confirmed by the author of the present paper in [18]. Apparently, the explanation of equality (38) can be found in the favourable conditions for blowing off the boundary layer from the road simulating disc in the aircraft wind tunnel. Besides that, the further application of the method along with wind tunnel testing has revealed that by reducing the flow speed in wind tunnel to 25-10 m/s, i.e. to the values comparable with initial and average speed values of the road tests, the coefficients c_x and c_w are converging (Table 5). More extended further studies should clear an effect of tunnel operational speed on c_w reduction like an effect of flow around sphere including crisis region.

Table 5: Aerodynamic drag coefficient as a function of full-scale wind tunnel operational speed (Test object: Jeep 4x4, F=3.028 m^2, c_x=0.5317)

V (m/s)	10.6	15.4	20.2	25.1
c_w	0.5222	0.5079	0.4991	0.4922
c_z	0.2337	0.2323	0.2169	0.2157
c_w/c_x	0.98	0.96	0.94	0.93
V (m/s)	30.1	35.0	40.0	45.6
c_w	0.4854	0.4778	0.4742	0.4751
c_z	0.2250	0.2259	0.2327	0.2247
c_w/c_x	0.91	0.90	0.89	0.89

It is noteworthy that for the case of additional vehicle configuration changes in order to reduce aerodynamic drag the road-tunnel data correlation degree decreases. This is illustrated by results for the 5050 mm length passenger car shown in Table 6. From these results it can be seen that maximum of discrepancy occurs when estimating an influence of drag reduction devices installed in the vehicle understructure (see, e.g., lines 2,3 of Table 6). Reduction of c_w in wind tunnels is lower than that of on-road c_x, as it had been recently shown by Le Good et al [25].

On the next stage of the method application the different full-size cars, trucks, buses, and articulated vehicles were tested. During the testing a correlation between c_x in road conditions and c_D of the scale models of these objects including bulky vehicles was examined (Fig.11).

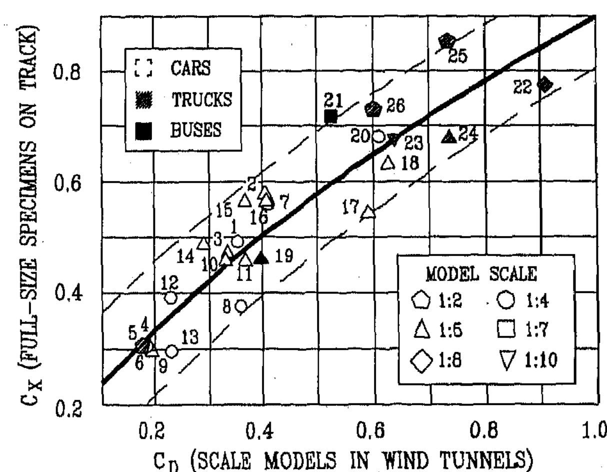


Figure 11: Correlation between aerodynamic drag coefficients of full-scale vehicles on the track and of its scale models in wind tunnels.

Symbol numbers correspond to the following vehicle makes with frontal area in m^2:

1 ZAZ-1102	1.71	15 UAZ-3160 *	3.03
2 VAZ-2102	1.74	16 UAZ-3160	3.03
3 VAZ-2108	1.87	17 UAZ-3170	3.06
4 Debute-1 (NAMI)	1.85	18 UAZ-469	3.39
5 Debute-1 (NAMI)	1.85	19 UAZ-452A	4.20
6 Debute-1 (NAMI)	1.85	20 GAZ-3302	3.78
7 Moskvitch-2140	1.78	21 LAZ-4207	7.36
8 Moskvitch-2141	1.89	22 ZIL-130	4.80
9 Okhta	2.25	23 KamAZ-5320	6.90
10 GAZ-24	2.28	24 KamAZ-5320	6.90
11 GAZ-24	2.28	25 Scania R142H (trailer-train)	9.20
12 GAZ-3105	2.09		
13 GAZ-3105 *	2.09	26 Scania R142H (with airshield)	9.20
14 GAZ-14	2.87		

Note: * Improved.

A large experience of automotive works in the field of wind tunnel testing of scale models being built during the preparation for production of the

Table 6 : Comparison of aerodynamic drag variations for road tests (Δc_x), aircraft wind tunnel TsAGI ($\Delta \bar{c}_w$), and automotive full-scale wind tunnel (Δc_w)

Changes of configuration and outfit	Explanation	Aerodynamic drag changes (%) *		
		Δc_x	$\Delta \bar{c}_w$	Δc_w
$c_x = 0.3534$	Sealing of headlight periphery 1 and of joint between front fenders and bumper 2	-9.9	-2.8	-1.0
$c_x = 0.3165$	Mounting of fog lights flush with spoiler 1. Installation of fairings 2 and 3 on spare wheel and rear suspension	-9.4	-4.5	-3.4
$c_x = 0.3095$	Windshield wipers put away behind hood deflector 1; deflector 2 radius enlargement	-1.8	-1.0	-1.0
$c_x = 0.2956$	Mounting of flat wheel caps, enhanced negative body pitching	-3.5	-3.6	2.3
$c_x = 0.2806$	Enhanced negative body pitching (second level)	3.8	-1.0	0.5
$c_x = 0.2738$	Removal of right-hand mirror	-1.7	-3.4	-1.5

Note: * Difference between $c_x (c_w)$ of each configuration and those of preceding configuration, divided on basic values: $c_x = 0.3925$ and $c_w = 0.3840$ of original outfit is assumed to by 100 %.

full-scale models has been utilized. As a result, on the base of the tests of 26 scale models in wind tunnel and of their full-scale analogs in road conditions the following correlation relationship for the forecast of c_x by c_D was obtained:

$$c_x = 0.9 \cdot \sqrt[8]{c_d^5} \pm 0.11 \qquad (39)$$

From the graphic data of Fig.11 and generalizing formula (30) related to the models without simulating air streams in the underbonnet space and in the ventilation systems, it suggests that for the full-scale models the error of c_x prediction can reach 15-35% (lower value for larger c_D).

ROLLING RESISTANCE - An effect of speed and tire inflation pressure on vehicle rolling resistance is shown in Figs. 12, 13, and 14.

These data demonstrate the method sensitivity to small variations of rolling resistance as affected by various factors. The experiments had confirmed the f_0 decreasing with the rising of inflation pressure. However, the same factors do not cause obligatory falling down of coefficient k_f that characterizes rolling resistance dependence upon speed (Fig.13). For a number of tire models this coefficient rises with p_t increasing, sometimes passing through the peak. For several tire models it passes through the valley which coincides with pressure value recommended by manufacturer for the correspondent load. This effect may be caused by the interaction of two opposite factors accompanied by raising of p_t, i.e. reduction of unsprung mass

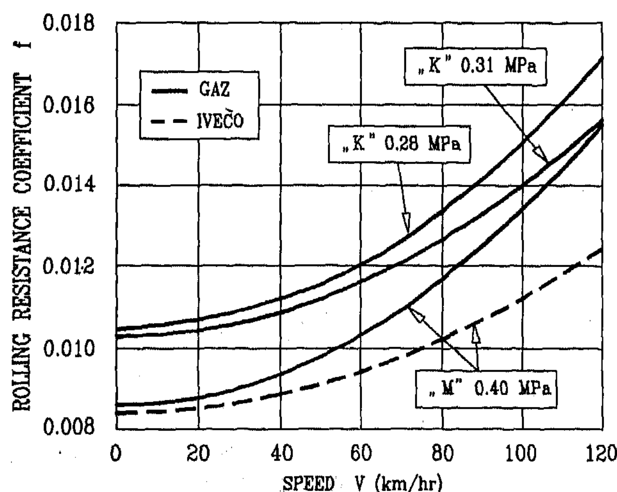


Figure 12: Typical rolling resistance coefficient versus speed for 4x2 1.5 t trucks GAZ-33022 and Iveco-Turbo-Daily 35.10-320 with tires 175/75R16 (at manufacturers "K" and "M" recommended pressure)

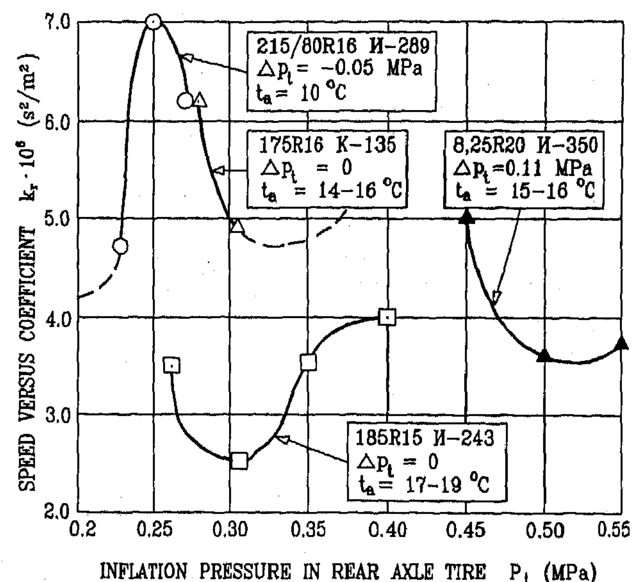


Figure 13: Typical k_f coefficient versus rear axle tire inflation pressure for several tire models of 2-axle truck (Δp_t - difference between inflation pressure of front and rear axle tires)

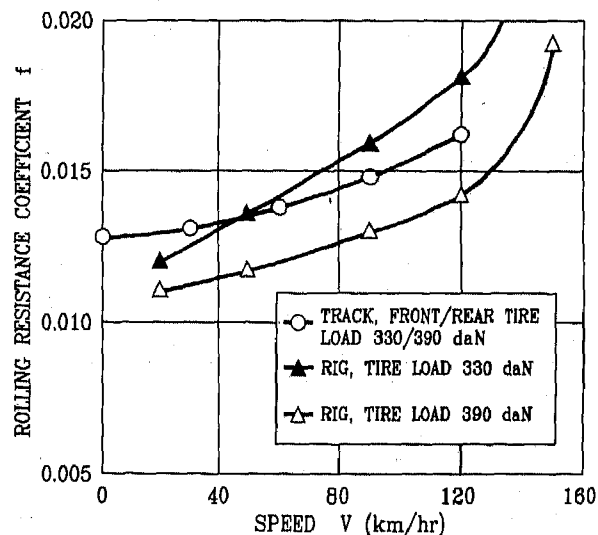


Figure 14: Rolling resistance coefficient of tire 165/180R13 versus speed on road and rig (wheel suspension system is the same for both cases)

oscillation amplitude and tire losses giving as a consequence a rise of sprung mass amplitudes as well as energy dissipation in shock absorbers. So, as opposed to the tire test data on the drum rig (with no vehicle suspension), energy dissipation for the tire rolling as a part of the tire-suspension-road system is expressed by coefficient k_f. It is of

importance in order to optimize parameters of the system. The tire of the given model may not be best among others being tested on the rig, but interacting with a suspension it can provide a significant reduction of rolling resistance in real road conditions in comparison with other models. Conducting the tire tests by the rig equipped with wheel suspension enhances the agreement between the bench and road results (see Fig.14).

DRIVELINE POWER LOSSES - The values that are reduced to the drive wheel circumference for direct summarizing with rolling resistance force are given in Table 7.

Table 7 : Driveline losses

Vehicle	Chassis type	Driveline losses (daN)	
		0 km/hr	90 (80)* km/hr
Passenger cars			
GAZ-3105	4x4	4.3	9.5
Peugeot-605 SR	4x2	0.3	5.3
Mercedes 260E	4x4	6.4	13.4
Ford Taurus	4x2	0.3	7.3
Renault-25 GTX	4x2	1.3	7.0
Ford Scorpio	4x4	3.6	11.7
GAZ-31029	4x2	1.8	5.5
Light-duty trucks and pickups			
GAZ-23042		1.3	6.7
GAZ-33022		0.5	3.8
GAZ-2705		2.4	5.3
GAZ-27052	4x2	0.8	6.4
GAZ-27053		2.5	5.4
Ford FT-190		1.2	7.1
Iveco 35.10-320		1.4	9.4
Medium-duty trucks			
ZIL-4331-10	4x2	3.3	25.5
Nissan L 80.14		2.5	10.2
Heavy-duty trucks			
Merc.-Benz 2238	6x4	2.4	66.7
Scania R142H		5.0	44.8
Buses			
ZIL-3207		4.7	14.2
LAZ-4207		4.7	15.6
Alterna 4215	4x2	4.3	18.5
Alterna 4216		5.1	15.1
LiAZ 5256**		6.7	67.0

Note : * For trucks;
** Gearbox with torque converter and disc clutch.

Considering these forces to be approximated by the linear dependence upon speed, two values at V=0 and V=90 or 80 km/hr are sufficient for its estimation.

The method allows altogether wide possibilities for studying relatively small resistance of this kind and also its dependence on oil temperature and running-in level of sealings and other driveline frictional pairs. In particular, in Figs. 15, 16 an estimation of the final drive temperature influence on the above mentioned losses is shown.

ROUTINE STUDY OF VEHICLE DRAG SHARES - The most typical opportunities and data obtained by the method developed during the studies of trucks and buses resistance parameters are given in Table 8.

Configurations 1,2 demonstrate 10% increase of passenger car c_x owing to 150 mm less overall length, unless tire profile width reduced.

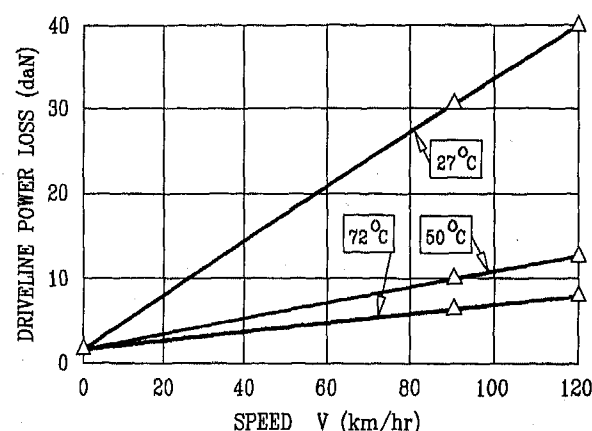


Figure 15: Driveline power loss versus speed (Volvo-760)

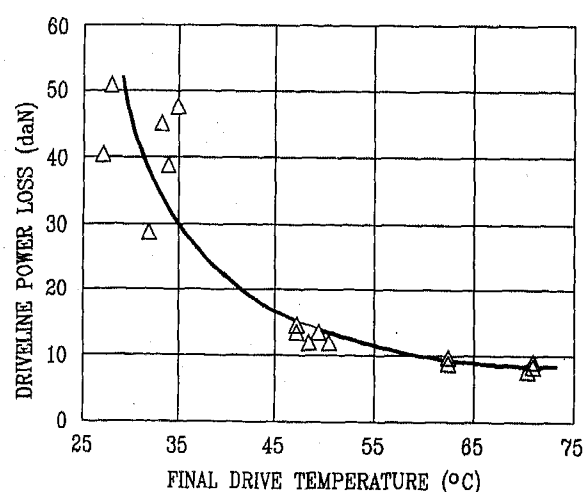


Figure 16: Driveline power loss versus oil temperature (Volvo-760 at speed tend to zero)

In configurations 3,4 an inadequate fairing design combined with pickup upper trunk results in 2% c_x rise.

Table 8 : Aerodynamic drag and rolling resistance parameters for different vehicle configurations

Vehicle (tire)	Parameters *
1. GAZ-3105 (205/60R15 Goodyear)	$c_x = 0.353$ (±3.9%) $c_x \cdot F = 0.62\ m^2$ $f_0 = 0.0114$ (±2.2%) $k_f = 7.0 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0158$ $t_a = 16.5\ ^oC$
2. GAZ-3104 (195/60R15 Michelin)	$c_x = 0.388$ (±5.1%) $c_x \cdot F = 0.81\ m^2$ $f_0 = 0.0127$ (±2.5%) $k_f = 4.7 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0156$ $t_a = 0.0\ ^oC$
3. GAZ-23042 "Burlack" (205/70R14 ИД-220)	$c_x = 0.558$ (±3.4%) $c_x \cdot F = 1.67\ m^2$ $f_0 = 0.0139$ (±2.8%) $k_f = 3.4 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0160$ $t_a = 3.0\ ^oC$
4. GAZ-23042 "Burlack" (205/70R14 ИД-220)	$c_x = 0.570$ (±6.4%) $c_x \cdot F = 1.73\ m^2$ $f_0 = 0.0137$ (±4.2%) $k_f = 2.9 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0155$ $t_a = 4.0\ ^oC$
5. GAZ-3302 (185/182R15 Я-288)	$c_x = 0.597$ (±3.4%) $c_x \cdot F = 2.15\ m^2$ $f_0 = 0.0119$ (±3.2%) $k_f = 3.8 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0143$ $t_a = 13.5\ ^oC$
6. GAZ-3302 (185/182R15 Я-288)	$c_x = 0.540$ (±2.6%) $c_x \cdot F = 2.67\ m^2$ $f_0 = 0.0113$ (±2.1%) $k_f = 5.2 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0146$ $t_a = 21.0\ ^oC$

Table 8 (Continued)

Vehicle (tire)	Parameters *
7. GAZ-33022 (175R16 K-135)	$c_x = 0.649$ (±2.2%) $c_x \cdot F = 3.57\ m^2$ $f_0 = 0.0092$ (±2.2%) $k_f = 5.0 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0123$ $t_a = 13.0\ ^oC$
8. GAZ-33022 (175R16 K-135)	$c_x = 0.540$ (±4.5%) $c_x \cdot F = 2.97\ m^2$ $f_0 = 0.0094$ (±3.8%) $k_f = 5.0 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0125$ $t_a = 13.0\ ^oC$
9. GAZ-33022 (175R16 K-135)	$c_x = 0.531$ (±2.8%) $c_x \cdot F = 2.92\ m^2$ $f_0 = 0.0108$ (±3.6%) $k_f = 4.9 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0139$ $t_a = -1.0\ ^oC$
10. GAZ-27052 (175R16 K-135)	$c_x = 0.503$ (±2.0) $c_x \cdot F = 2.22\ m^2$ $f_0 = 0.0104$ (±1.8) $k_f = 6.6 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0145$ $t_a = 14.0\ ^oC$
11. GAZ-27052 (175R16 K-135)	$c_x = 0.458$ (±3.7%) $c_x \cdot F = 1.84\ m^2$ $f_0 = 0.0108$ (±1.5) $k_f = 7.9 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0158$ $t_a = 8.0\ ^oC$
12. GAZ-27053 (175R16 K-135)	$c_x = 0.406$ (±2.7%) $c_x \cdot F = 2.07\ m^2$ $f_0 = 0.0105$ (±1.3%) $k_f = 5.0 \cdot 10^{-6}\ s^2/m^2$ $f_{90} = 0.0136$ $t_a = 15.0\ ^oC$

Table 8 (Continued)

Vehicle (tire)	Parameters *
13. Ford Transit FT-190 (185R14 Goodyear)	c_x = 0.449 (±2.4%) $c_x \cdot F$ = 1.80 m^2 f_0 =0.0089 (±5.1) $k_f = 6.2 \cdot 10^{-6}$ s^2/m^2 f_{90} = 0.0119 t_a = 14.0 °C
14. Iveco Turbo Daily 35.10-320 (175/75R16 Michelin)	c_x = 0.422 (±5.1%) $c_x \cdot F$ = 2.17 m^2 f_0 =0.0084 (±6.4%) $k_f = 3.7 \cdot 10^{-6}$ s^2/m^2 f_{90} = 0.0114 t_a = 16.0 °C
15. ZIL-4331-10 (260x508R ИН-142Б)	c_x = 0.674 (±4.9%) $c_x \cdot F$ = 3.51 m^2 f_0 =0.0092 (±2.7%) $k_f = 4.9 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0114 t_a = -3.0 °C
16. Nissan L 80.14 (205/75R17.5 Michelin)	c_x = 0.876 (±2.4%) $c_x \cdot F$ = 3.95 m^2 f_0 =0.0075 (±1.8%) $k_f = 4.6 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0098 t_a = 2.0 °C
17. ZIL-3207 "Younost" (250/70R16 И-283)	c_x = 0.631 (±6.2%) $c_x \cdot F$ = 2.54 m^2 f_0 =0.0082 (±4.9%) $k_f = 6.9 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0116 t_a = 1.0 °C
18. ZIL-32071 "Younost" (250/70R1216 И-283)	c_x = 0.682 (±5.9%) $c_x \cdot F$ = 2.86 m^2 f_0 =0.0108 (±4.9%) $k_f = 7.7 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0146 t_a = -1.0 °C
19. LAZ-4207 (11.00/70R22.5 Д-1М)	c_x = 0.717 (±4.2%) $c_x \cdot F$ = 5.28 m^2 f_0 =0.0060 (±3.8%) $k_f = 4.5 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0082 t_a = 19.0 °C
20. "Alterna" - 4216 (280-508R ОИ-73Б)	c_x = 0.870 (±5.0%) c_x = 0.798 ** $c_x \cdot F$ = 6.30 m^2 f_0 =0.0075 (±3.8%) $k_f = 5.7 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0103 t_a = 9.0 °C
21. "Alterna" - 4215 "Tourist" (280-508R ОИ-73Б)	c_x = 0.747 (±3.0%) c_x = 0.731 ** $c_x \cdot F$ = 5.79 m^2 f_0 =0.0079 (±3.3%) $k_f = 5.4 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0109 t_a = 5.0 °C
22. LiAZ-677 (280-508R И-73А)	c_x = 0.753 (±5.8%) $c_x \cdot F$ = 4.92 m^2 f_0 =0.0059 (±3.6%) $k_f = 5.6 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0087 t_a = 20.0 °C
23. LiAZ-5256 (11/70R22.5 Д-1М)	c_x = 0.662 (±3.4%) $c_x \cdot F$ = 4.50 m^2 f_0 =0.0055 (±5.5%) $k_f = 4.6 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0078 t_a = 24.0 °C

Table 8 (Continued)

Note : * The values in parentheses represents half of significant error at 0.95 confidence level.
** Shields around radiator of engine are installed; frontal area of air intake of radiator is decreased by 20 %.

In configurations 5,6 coefficient c_x is reduced through the tent mounting over the cargo-carrying platform of the 1.5 t truck as compared to the open unloaded platform, unless c_xF of the latter is of lesser value.

In configurations 7,8,9 coefficient c_x of the closed truck is successively reduced by fitting, firstly, the shield fairing on the cabin roof (see configuration 8) and then the three-dimensional ellipsoidal one.

In configuration 10 of a police bus a light-sound signal device width as broad as 0.85 bus overall width is installed on the bus roof, that results in 10% c_x rise as compared to configuration 11.

In configuration 12 the same bus as in configuration 10, but with the higher roof has 24% c_x and 7% c_xF fall, unless the larger frontal area. This fact allows to come to a very important conclusion that an increase of vehicle height usually brings to considerable c_x reduction. Due to the lack of space for the detailed analysis of this effect, a distant but clear analogy of the vertical wing near the road having less drag than the horizontal one may be used. Likewise this effect may be noticed while comparing configurations 13, 14 of the van-type trucks with high and low roofs, despite of worse aerodynamic outline of Iveco-Turbo-Daily cabin. The similar effect is demonstrated by comparison of two unificated buses (configurations 20, 21). Configuration 21 is a tourist version of configuration 20, being 7% higher (but only 1% longer) and, consequently, has 14% lesser c_x.

As a whole, these results may be considered as unpremeditated embodiment of the old Rumpler's concept [30] at the modern vehicle design stage. The vehicle he had built was relatively high, and its horizontal plan was of wing profile form. As Hugo reports [30], in 1979 Buchheim had obtained c_W=0.28 on the museum specimen of Rumpler's vehicle in a large wind tunnel.

From configurations 17, 18 of the comfortable 18-seater bus an effect caused by displacement of saloon ventilation system air intake from the area of the windshield upper edge (configuration 15) to the rear roof part can be estimated. During the tests the chokes of the air intakes were fully opened. In configuration 16 ventilation system efficiency increased for more than 2.5 times resulting in 8% c_x rise. As far as f_0=0.0146 concerned account must be taken that the tests of the bus with improved ventilation system were performed on the tires not fully rolled.

Recurring to unificated city and tourist buses (configuration 20, 21), c_x reduction owing to the reduced area of engine front radiator intake of the first model for 8% and of the second one for 2% is observed.

A variety of the method employment illustrations is continued in the Table 9 including the articulated trailer trains.

Trailer train with six-wheel traction truck Mercedes-Benz 2238 of configuration 1 indicates 28% difference of vehicle aerodynamic drag with and without aerodynamic fairing. In configuration 2 of the similar trailer train, but with traction truck Scania R142 H, this difference being somewhat lower is caused by the enlarged distance from cabin to container, despite of the three-part fairing set on the cabin. The semitrailers of both trailer trains were equipped with twin wheels. A significance of semitrailer single wheel effect on the frontal aerodynamic drag reduction is shown by the test results of the trailer train (see configuration 3-MAZ-64227), despite a simplified fairing on the cabin. It can be explained by a strong influence of turbulent wake behind the truck on the twin wheel axle having a larger frontal area. For single wheels a high turbulence stream flows into the space between semitrailer wheels with the less aerodynamic energy loss.

Finally, comparison of c_x values of a concept truck (configuration 4) and of the trailer train on its base (configuration 5) indicates that c_x reducing due to significant test object lengthening is unexpectedly not so large as against to the negative influence of friction resistance, as well as resistance of enhanced number of protruding parts, particularly in the underbody space, and also to the influence of trailer train sections on each other.

In Table 9 a significant rolling resistance reduction of large cross section single tires on the 22.5" rim (items 4, 5) as compared to the twin wheels (items 1-3) is shown.

COAST DOWN TESTS WITH RECORD INTERRUPTION - To assess the possibilities of the method application for trials on the straight-line measuring sections alternating with curved track segments (Fig.5a), or turns on the ends of the measuring sections where data record is interrupted, the tests were undertaken by employing the measurement schemes shown in Fig. 5b realized on one extended 1400 m length road segment. To provide the opportunity not only of calculating time T_c but also of its measuring the length of segment 4-4' was chosen successively equal to 0, 200 and 400 m (Table 10).

It is obvious from these results that straight test sections having length of 600-1000 m, or 2-2.5 times shorter than total coasting distance of the object examined may be used. Since distance and time of record break are enhanced, some deterioration of the data convergence may be expected. It is especially affected by significant

Table 9 : Aerodynamic drag and rolling resistance parameters for different train vehicle configuration

Trailer train (tires)	Parameters *
1. "Mercedes-Benz" 2238 with semitrailer "Fruehof" FB6F2-40 (11.00R20 Goodyear) (12.00R20 И-150N twin disc)	c_x = 0.718 (±5.0%) c_x = 1.001 ** $c_x \cdot F$ = 5.88 m^2 f_0 = 0.0063 (±1.4%) k_f = $2.0 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0073 t_a = 16.0 °C
2. "Scania" R142H with semitrailer MAZ-93891 (11.00R20 Michelin) (320-508 ИЯВ-12Б twin disc)	c_x = 0.729 (±4.2%) c_x = 0.854 ** $c_x \cdot F$ = 6.93 m^2 f_0 = 0.0076 (±1.8%) k_f = $4.4 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0098 t_a = 15.0 °C
3. MAZ-64227 with semitrailer MAZ-9398 (300-508R) (300-508R single disc)	c_x = 0.435 (±3.1%) $c_x \cdot F$ = 4.09 m^2 f_0 = 0.0081 (±2.6%) k_f = $4.7 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0104 t_a = 13.0 °C
4. NAMI-0286 ("Typhoon") (350/80R22.5 И-338 front, 420/80R572 И-325 rear - all single disc)	c_x = 0.347 (±2.4%) $c_x \cdot F$ = 3.31 m^2 f_0 = 0.0055 (±4.2%) k_f = $5.2 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0081 t_a = 5.5 °C
5. NAMI-0286 ("Typhoon") with semitrailer NAMI-0287 (350/80R22.5 И-338 front, 420/80R572 И-325 rear) (350/80R22.5 И-338)	c_x = 0.400 (±3.4%) $c_x \cdot F$ = 3.82 m^2 f_0 = 0.0062 (±2.7%) k_f = $5.2 \cdot 10^{-6}$ s^2/m^2 f_{80} = 0.0089 t_a = 6.2 °C

Note : * The values in parentheses represents half of significant error at 0.95 confidence level.
** Fairing from cabin roof is removed; f_{80} - rolling resistance coefficient at V = 80 km/hr .

Table 10 : Comparison between calculated and measured value of concordance time T_c (retrieval of 3 worst test data from 6 measurements for each vehicle)

Vehicle, weight (daN) ΔS_c* (m)	T_c values (s)			Max. error $\frac{\Delta T_c}{T_4 + T_c}$ ** %
	Calculated	Measured	Difference ΔT_c	
GAZ-2705	0.2698	0.0000	-0.2698	-0.29
G=2929	-0.2635	0.0000	-0.2635	-0.29
$\Delta S_c = 0$	-0.1365	0.0000	-0.1365	-0.15
GAZ-33022	14.0980	14.6763	-0.5783	-0.47
G=2909	15.4757	14.8554	0.6203	-0.50
ΔS_c=200	14.6319	15.1885	-0.5566	-0.45
GAZ-27052	25.6289	26.8851	-1.2562	-0.98
G=3194	25.2249	27.9237	-2.0988	-1.63
ΔS_c=400	25.6481	27.1511	-1.5030	-0.15

Note : * ΔS_c - distance of time record interruption;
** T_c and T_4 - see fig.5b.

speed variation on the record break sections which may cause variation of tire temperature and, consequently, of rolling resistance.

The most promising comparative data for a number of vehicles are represented in Table 11.

So, the future progress of the coast down techniques presented is associated with the reliability of the instruments for tire temperature control during vehicle motion.

VEHICLE WHEEL ROLLING RADIUS - As referred above it can be measured by the time interval meter with two sensors hooked up (see Fig.9). One of them responds to the road marks of the measuring section start and finish, and the other senses wheel marks to count revolution number including its fractions on both ends of measuring section. An employment of such measuring scheme indicated the maximum 0.05% rolling radius variation for the tire inflation pressure of trucks and buses higher than 0.55 MPa, in the speed range from 0 to 120 km/hr and tire temperature range 20-60^{0}C.

The testing of the vehicle of configuration 7 (see Table 8) has shown that for such conditions a scheme with wheel revolution number counter can be employed for coast down tests similarly the driveline testing. Thus, for express evaluation of drag components for the vehicles with tire pressure higher than 0.5 MPa the track marking may be not in use.

Table 11 : Comparison between drag coefficients obtained by measurements with interruption (Fig. 5b) and basic method

Vehicle model, G (daN), ρ (kg/m^3), t_a (°C)	Parameters	Results by methods:		Relative discrepancy (%)
		With interruption	Basic	
GAZ-2705	f_0	0.0150	0.0150	0.0
G=2929	$c \cdot 10^4$	0.4707	0.4675	0.9
ρ=1.258	δc %	2.2	1.7	
t_a= 1.0	$k_f \cdot 10^6$	3.98	4.04	0.0
	f_{120}	0.0194	0.0195	-0.5
	c_x	0.4901	0.4857	0.9
GAZ-33022	f_0	0.0110	0.0110	0.0
(fairing)	$c \cdot 10^4$	0.6494	0.6549	-0.8
G=3220	δc %	5.1	4.7	
ρ=1.304	$k_f \cdot 10^6$	5.25	5.18	1.4
t_a=-0.5	f_{120}	0.0168	0.0168	0.0
	c_x	0.5255	0.5310	-1.0
GAZ-33022	f_0	0.0108	0.0108	0.0
(fairing)	$c \cdot 10^4$	0.7151	0.7226	-1.0
G=2887	δc %	1.2	1.6	
ρ=1.306	$k_f \cdot 10^6$	4.82	4.88	-1.2
t_a=-1.0	f_{120}	0.0162	0.0162	0.0
	c_x	0.5255	0.5310	-1.0
GAZ-27052	f_0	0.0107	0.0107	0.0
G=2882	$c \cdot 10^4$	0.5241	0.5071	3.4
ρ=1.273	δc %	7.8	5.5	
t_a=9.0	$k_f \cdot 10^6$	5.26	4.87	8.0
	f_{120}	0.0165	0.0161	2.5
	c_x	0.5319	0.5171	2.9
GAZ-27052	f_0	0.0108	0.0107	0.9
G=3194	$c \cdot 10^4$	0.4695	0.4571	2.7
ρ=1.276	δc %	5.1	4.6	
t_a=8.0	$k_f \cdot 10^6$	4.31	4.25	1.4
	f_{120}	0.0159	0.0154	3.2
	c_x	0.5319	0.5171	2.9

Note : f_{120} - rolling resistance coefficient at speed 120 km/hr .

SUMMARY AND CONCLUSIONS

1.The method of integrating the vehicle coast down equation eliminates the traditional need for measuring or calculating vehicle velocity or deceleration to determine vehicle drag and rolling resistance. In particular, the explicit form of the speed-time and distance-time functions are derived with resistance parameters and their combinations in coefficient form. The equations are based on constant coordinates of separate points of the distance-time curve.

2.By excluding speed recording the coast down test procedure was simplified, and an acceptable sensitivity to small changes of aerodynamic vehicle configuration, and also of tire and transmission parameters, was achieved.

3.It has been demonstrated that parameter measurements can be reduced to time recording only owing to replacement of distance measuring by precision track marking. The research equipment is simplified, and achievement of measurement accuracy, reliability, and replication presented is obtained under 6 coast-downs only.

4.Aerodynamic drag and rolling resistance separation by 2- or 3-level vehicle weight varying being combined to distance-time dependencies and widely verified by the tests had shown its effectiveness for examining motion resistance of these kinds.

5.By means of the method developed a considerable quantity of test results was accumulated that allowed to set the stable correlation between coefficient c_W of full-size passenger cars, and also of the scale models of trucks and buses in wind tunnels, on the one hand, and c_x of the correspondent full-scale vehicles in road conditions, on the other.

6.The method is widely used for investigation and comparative testing of different models of passenger cars, buses, trucks, and articulated vehicles as well as for their updating and operational development.

ACKNOWLEDGEMENT

The author wishes to thank Chief Designer Board of GAZ corporation and Board of Directors of the Central Automotive Proving Ground (Dmitrov, Russia) for their co-operation, and also Mr. N.Hur and Mr. I.Ahn - from Korea Institute of Science and Technology for their interest to the author's investigation, especially for Mr. Hur's initiative and idea [27] of application of the method developed to the task of coast down testing on the short tracks.

REFERENCES

1. Hoerner S. The determination of aerodynamic resistance of vehicles from free motion method. VDI, 79 (1935), 1028-1033.
2. Фалькевич Б.С. Дорожные испытания автомобилей (Vehicle Road Testing). Гострансиздат. Москва-Ленинград. 1936.
3. White R.A. & Korst H.H., The Determination of Vehicle Drag Contributions from Coast-Down Tests. SAE 720099, 1972.
4. Bez U., Bestimmung des Luftwiderstandsbeiwertes bei Kraftfahrzeugen durch Auslaufversuch. ATZ, N 11, 1974, pp. 345-350.
5. Walston W.H. et al, Test Procedures for the Evaluation of Aerodynamic Drag on Full-Scale Vewhicles in Windy Environments. SAE 760106, 1976.
6. Buckley F.T.,Jr. et al, Analysis of Coast-Down to Assess Aerodynamic Drag Reduction on Full-Scale Tractor-Trailer Trucks in Windy Environments. SAE 760850, 1976.
7. Yasin T.P., The Analytical Basis of Automobile Coastdown Testing. SAE 780334, 1978.
8. Roussillon G., Contribution to Accurate Measurement of Aerodynamic Drag on a Moving Vehicle from Coast-Down Tests and Determination of Actual Rolling Resistance. Proceedings of 4th Colloquium on Industrial Aerodynamics, Aachen, 18-20 June 1980.
9. Петрушов В.А. Новый метод определения сопротивления движению автомобиля (New Method of Vehicle Drag Determination). "Автомобильный транспорт", 1982, No.11.
10. Staska G., Quantifying Resistance in Road Tests, Automobiltechnische Zeitschrift, 86, 1984.
11. Петрушов В.А. Способ определения сопротивления движению колесного транспортного средства (Method of Wheeled Vehicle Drag Determination). Patent of Russia No.1150512, 1985.
12. Петрушов В.А. Оценка аэродинамических качеств и сопротивлений качению автомобиля в дорожных условиях (Estimation of Vehicle Aerodynamic Efficiency and Rolling Resistance in Road Conditions). "Автомобильная промышленность", 1985, No.11.
13. Петрушов В.А. Решение задачи интегрирования уравнения затухающего движения автомобиля в переменных путь-время и ее практические приложения (Solution of Problem of Integrating Vehicle Coasting Equation in Time-Distance Variables). Труды НАМИ. 1986.
14. Lucas G.G. & Emtage A.L., A New Look at the Analysis of Coast-Down Test Results, Proc Inst of Mech Engineers, Vol.201, No.D2, 1987.
15. Петрушов В.А. Способ определения аэродинамического сопротивления и сопротивления качению движителя колесного транспортного средства (Method of Determination of Vehicle Aerodynamic Drag and Rolling Resistance of Wheeled Propulsor). Patent of Russia No.1386862, 1988.
16. Eaker G.W., Wind Tunnel-to-Road Aerodynamic Drag Correlation. SAE 880250, 1988.

17. Remenda B.A.P. et al, Vehicle Coastdown Resistance Analysis under Windy and Grade-Variable Conditions, SAE 890371, 1989.
18. Петрушов В.А. и др. Автомобиль в аэродинамической трубе и на дороге (Vehicle in Wind Tunnel and on the Road). "Автомобильная промышленность", 1990, No.1.
19. Petrushov V.A. Vehicle C_x Measured on the Road. Avtoexport round up. 1990, No.4.
20. Passmore M.A. & Jenkins E.G., Measuring Vehicle Drag Forces Using an On-Board Microcomputer, Proc Inst of Mech Engineers, DO3389 Vol.204, 1990.
21. Korst H.H. & White R.A., Coastdown Tests: Determining Road Loads Vedsus Drag Component Evaluation, SAE 901767, 1990.
22. E.D.-M.P.O., Cosi le Nostre Prove, Quattroruote, No.424, 1991.
23. Passmore M.A. & Le Good G.M., A Detailed Drag Study Using the Coastdown Method, SAE 940420, 1994.
24. Buckley F.T.,Jr., ABCD - An Improved Coast Down Test and Analysis Method, SAE 950626, 1995.
25. Le Good G.M. et al, On-Road Aerodynamic Drag Measurements Compared with Wind Tunnel Data, SAE 950627, 1995.
26. Petrushov V.A., N.G.Hur, I.K.Ahn, High Accuracy Coastdown Test Method by Distance-Time Measurement: I.Theoretical Background and Discussions on Accuracy Improvements. Korea Society of Automotive Engineers, Inc. Paper No. 953719, 1995.
27. Petrushov V.A., N.G.Hur, I.K.Ahn, High Accuracy Coastdown Test Method by Distance-Time Measurement: II.Development of Short Distance Method and its Evaluation. Korea Society of Automotive Engineers, Inc. Paper No. 953730, 1995.
28. Cogotti A. et al, Comparison Tests between Some Full-Scale European Automotive Wind Tunnels - Pininfarina Reference Car. SAE 800139, 1980.
29. Buchheim E. et al, Comparison Tests between Major European Automotive Wind Tunnels. SAE 800140, 1980.
30. Hucho W. H. (Herausgeber) Aerodynamik des Automobils. Vogel-Verlag, Wursburg, 1981.

970451

PAC-ITS: Towards AHS and 150 mph Highways

Jim Haugen, Sr.
Haugen Associates

ABSTRACT

A unique approach to intercity highway transportation entitled PAC-ITS: Packet Autopiloted Cruiseway - Intelligent Transportation System, has been detailed in a recently completed contract for the National Automated Highway System Consortium (NAHSC). The PAC-ITS concept focuses on: (1) improving the productivity of intercity travel (2) facilitating increased highway speeds in a safe and environmentally permissible manner; and (3) minimizing automation complexities. Key concept features include:

- Use of PAC-ITS convoys which are composed of a mix of 15 or 20 vehicles - personal cars, buses and freight units - electronically and mechanically linked together.
- Interconnection of all vehicle power trains and brakes enabling unified convoy acceleration and braking.
- Individual active lateral guidance for each convoy vehicle
- Professional "pilot" control of each convoy from a specially equipped lead vehicle.

This paper outlines the functional features of PAC-ITS; describes the rationale behind the concept; and structures a three phase evolutionary development and deployment scenario leading to 150 mph operations.

This paper draws upon the author's years of involvement in strategic planning and R&D of advanced transportation systems at U.S auto companies, as well as his more recent consulting work in Intelligent Transportation Systems.

INTRODUCTION

- Must U.S. highway and vehicle progress permanently plateau within the constraints of the 70 or 75 mph Interstate Highway System?
- Shouldn't R&D progress continue to upgrade or supplant the Interstate?

The PAC-ITS concept provides a basis for initiating a dialogue on these issues; especially important as we mark the beginning of the second 100 years of the U.S. auto industry.

Construction of the Interstate highway network was a major milestone in highway transportation. Its many benefits are best summarized in two words - improved productivity. PAC-ITS seeks to continue to extend highway productivity through the use of emerging technology capabilities. Furthermore, the introduction of such a system is expected ignite a new era of transportation innovation and growth - in infrastructure construction; control and communications development; higher performance vehicle engineering; and business and organizational change.

CONCEPT DESCRIPTION

PAC-ITS concept basics are simple:

- Temporarily link a group of 15 or 20 vehicles together for highway trips, thereby saving energy through train type aerodynamics. Operate such convoys on a single dedicated lane at constant speed to simplify operations and minimize construction costs.
- Establish standards for performance and physical design constraints for all convoy vehicles - cars, trucks and buses.
- Use advanced electronic controls to link all the power trains; guide all the vehicles in the same path; convey individual vehicle status through diagnostics; and keep an eye on all the convoys on the highway.
- Place a single trained pilot up front in a special vehicle to lead the convoy; maximize his productivity and salary through the number and value of trailing vehicles; and provide him with adequate sensing and monitoring devices.
- Design this advanced system so it can be initially operated on the Interstate freeways and also evolve into a future nation-wide high-speed (possibly 150 mph) network.

PAC-ITS can be viewed as equivalent to a train, but with rubber tire advantages: the system is dual-mode, so all PAC-ITS vehicles can also be conventionally driven on all existing roadways; flexible convoys of 20 or so highway vehicles can

be assembled on a demand basis; departures and convoy spacing can be at intervals of less than a minute.

The PAC-ITS concept is illustrated in Figures 1 through 3.

- Figure 1 shows a close-up of three of the PAC-ITS vehicles in a convoy, illustrating:
  - -- The unique Pilot vehicle which, in this case, is also carrying passengers. The pilot sits up high for best visibility and is assisted by (1) a laser based Adaptive Cruise Control system; and (2) a laser beam or machine vision system controlling lateral position with respect to roadway markers.
  - -- A four wheel drive vehicle and a sports sedan linked up behind the pilot vehicle, illustrating the versatility of the concept to allowing mixing of vehicle types.

- Figure 2 is similar to Figure 1, only projected forward some 15 years. PAC-ITS convoys are now operating on an exclusive lane at speeds in excess of 100 mph. Features shown are:
  - -- Translucent flexible adapter pieces unfolded between each pair of vehicles for improved high speed aerodynamics.
  - -- The use of traffic barriers separating the dedicated lane from any manually driven traffic.

- Figure 3 illustrates three different convoy arrangements as compared to a double bottom semi configuration: (1) a Phase I convoy where all the freight vehicles have driver positions; (2) Phase II freight vehicles with no driver positions, assuming remote control of these vehicles at the freight center; and (3) a string of minivan size passenger vehicles behind the larger lead vehicle.

PAC-ITS OPERATIONS

Figure 4 illustrates a three step roadway entry process for PAC-ITS vehicles. In step 1, vehicles enter a queuing area from an arterial roadway, selecting the appropriate queuing lane for a designated convoy destination. In step 2, vehicles from one of these lanes move forward to link up with a lead vehicle to form a convoy. Finally, in step 3, the convoy receives a go ahead signal, accelerates and merges into an open slot on the dedicated freeway lane. The acceleration and merge operations are expected to take place under the manual control of the convoy pilot, except for the steering which is automatic. Once on the dedicated lane the pilot's role becomes supervisory only, monitoring convoy headway, automatic lane keeping, potential roadway obstacles, weather conditions and individual vehicle diagnostics.

PAC-ITS R&D NEEDS

Certainly any new system such as PAC-ITS requires extensive R&D and design efforts. Examples of some of the important areas include:

<u>VEHICLE TECHNOLOGY</u>

- Lateral guidance system - Electronic sensing based, such as lasers or machine vision; or mechanical based, such as the use of fold-out guidance arms to provide positive vehicle roadway containment.
- Vehicle linking - Development of low speed automatic coupling; coupling tie-in to the vehicle structure; throttle and brake interconnection and minimum accel/decel performance standardization.
- Vehicle design changes - Intake and exhaust re-routing; electronic steering and braking; remote diagnostics.

<u>ROADWAY TECHNOLOGY</u>

- Cooperative vehicle guidance - Addition of guidance system targets or guidance arm rails.
- Road construction - Smooth concrete construction for 150 mph operation, or use of plastic "smoothing" track inserts (since all vehicles follow the same path).
- Barrier design - Use of barriers to exclude manual vehicles and animals; to reduce side wind gusts; permit entry of emergency service vehicles.

<u>CONTROLS AND COMMUNICATIONS</u>

- Vehicle controls - Automatic steering; vehicle and convoy adaptive cruise control; remote maneuvering and automatic parking (especially for freight vehicles); automatic user "fare" payment.
- Convoy controls - Slot reservation techniques for ease of merging; convoy tracking and management; entry/exit management.

AHS RELATIONSHIP

The general concept of the automated highway system is often referred to as 'hands-off", "feet-off" driving on exclusive guideway lanes, with emphasis on solving urban area congestion. However, this concept has important weaknesses:

- Physical constraints, high infrastructure costs and marginal benefits make its deployment in urban areas doubtful
- A driver's desired "eyes-off" or :brain-off" function is doubtful because of emergency intervention concerns with very short headway operations
- Driver trust in total electronic roadway and traffic monitoring and control is yet to be proven
- The need for increasing headway gaps as speeds are raised preclude good aerodynamics and fuel consumption at high-speeds.

PAC-ITS CONCEPT RATIONALE

The PAC-ITS concept focuses on improving the productivity of interstate highway travel through high speed operations with mixed vehicle convoys. The rationale behind this approach is as follows:

- SYSTEM DESIGN SIMPLICITY - The PAC-ITS convoy approach simplifies many facets of advanced highway system design:
  - -- Minimizes complexity - single lane operations; no lane change maneuvers; all vehicles at a constant speed with no passing
  - -- Vehicle linking provides inherent system redundancy - if, for example, one vehicle's engine fails, the convoy tows the failed vehicle to the next exit
  - -- Linking the power trains of 20 vehicles can have large traction benefits, allowing safe operations in deteriorated weather conditions

- PROFESSIONAL PILOT VEHICLE CONTROL - Each freeway train will have a trained professional pilot in the lead vehicle, with many advantages:
  - -- History has shown that is impossible to design a perfect automatic system; human supervision and intervention, in systems ranging from subways to airplanes to spacecraft, is always needed for unforeseen problems.
  - -- Use of a human pilot minimizes the cost and complexity of sensor and system electronics additions to individual vehicles
  - -- People already have confidence in the use of trained pilots from airline experience
  - -- The trained pilot can monitor the diagnostics of all convoy vehicles for failures and apply human judgment to take appropriate action

- USER FRIENDLINESS AND CONFIDENCE - The PAC-ITS concept emphasizes a user orientation, supplemented by technology:
  - -- PAC-ITS use is intended to be simple - personal cars do not have to be turned into the equivalent of jet aircraft cockpits; no user special training required
  - -- Hands-off, feet-off, and eyes-off trip convenience
  - -- Executive jet equivalency in comfort and safety is expected as new personal vehicle designs emerge
  - -- The use of a human pilot achieves "go to sleep" confidence; frees drivers for on-road productive activities
  - -- A relative public feeling of safety is natural, given cultural familiarity with mechanically linked trains

- COMMERCIAL VEHICLE IMPROVEMENTS - The use of PAC-ITS "freeway trains" has special advantages for commercial vehicle users:
  - -- "Slipstreaming" will reduce overall drag by a significant amount, lowering operating costs
  - -- Productivity will be improve greatly since one driver can handle a payload as much as 10 times as large or valuable, with even better on-time delivery
  - -- As safety benefits are demonstrated, speeds can be increased, further increasing productivity
  - -- The public image of "killer trucks" can be radically improved by a combination of lower profile, aerodynamic trailers; precise lateral guidance eliminating trailer sway
  - -- Reduces commercial vehicle turnover problems and provides new higher paying jobs - PAC-ITS pilots instead of truck drivers

- POTENTIAL FOR HIGH SPEED OPERATIONS - PAC-ITS makes future highway travel at higher speeds a practical possibility:
  - -- PAC-ITS trains inherently minimize vehicle drag - drag is highway vehicles' biggest enemy; expensive in fuel consumption and emissions; a major factor in preventing the move to higher freeway speeds
  - -- The mechanically fixed vehicle spacing allows transition shrouds between vehicles, further enhancing high speed aerodynamics
  - -- Increasing speed significantly, as long as safety also improves, reaps major highway travel benefits:
    1) Car and bus competition with air travel
    2) Increased rural travel range
    3) More life style options
    4) Much higher freight hauling productivity

EVOLUTION PHASES

PAC-ITS Intercity deployment might occur over three phases, as postulated in the Deployment Schedule of Figure 5.

- Phase I - Deployment of limited size PAC-ITS packet convoys on Interstate freeways in mixed traffic.
  - -- As shown in the figure, first operations in mixed traffic, on a limited scale, are assumed to begin within 11 years - in January, 2008.
  - -- Mixed traffic operations would then be extended to other regions over the following decades as interest grows.

- Phase II - Interstate deployment expands; adding exclusive PAC-ITS lanes in some locations; adding low speed off-guideway maneuvering:
  - -- Automated Convoy Formation - This feature allows drivers to switch to automatic control upon entering the convoy formation area. The driver's vehicle is then automatically routed to the correct lane and automatically docked to other vehicles in the convoy.
  - -- Automated Parking and Yard Operations - This feature allows PAC-ITS equipped vehicles to be automatically routed and parked without a driver on board. It allows automatic storage of vehicles in parking areas or decks; remotely controlled movement of freight vehicles within a freight terminal.

- Phase III - PAC-ITS deployment on separate right-of-way with staged evolution to higher speeds.
  - -- Research and Development on increasing PAC-ITS speeds continues through a span of several decades, as shown.
  - -- Approval and funding - public and/or private - for construction of a national PAC-ITS type AHS network is shown occurring in the year 2014. This timing assumes expected positive reactions to initial Interstate based operations, combined with demonstrations of high speed operations feasibility. Within four years of approval, the first link begins 100 mph operations.
  - -- PAC-ITS operating speeds are shown being increased at a programmed rate of 10 mph every three years, based on meeting safety, energy consumption and emissions targets. This planned evolution of operating speed yields a top speed of 150 mph by the year 2033, some 25 years after the initiation of the first PAC-ITS operations. It also gives vehicle manufacturers known time frames in which to schedule corresponding higher performance vehicle introductions.

CONCLUSIONS

The National Automated Highway System program has, thus far, focused on the possibility of alleviating urban congestion using a total electronic automation approach. The PAC-ITS concept offers a different direction - a focus on improving intercity transportation productivity and the use of an Autopiloted approach which greatly simplifies system design and provides users with the confidence of a pilot in charge.

Electronics plays an assist and support role, similar to its proven role in airline operations.

The PAC-ITS concept raises important issues which should provide the basis for a national dialogue regarding continuing Interstate system development. Important questions in this dialogue are:

- Should the U.S. have an Interstate II design, development and deployment program over the next 40 or 50 years?
- Will AHS with a total automation approach ever be feasible and capable of achieving user confidence?
- What relative priorities should be placed on urban congestion alleviation vs intercity productivity improvement?
- What are the relative costs, benefits, feasibility and risks?
- Should we endeavor to push the speed envelope of cars and trucks, achieving requisite service gains, by taking advantage of advancing technology to do so, while maintaining environmental sensitivity?

REFERENCES

(1) Haugen, Jim, "PAC-ITS: Packet Autopiloted Cruiseway - Intelligent Transportation System", Study for the National Automated Highway System Consortium, January 11, 1996.



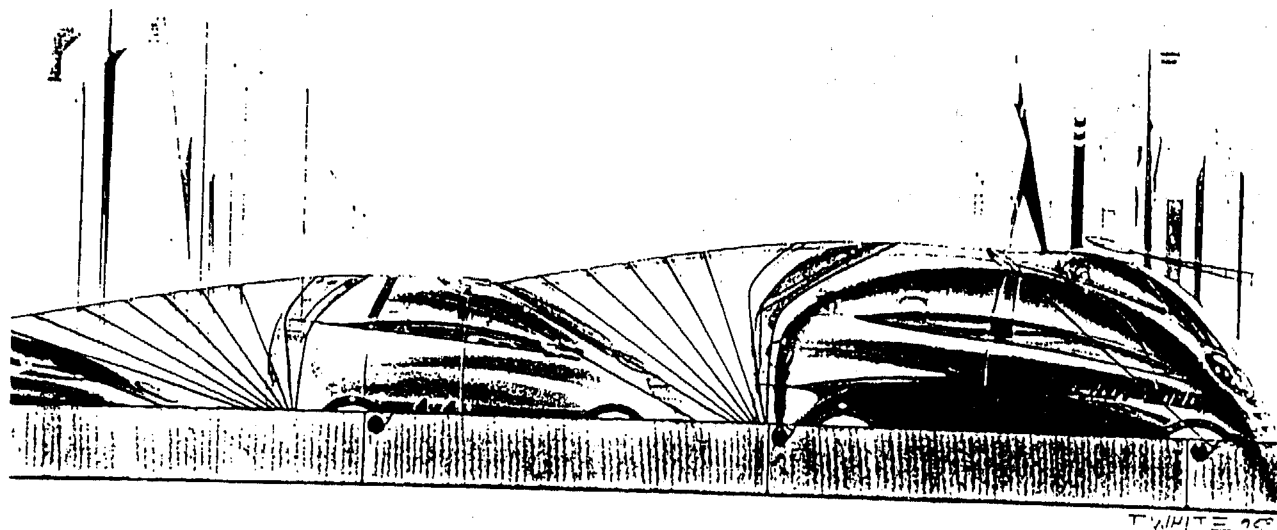


Figure 2: PAC-ITS PHASE 2 CONVOY



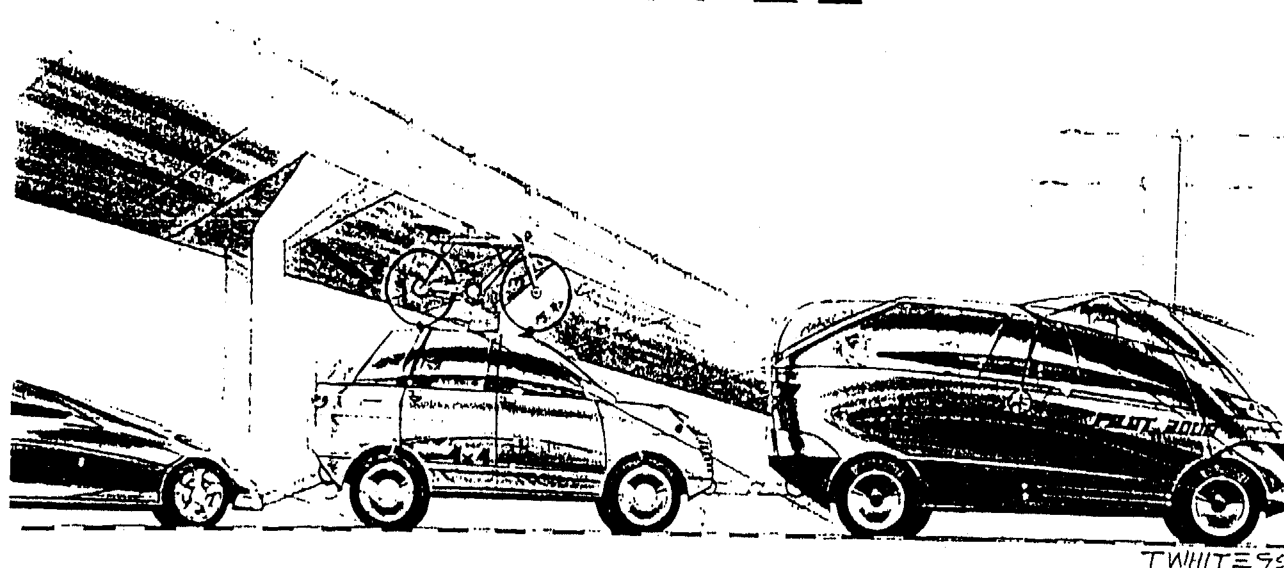
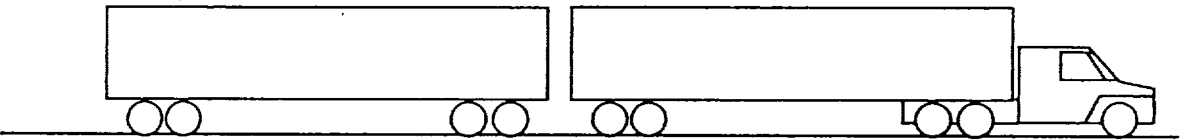
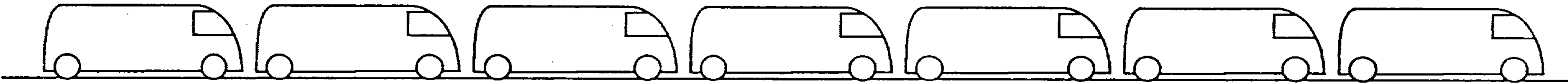


Figure 1: PAC-ITS PHASE 1 CONVOY

Figure 3: PAC-ITS CONVOY COMPARISONS



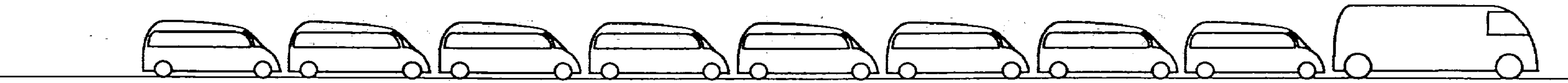
110 FOOT DOUBLE BOTTOM SEMI



PAC-ITS TRAIN: PHASE I VEHICLES WITH DRIVER POSITIONS

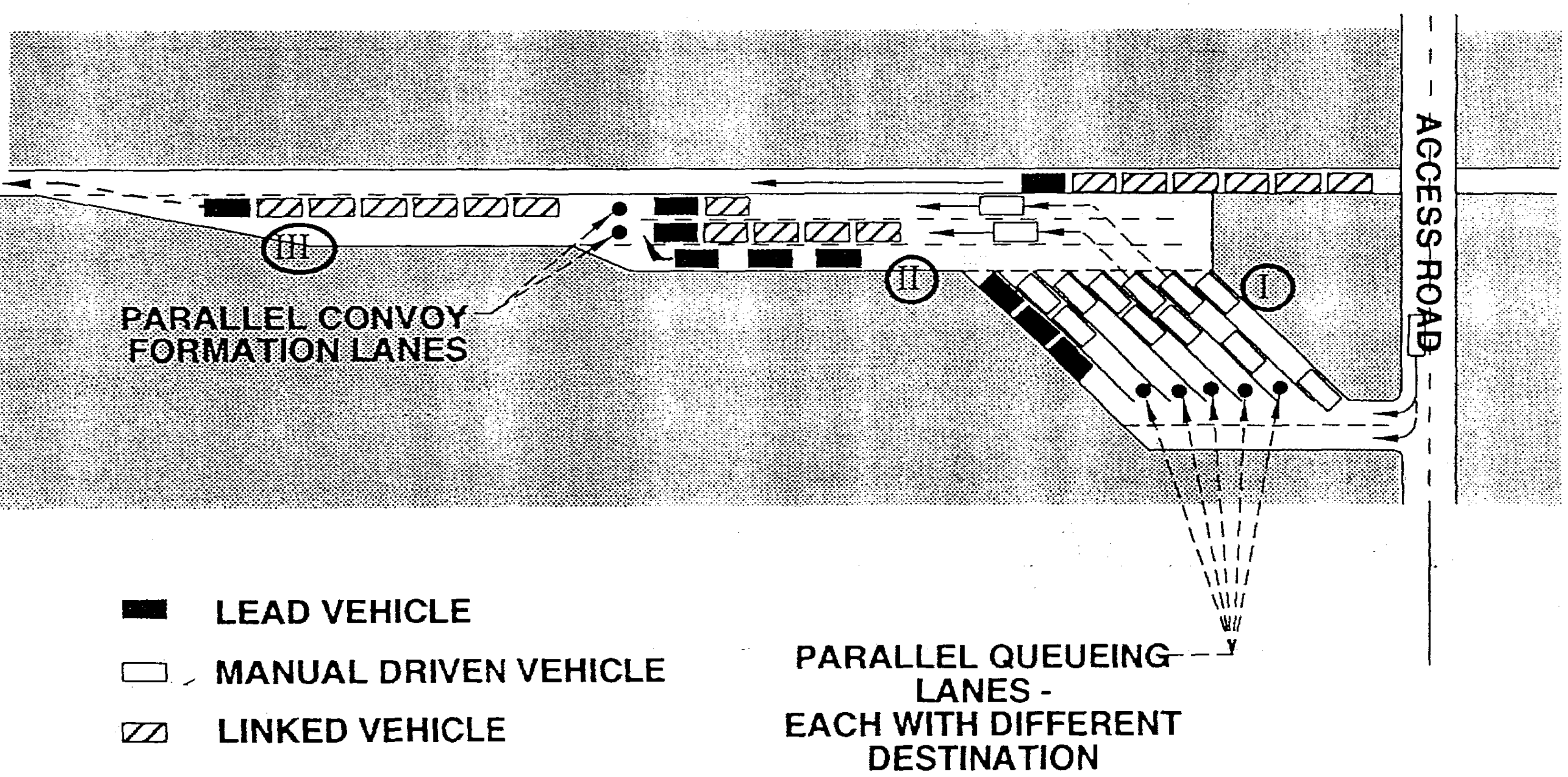


PAC-ITS TRAIN: PHASE II VEHICLES WITH NO DRIVER POSITIONS



MINIVAN PAC-ITS TRAIN

Figure 4: PAC-ITS ENTRANCE CONFIGURATION



ID	Name
1	**PHASE I INTERCITY DEPLOYMENT**
2	INITIAL R&D
3	DEMOS
4	MIXED TRAFFIC FIELD TESTS
5	CONSTRUCT ACCESS FACILITIES
6	DESIGN PRODUCTION VEHICLES
7	PHASE I OPERATIONS
8	**PHASE II INTERCITY DEPLOYMENT**
9	BUILD EXCLUSIVE LANES
10	PHASE II OPERATIONS
11	**AUTO. TRAIN FORMATION**
12	R&D
13	DEMOS
14	FIELD TESTS
15	OPERATIONAL PRODUCTION VEHICLES
16	**AUTO. PARKING & YARD OPS.**
17	R&D
18	DEMOS
19	FIELD TESTS
20	CONSTRUCT FACILITIES
21	DESIGN PRODUCTION VEHICLES
22	BEGIN OPERATIONS
23	**PHASE III INTERCITY DEPLOYMENT**
24	HIGH SPEED R&D
25	100-150 MPH DEMOS
26	PAC-ITS H.S. NETWORK START
27	CONSTRUCTION & FIELD TESTS
28	BEGIN 100 MPH OPERATIONS
29	BEGIN 110 MPH OPS.
30	BEGIN 120 MPH OPS.
31	BEGIN 130 MPH OPS.
32	BEGIN 140 MPH OPERATIONS
33	BEGIN 150 MPH OPERATIONS

Figure 5: PAC-ITS PHASES AND TIMING

970455

Analytical Methodology for Design and Performance Assessment of Run-Off-Road Collision Avoidance Systems

V. K. Narendran, D. B. Pape, J. A. Hadden, and J. H. Everson
Battelle

D. A. Pomerleau
Carnegie Mellon Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Research in Intelligent Transportation Systems (ITS) has been increasingly focussed on the development of Collision Avoidance Systems (CAS). A CAS would reduce the incidence of collisions by providing warnings to the driver to take evasive action. Because single vehicle roadway departures, also known as Run-off-Road (ROR) events, are a cause of a significant portion of vehicle accidents and fatalities, an effective CAS for ROR can potentially improve highway safety dramatically.

The development of performance specifications for CAS for ROR events is a part of an ongoing three-phase program for NHTSA (National Highway Traffic Safety Administration). This paper focusses on the development and application of a powerful simulation tool, RORSIM, for CAS assessments over a wide range of environmental, roadway, driver, vehicle and CAS operating conditions. The results of CAS effectiveness studies are presented. These results indicate that such systems can improve highway safety significantly by providing consistent and timely warnings to the driver.

INTRODUCTION

There are existing prototypes of ROR CAS such as the RALPH system, developed by Carnegie Mellon University, [Pomerleau, 1995], and the CAPC [Ervin, 1995] system developed by University of Michigan Transportation Research Institute, which use a time-to-line crossing (TLC) algorithm to warn drivers of incipient ROR events. A dynamic simulation capability is a powerful supplement to on road and test track evaluation with such systems, which can be time-consuming, costly, and limited in the range of conditions that can be assessed.

The goal of the ROR program is to develop a simulation tool to establish normal driving scenarios and study the effectiveness of a ROR CAS when brief periods of driver inaction are imposed during the normal driving. In the first phase of study of CAS for ROR events [Pape et. al., 1996], an analytical approach was developed specifically for evaluating CAS effectiveness in preventing ROR incidents. To meet the need for the ability to evaluate large numbers of driving scenarios and the influence of variations in several vehicle, driver and CAS parameters, the analytical approach was expanded. Distributions were defined for a wide range of input parameters. Effectiveness criteria were defined for the CAS and, in order to allow the user to study CAS performance over a wide range of scenarios, a "Monte Carlo" like technique was introduced. Results of simulation studies involving tens of thousands of cases are presented in this paper.

ANALYTICAL METHOD

This simulation tool was implemented in the software package RORSIM. RORSIM is a menu-driven time-domain simulation program that predicts driver/vehicle/CAS dynamic interaction. It is an enhancement to VDANL (Vehicle Dynamic Analysis, Non-Linear), which is a general-purpose rubber-tired vehicle simulation program developed for NHTSA by Systems Technology, Inc. in Hawthorne, California [Allen et. al. 1992]. VDANL provides the basic vehicle dynamics model for the simulation as well as the closed-loop driver model. VDANL includes a 17-degree-of-freedom model of a general vehicle.

In addition to representation of the dynamics of vehicle motion, the model also captures the realistic dynamic interactions of vehicle-road, driver-vehicle, and CAS-driver interfaces. An explicit driver model takes into account the delays associated with the different response modes of a human driver such as driver reaction times to braking and steering commands. Nominal driver lane-keeping behavior, aggressiveness of response to warnings issued by a CAS, and brief periods of driver inaction also have been modeled.

The CAS includes a sensor system that acquires the data required by the decision module of the system. The decision module then uses these data with available vehicle state information to determine whether the driver requires a warning. Several CAS methodologies have been proposed for studying ROR events. A detailed description of these methods and their benefits was included in [Pape et. al., 1995]. In the present study only the lateral, downward-looking system (TLC) was used. However, the methodology developed could be applied directly to evaluate other CAS methodologies.

The RORSIM package can simulate a complete driving scenario: a situation develops, it is sensed by the CAS, the driver responds to the warning and regains safe control of the vehicle. When applied in this manner, RORSIM is useful for demonstrating that a CAS can successfully prevent a ROR accident for a given driving situation.

MONTE CARLO SIMULATION TECHNIQUE - The scope of the RORSIM simulation studies was to conduct extensive parameter studies using a Monte Carlo like approach. This provided a more comprehensive characterization of the effectiveness of CAS schemes in incipient ROR events.

A statistical approach was used in which all the key parameters were varied simultaneously according to a scheme that ensured good coverage of the range and distribution of each parameter value. A scheme known as the "Latin hypercube" approach [Stein, 1987 and McKay et.al. 1979], essentially similar to but more efficient than a pure Monte Carlo approach, was used to generate the input variables for the simulation runs. The distributions for defining the various input parameters have been outlined in this section.

Latin hypercube sampling is a simulation procedure which is an appealing alternative to conventional Monte Carlo methods because it generally produces estimates with a lower variance than simple random sampling of input vectors. Latin hypercube sampling stratifies each marginal distribution of the input vector as much as possible but otherwise picks the vectors randomly. Approximately six hundred random input vectors, (i.e., values for vehicle speed, lane keeping performance amplitude, driver reaction time etc.), were generated using a series of Latin hypercube samples.

PARAMETER DISTRIBUTIONS - Table 1 lists the input parameters that were varied randomly by the Latin hypercube method. The range of values considered and corresponding distributions selected for each parameter have also been included. The distributions were selected based on reviews of existing literature, and standard highway design handbooks. A brief discussion on the selection of distributions for some of the key input parameters is presented below.

Table 1. Distributions of parameters for the Latin hypercube simulation studies

Parameter	Range of Values	Distribution
Roadway Curvature	250 ft, 500 ft, 1000 ft, 2000 ft, straight	Uniform
Lane Width	12 ft	Fixed
Road Friction	0-1	Beta
Shoulder Friction	0-1	Beta
Shoulder Rolling Resistance	0.015-0.9	Log normal with an offset
Vehicle Speed	(30-110) feet/second	Normal
Driver Lane-Keeping Performance	0.1-0.35	Uniform
Driver Reaction Time	0.01-2.0 seconds	Log Normal
Driver Aggresiveness of Response	0.4-2.5	Uniform
Initial Time of DriverInattentiveness	2-17 seconds	Uniform
Duration of Driver Inattentiveness	2.2 - 13.2 seconds	Uniform
TLC Threshold	-0.6, 0.0, 0.6, 1.2, 1.8 seconds	Fixed for each run
CAS Accuracy	0.0, +/-0.25, +/-0.5, +/-1.0	Uniform

<u>Roadway Design and Vehicle Speed</u> - Nine different roadways were adopted for these studies. These included a straight road and roads with left-hand and right-hand curvatures of 250 ft, 500 ft, 1,000 ft and

2,000 ft. The curved road consisted of a straight section followed by a curved section with entry and exit spirals. Vehicle speeds for each road design were normally distributed with a standard distribution of 3 fps. The mean speed on the tighter curves (250 and 500 ft radius) was 10 percent above the rated speed, determined using AASHTO guidelines, [AASHTO, 1994]. The mean speeds on the remaining curves were selected to provide reasonably uniform coverage over the entire range of speeds.

Driver Lane Keeping Performance - This feature was modeled by applying a randomly-varying lateral wind force. The vehicle path approximates the paths followed by drivers in real vehicles. Higher amplitudes were used to model drivers who, for one reason or other, do not follow as straight a path as others.

Driver Aggressiveness of Response to Alarm - This driver is characteristic was modeled by changing the gain in the steering feedback loop after the alarm. A high multiplicative factor corresponds to a driver who panics when an alarm is issued and over corrects. A lower factor represents a driver who responds more sluggishly to an alarm. This effect was applied in addition to the steering reaction time which is a delay between the alarm and the driver's beginning of response.

For each curved roadway, assuming a constant speed of travel, the time of onset of driver inattentiveness and the duration of inattentiveness were selected to ensure that a ROR event can occur in any portion of the roadway with uniform probability. This allowed for the possibility of a ROR event on entry and exit spirals as well as curved sections of roadways.

EFFECTIVENESS MEASURES

The primary function of RORSIM for these studies was to estimate CAS effectiveness for a wide range of driving scenarios. To serve this function a suite of normal driving cases was selected. In each case driver inattention was imposed to generate potential run-off-the-road events. The effectiveness of the CAS was evaluated by studying the same scenarios with and without CAS protection.

RUN-OFF-ROAD CRITERIA - A key issue in the development of an effective ROR countermeasure system that is not overly restrictive or annoying is the establishment of limits of acceptable vehicle excursions. Perhaps the least ambiguous but most restrictive ROR criterion is any situation in which at least one tire crosses over the lane edge. This criterion, which for the purposes of this study is defined as the "1-Tire ROR Criterion" implies that it is unsafe for the vehicle to use any part of the shoulder or adjoining lane during normal driving or for maneuvering to avoid a potential accident. This criterion is particularly relevant to situations where no shoulder exists (e.g., under bridges or through construction zones).

A more forgiving ROR criterion is one that allows excursions of the vehicle onto the shoulder (e.g., "lane-cutting" during curving). This "2-Tire ROR Criterion" defines a ROR as an event in which both front tires are outside the original lane of travel and permits more vehicle lateral excursion before the result is considered to be an ROR event.

PERFORMANCE CRITERIA - A CAS must be extremely reliable and at the same time must not be a nuisance to the driver. The performance criteria for the CAS address these issues. Post processing of the data obtained from the Latin hypercube simulation runs reveals a wealth of information. This information can be tested against the performance criteria defined in this section to evaluate the overall effectiveness of the system.

Each pair of simulations (one situation with and without CAS assistance) was categorized as one of six outcomes to determine the effectiveness of the CAS system:

Safe Correct Detection (SCD) - The system issued a warning of an incipient ROR situation. In response to the warning, the driver maneuvered the vehicle, avoiding a ROR situation.

Late Correct Detection (LCD) - The system issued a warning of an incipient ROR situation. In response to the warning, the driver maneuvered the vehicle, resulting in a ROR situation.

Missed Detection (MD) - The system failed to issue a warning of an incipient ROR situation.

Correct Non Detection (CND) - The system did not issue a warning and none was required.

Safe False Alarm (SFA) - The system issued a warning, when there was no incipient ROR situation. The driver maneuvered the vehicle, and no ROR occurred.

Unsafe False Alarm (UFA) - The system issued a warning when there was no incipient ROR situation. The driver, overreacting to the alarm, drove the vehicle off the road, causing an ROR event.

The simulation cases were designed so that the driver could react to a warning only during an inattentive period. Thus, LCDs and UFAs were possible only during the inattentive period. Further, MDs were possible only (1) if the CAS was not perfectly accurate or (2) if the TLC threshold was set to a negative value (corresponds to a case where an alarm is issued only after the vehicle is paretically off the road).

RESULTS

Each of the 591 ROR test cases was run with an operating CAS, and the characteristics of the CAS were varied to evaluate the influence on its effectiveness. Five thresholds of the system were tested, each with seven levels of error in the sensor, (including a 0 CAS error case). Note that the negative 0.6 TLC case corresponds to 0.6m distance beyond the lane edge. The results have been plotted a number of ways to study different trends, but the basic question underscoring the analysis is whether the vehicle in a particular simulation left the 12-ft lane, and if so, by how much. In all the figures, the performance of the TLC-based CAS is compared to that for the Sonic Nap Alert Pattern (SNAP), a "rumble strip" feature that has been implemented by the Pennsylvania Turnpike Commission. The SNAP feature is installed just inside the roadway shoulder and in the plots corresponds to a TLC of -.25m.

INFLUENCE OF TLC THRESHOLD - Cumulative Frequency Distributions (CFDs) of maximum lateral excursion are shown in Figure 1 for inattentive driver cases without a CAS and with a CAS with TLC thresholds set at several levels. As shown in this figure, the presence of the CAS generally reduces the population of large vehicle lateral excursions. For example, about seventy five percent of the maximum lateral excursions for the cases without a CAS were above 10 feet. In contrast, only about 10 percent of the cases were above 10 feet using a CAS with a TLC threshold of 1.8 seconds, and the percentage of these cases increased with decreasing TLC threshold. These results provide a good first indication of the potential of a CAS in reducing ROR events by limiting vehicle lateral excursions.

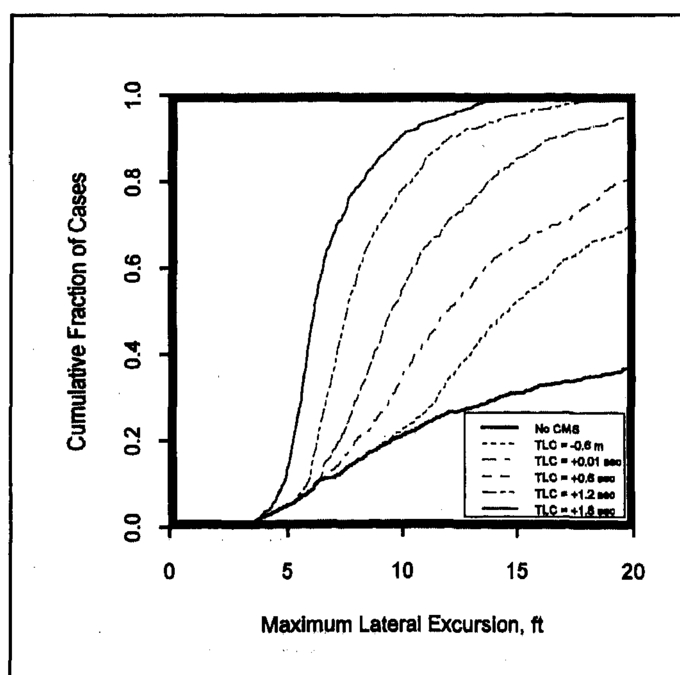


Figure 1. CFDs of maximum lateral tire excursion for inattentive driving with a CAS set at several TLC threshold levels

Lateral position of the front tire is plotted against the different TLC thresholds in Figure 2 at the 50th percentile levels, along with 90% confidence limits. The tight confidence intervals indicate that using sets of 591 simulation cases in this study was sufficient to provide good esttimates of the lateral excursoins. The percentile curves are compared with the no CAS assisted cases in Figure 2, indicating the substantial decrease in maximum lateral excursion levels caused by using the CAS. Also indicated in the figure are the lateral deviations that correspond to the case when 1 tire or both tires are on the shoulder. It is clear that sufficiently high TLC thresholds are required to ensure that not even 1 tire will stray onto the shoulder.

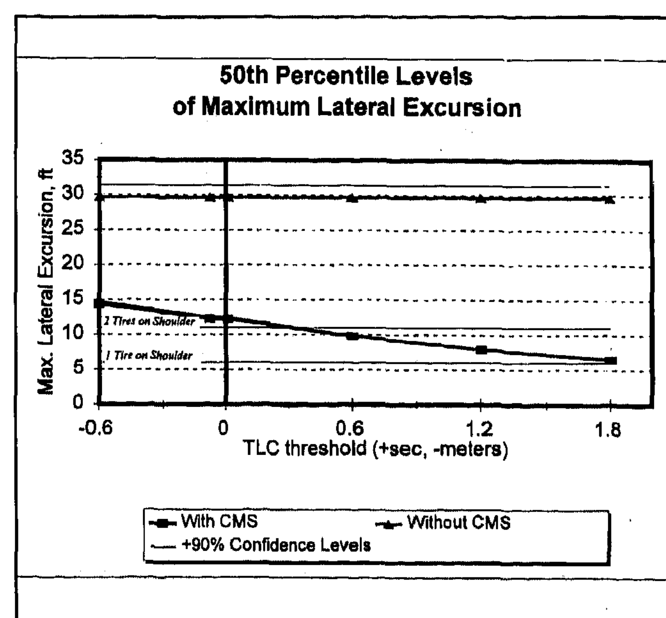


Figure 2. Influence of TLC threshold on 50th percentile values of maximum lateral excursion with 90 percent confidence levels (The very high excursions are only for comparison. The vehicle is well beyond the shoulder when the excursion is more than about 20 ft.)

INFLUENCE OF ROR CRITERION - Both the "1-Tire" and "2-Tire" ROR Criteria were applied to evaluate CAS effectiveness. The results of this exercise are summarized in Figure 3.

In Figure 3, the fraction of cases with maximum lateral excursions exceeding 6 ft (1-Tire ROR Criterion) and 11 ft (2-Tire ROR Criterion) are plotted versus TLC threshold. For comparison, the fraction of cases that exceeded these levels for inattentive drivers without a CAS are shown. The results imply that the perceived effectiveness of a CAS depends strongly on how a ROR event is defined.

Figure 4 highlights the influence of shoulder widths on CAS performance. The percentage of ROR events prevented at different TLC thresholds have been shown for 3 different shoulder widths. The 13 ft shoulder is the nominal shoulder width adopted by the

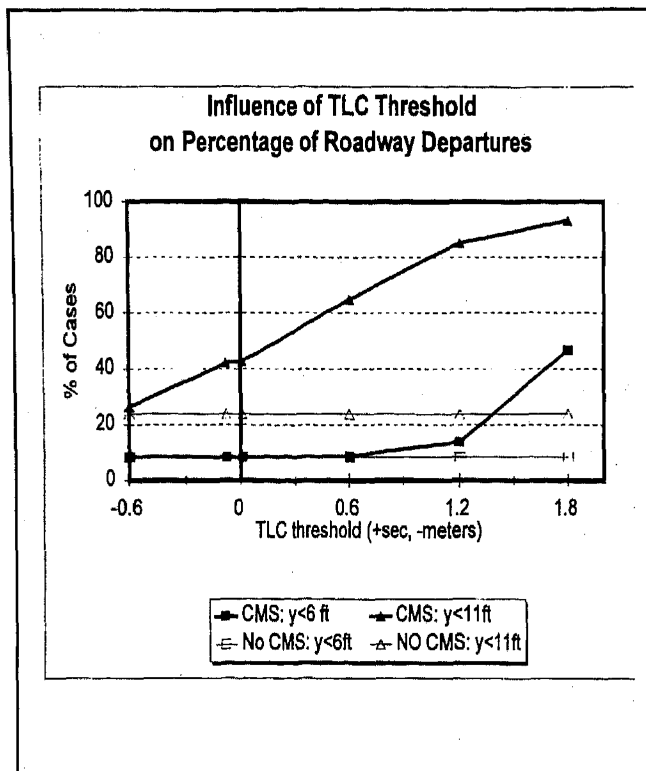


Figure 3. Influence of TLC threshold on CAS performance - comparison of roadway departures based on 1-tire and 2-Tire ROR criteria

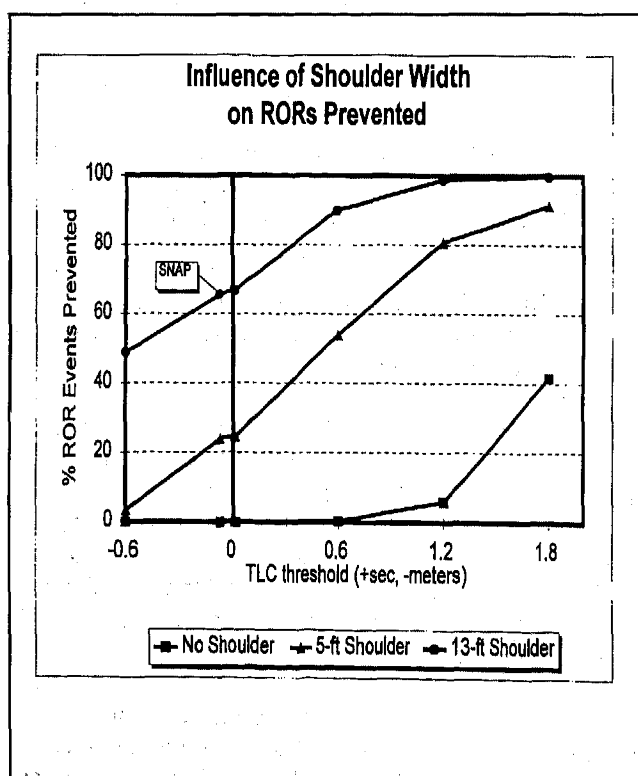


Figure 4. Influence of shoulder width on ROR events prevented at different TLC threshold levels

Pennsylvania Turnpike Commission. As expected the number of ROR events prevented shows a marked with the increase in width of the shoulder. If the driver is allowed to use the shoulder to maneuver the vehicle to avoid a ROR event, then the CAS can be credited with many more ROR preventions compared to a policy of defining any instance of a tire on the shoulder as a ROR event.

Additional results based on application of the 2-Tire ROR Criterion are shown in Figures 5 and 6. In Figure 5, the percentage of ROR cases involving Safe Correct Detects, Late Correct Detects and Missed Detects are plotted as a function of TLC threshold. In Figure 6, the percentage of non-ROR cases involving Safe False Alarms, Unsafe False Alarms and Correct Nondetects are plotted as a function of TLC threshold. As indicated in these figures, the percentage of Safe Correct Detects increases monotonically with increasing TLC threshold, while the percentages of Late Correct Detects and Missed Detects decrease. However, this beneficial trend of increasing Correct Detects is offset by an increase in the percentage of False Alarms (and corresponding decrease in Correct Nondetects). These results stress the need for careful optimization of a CAS to provide the best combination of high correct detections, tolerable false alarm rates and minimal missed detections.

The effect of road curvatures on CAS effectiveness can be seen in Figure 7 which shows the percentage of ROR events prevented for curved, straight and all roads(selected for the study). The percentage of ROR events prevented is higher in the straight than curved roads for higher TLC thresholds. This study

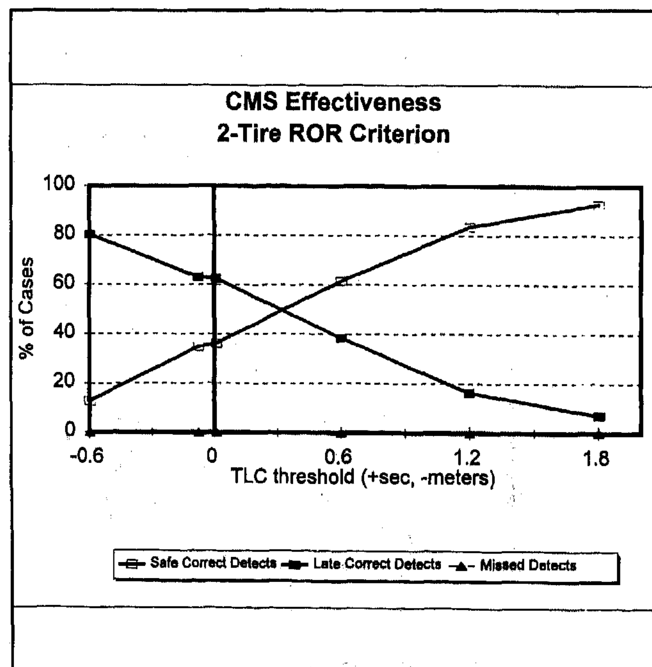


Figure 5. Influence of TLC threshold on CAS performance - comparison of safe and late correct detections and missed detections based on the 2-Tire criterion

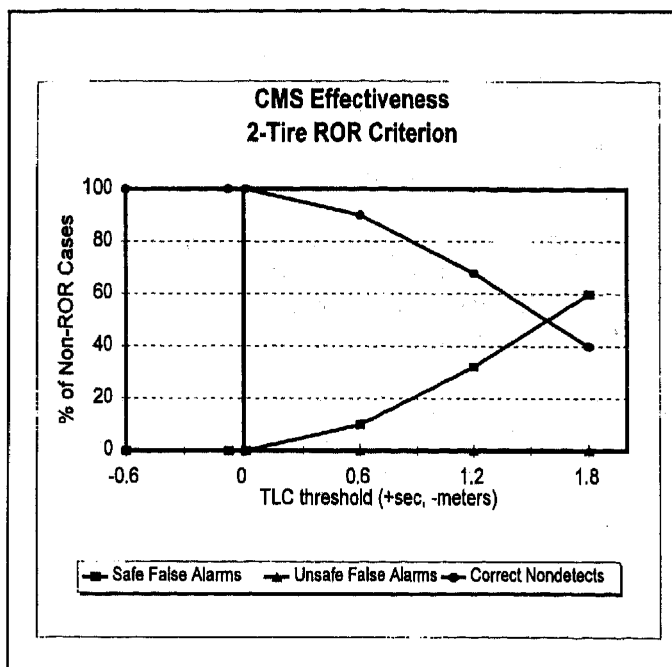


Figure 6. Influence of TLC threshold on CAS performance - comparison of false alarms, detections and non-detections based on the 2-Tire criterion

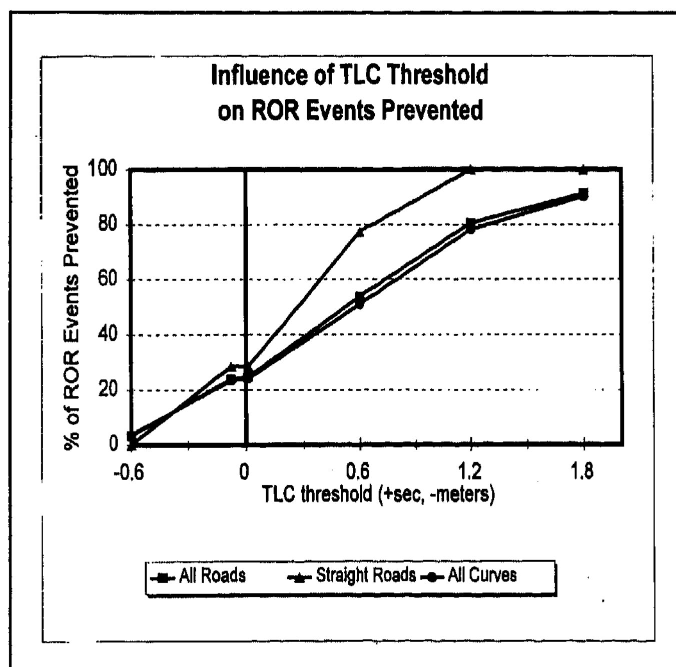


Figure 7. Comparison of CAS performance for different TLC thresholds for straight and curved roads

involves steering control of the vehicles in response to a ROR warning. Future studies reflecting a combination of steering and throttle control should mitigate the speed-curvature effects on CAS effectiveness.

Several thousand simulations were also run to evaluate the influence of CAS accuracy on CAS effectiveness. The studies revealed that lane postion sensors with errors under 6 inches caused no appreciable change.

The TLC based electronic CAS shows consistently better effectiveness than the SNAP system for thresholds greater than 0. This indicates that a CAS with a properly selected TLC threshold could potentially prevent more ROR accidents than such a rumble strip feature, primarily because a CAS provides an earlier warning of incipient danger.

Conclusion

The following conclusions have been drawn from the results of the foregoing simulation studies:

1. The CAS configuration evaluated in the studies showed considerable potential for effectively reducing a vehicle's probability of a ROR accident compared to a vehicle not equipped with a CAS.

2. The TLC method for warning of incipient ROR conditions is effective and is relatively insensitive to errors in lane position measurement (bias) error. A TLC threshold level in the range 1-1.4 seconds may have the most overall effectiveness.

3. The effectiveness of a CAS depends strongly on the amount of clear shoulder available for maneuvering. Therefore, the reported effectiveness depends on the criterion used to define a ROR event. The results of this study indicate that the CAS is much more successful (on the basis of accidents prevented) if instances of the vehicle riding partly on the shoulder are not considered as a ROR event. That is, if the driver is allowed to use some of the shoulder to maneuver the vehicle back into the lane, the success of the CAS is quite good. This strongly suggests the need for information on available shoulder widths, the presence of culverts and overpasses, etc., in order to ensure that proper ROR warnings are issued to the driver on a site-specific basis. In cases where obstacles are unusually close to the lane edge, an occasional early warning would be appropriate.

Acknowledgment

This work was funded by the Office of Collision Avoidance Research, National Highway Traffic Safety Administration as a part of Contract DTNH22-93-C-07023. The authors gratefully acknowledge the support and advice provided by NHTSA during the course of this work.

References

[1] AASHTO (American Association of State, Highway, and Transportation Officials) [1994]. *A Policy on Geometric Design of Highways and Streets.*

[2] Allen, R.W., Szostak, H.T., Klyde, D.H., Rosenthal, T.J., and Owens, K.J., [1982], "Vehicle Dynamic Stability and Rollover." Contract No. DTNH22-88-C-07384.

[3] Ervin, R.D., (Ed.): "The Crewman's Associate for Path Control (CAPC)," Final report for TACOM Contract DAAE07-93-C-R124, University of Michigan Transportation Research Institute, Report No. UMTRI-95-35, 1995.

[4] McKay, M.D., Beckman, R.J., and Conover, W.J., [1979], "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, Vol. 21, No. 2, pp. 239-245.

[5] Pape, D.B., Narendran, V.K., Koenig, M.J., Hadden, J.A., and Everson, J.H., [1996], "Dynamic Vehicle Simulation to Evaluate Countermeasure Systems for Run-Off-Road Crashes." SAE Paper 960517.

[6] Pape, D.B., Narendran, V.K., Hadden, J., Everson, J., Koenig, M., and Pomerleau, D., [1995], "Run-Off-Road Collision Avoidance Using IVHS Countermeasures: Task 4--Volume I." Contract No. DTNH22-93-C-07023.

[7] Pomerleau, D.A., [1995], "RALPH: Rapidly Adapting Lateral Position Handler." Proceedings of the Intelligent Vehicles '95 Symposium, September 25-26, 1995, Detroit, pp. 506-511.

[8] Stein, M.I., [1987], "Large Sample Properties of Simulations Using Latin Hypercube Sampling," Technometrics, Vol. 29, No. 2, pp. 143-151.

970456

Forward-Looking Collision Warning System Performance Guidelines

Terry B. Wilson and Walker Butler
Frontier Engineering Sciences

Dan V. McGehee
University of Iowa

Tom A. Dingus
Virginia Polytechnic Institute and State University

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This document presents preliminary guidelines for a forward looking collision warning system. The intent of a forward looking collision warning system is to eliminate or mitigate vehicular rear-end collisions through driver notification or warning. All aspects of performance are addressed including general system requirements, driver / vehicle interface methodology, collision dynamics, standardized testing and estimation of associated benefits.

This guideline is intended to be used by manufacturers and developers of vehicular based forward looking collision warning systems as a tool to 1) standardize system requirements; 2) standardize driver interface and control among systems developed by different manufacturers; 3) standardize testing to be used in verifying proper operation of the forward-looking collision warning systems and; 4) measure the system benefits prior to widespread deployment of such systems.

PREFACE

The document that follows presents preliminary performance guidelines for forward-looking collision warning systems to improve vehicular safety by eliminating or mitigating rear-end collisions through driver notification or warning. Numerous variations exist within this category such as adaptive cruise control, automatic collision avoidance, and driver warning system types. This document is specific to those devices designed to detect potential rear-end collisions and provide a warning to the driver to avoid the collision. Automatic collision avoidance is excluded from the scope of this document, although some braking of the vehicle is allowed to provide a warning to the driver. Automatic collision avoidance and adaptive cruise control are covered in separate guidelines being developed as part of the research on this contract.

This document represents the second generation of performance specifications associated with this forward-looking collision avoidance research. It represents a significant change in philosophy from the previous version that provided detailed parametric specifications for a forward-looking collision warning system. This change in philosophy is a result of a peer review workshop that was held in December 1995 to receive feedback from industry participants. The consensus was to provide guidelines associated with system performance rather than specifications. This and subsequent versions of the performance guidelines will focus on "what" the forward-looking collision warning system is supposed to do, but not "how" to do it.

The parameters specified within this document should be considered guidelines for acceptable performance. Since forward-looking collision warning systems are extremely complex, especially the man-machine interface, numerous tradeoffs exist that allow the system designer to balance performance over variations in system parameters. System performance that deviates from these guidelines may degrade system effectiveness.

Numerous technical details, beyond the scope of this research must be resolved before a cost effective forward-looking collision warning system can be deployed. The technology required for a forward-looking collision warning system has relatively recent origins in expensive military hardware but must be made very low cost and reliable for entry into the vehicular market.

The major key to success of a forward-looking collision warning system is the man-machine interface. User acceptance is critical to gaining the publics trust of these systems. Technical problems can eventually be overcome but it is impossible, and not very cost effective, to determine all associated driver responses to a properly designed system. A great deal of time and effort has been expended to research the basis of the driver warning methodology presented within this guideline. The results are intended to be broad enough to allow vehicle manufacturers latitude in creating warnings / displays that are acceptable to both vehicle manufacturer and buyer. However, they must be restricted enough to make forward-looking collision warnings salient and unique and at the same time prevent them from interfering with warnings from other systems. Coupling a forward-looking collision warning system with an adaptive cruise control system may be one way to boost public acceptance of this technology.

One result discovered in developing these guidelines was the need for additional research as well as limited deployment of these types of systems to gain real world data. Very little empirical data is currently available to assess wide spread user acceptance of such devices. This should be the focus of future effort: gathering long term data and assessing user acceptance as well as aiding near term deployment.

In developing these guidelines, extensive mathematical modeling and simulation was performed specifically for rear-end collision avoidance. These mathematical models and simulations form the basis for much of the information presented, especially in the area of benefit estimation. Additional research under this contract will verify and refine the models with on-road testing using a test bed system. The simulations will then be re-run and this document updated. Once completed,

additional peer review workshops will be held to discuss the results with industry participants and receive feedback as to the utility of the guidelines.

The guidelines presented herein, allow for both cooperative and autonomous systems as long as the required information can be received, processed and the driver warned in a timely manner. Current industry focus is primarily on autonomous systems.

INTRODUCTION

This document establishes the performance guidelines for a forward-looking collision warning system for improving vehicular safety. A forward-looking collision warning system assists the driver in avoiding or mitigating rear-end collisions. A rear-end collision is defined as a two or more vehicle collision in which vehicles are moving in the same direction prior to the collision or one in which the struck vehicle is stationary.

PURPOSE

The purpose of this document is to present preliminary guidelines for a forward-looking collision warning system for improving vehicular safety.

OBJECTIVE

The objective of a collision warning system following these guidelines is to reduce deaths, injuries and economic losses resulting from rear-end collisions. This is accomplished by aiding in development of forward-looking collision warning systems. This document establishes the necessary system parameters as an aid to system designers and manufacturers of forward-looking collision avoidance systems. This document is intended to provide a standard for driver interface and control aspects of the forward-looking collision warning system and a set of test scenarios to be used to judge the ability of a particular system to meet the guidelines contained herein. A methodology to estimate overall effectiveness and associated benefits of a forward-looking collision avoidance system is also included.

SCOPE

This document presents preliminary performance guidelines for forward-looking collision warning devices to improve vehicular safety by eliminating or mitigating rear-end collisions through driver notification or warning. Numerous variations exist within this category such as adaptive cruise control, automatic collision avoidance, and driver warning system types. This document is specific to those devices that detect potential rear-end collision circumstances and provide a warning to the driver to avoid the collision. Automatic collision avoidance is excluded from the scope of this document, although some braking of the vehicle is allowed to provide warnings to the driver.

BACKGROUND

According to data from the General Estimates System (GES) and Fatal Accident Reporting System (FARS) databases, rear-end collisions are the second largest single category of collisions. They represent 23% of all collisions, 1.45 million per year out of 6.26 million police reported crashes per year. In greater than ninety percent of rear-end collisions, driver inattention/distraction and/or following too closely were contributing factors. This information leads to the conclusion that a forward-looking collision avoidance system should be beneficial in reducing the total number of vehicular crashes and that a system that aids the driver's capabilities, by giving a warning of an impending collision situation or maintaining a safe headway for example, could provide this service.

LIMITATIONS

To the greatest extent possible, the guidelines presented herein are based on empirical research data, detailed modeling and simulation, and discussions with industry experts and/or system developers. Because the basis for this type of research is relatively new and this is the first publicly funded program of this type, the basis for many guidelines is simply sound engineering and/or human factors judgment. Additionally, since this research is a bridge from concept to prototype realization, little data is available to verify the wide spread deployment issues associated with these types of systems.

GUIDELINES

This section presents the guidelines for a forward-looking collision warning system. The guidelines are operating performance parameters that should be included as part of the system. The need for a forward-looking collision warning system is based primarily on driver inattention. However, following too closely is a secondary causal factor. Therefore, the guidelines presented herein contain an inattentive driver warning system that warns the driver of potential crash situations **and** a following-too-closely warning system that provides situational awareness to the driver for following too closely situations and attempts to modify driver behavior to provide a safer driving environment.

A block diagram of a representative forward-looking collision warning system is shown in Figure 1. The forward looking collision warning system uses a forward looking sensor to detect vehicles or objects in the forward path. A collision warning algorithm processes the raw sensor input and determines if a collision threat exists, and if so, warns the driver.

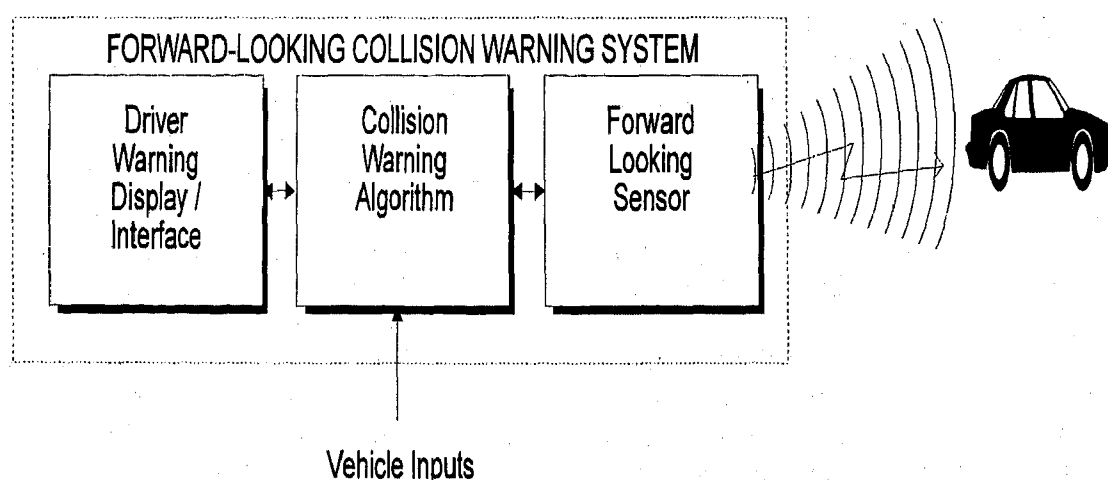


Figure 1 Driver Warning System Block Diagram

SYSTEM GUIDELINES

The system guidelines are those guidelines that are key to bounding the overall system operating environment but do not include issues of driver interface or driver control.

Moving versus Stationary Targets -- A forward-looking collision warning system must detect and warn the driver for all kinematics and dynamics of the host and

lead vehicles especially, stopped vehicles in the forward path. Most rear-end collisions (approximately 70%) occur with the vehicle in the forward path stopped while the remainder occur with the vehicle in the forward path moving (decelerating, constant slower speed, etc.). The ability to detect and warn the driver for stopped vehicle scenarios is the main parameter that distinguishes a forward-looking collision warning system from other types of systems. Other systems such as Adaptive Cruise Control are able to sense stopped vehicles, but ignore these stopped vehicles in order to provide smoother longitudinal control of the host vehicle.

Vehicle Types -- The forward-looking collision warning system shall detect all licensable vehicles. Because of the statistical nature of the forward-looking collision warning system, it is unrealistic to require the system to detect all licensable vehicles at maximum range. This is due to the potentially small "effective" cross-sectional area of some vehicles, such as motorcycles and sports cars. However, the forward-looking collision warning system must detect and react to low cross-section targets to the greatest extent possible. The system must be constructed to allow detection of all licensable vehicles at some (unspecified) range.

It is unrealistic to require each system be tested against every vehicle type for system verification and qualification. Therefore, for qualification testing, representative vehicles shall be used as standard targets.

Acquisition Range -- The forward-looking sensor acquisition range should be ≥ 130 meters. The minimum range at which the forward-looking sensor acquires targets should be 2 meters. The acquisition range is defined as the point at which a vehicle in the forward path has been detected reliably enough to permit tracking. An acquisition range of ≥130 meters is recommended because a system effectiveness of 98% is sufficient to eliminate or mitigate most collisions as shown in the section on benefits estimation. No statistical data exists for minimum acquisition ranges, although a large number of rear-end collisions are known to occur at low velocities and short ranges. As a guideline, the minimum acquisition range should be no greater than two meters. Care must be taken to ensure that the acquisition range is not degraded due to mud, dirt, snow, etc., on the sensor aperture.

Limitations in acquisition range will allow drivers to "drive beyond the sensor limit" of the forward-looking collision warning system. The worst scenario for this problem is the lead vehicle stopped scenario. For example, with a 0.9 second warning time and the host vehicle traveling at 96 kph (60 mph) the required warning distance is 131 meters. The ability to drive beyond the sensor limit may warrant a warning to the driver.

Horizontal Field-of-Regard --The collision warning sensor should have a minimum of ± 8° horizontal field of regard. This is necessary to prevent crashes due to roadway curvature and horizontal field of view limitations. To minimize nuisance alarms from vehicles in adjacent traffic lanes, parked vehicles, roadway signs, etc., the horizontal resolution needs to be fine enough to discriminate vehicles in adjacent lanes out to the acquisition range of the system. For the forward-looking collision warning system to function effectively, it must be capable of reacting properly to vehicles in

the forward path on curves. By providing this capability, the number of nuisance alarms received by the driver should be reduced in relation to curved roadways. Also, a number of rear-end collisions (approximately 9%) occur on curves and a system that ignores road curvature will have a resulting decrease in system benefit. Due to the statistical nature of forward-looking collision warning systems, it will not be possible to respond properly to all curved road scenarios. The guideline of at least ±8° should provide for 97% sensitivity in system effectiveness to horizontal field of view limitations and road curvature, see Figure 6. Scenarios on curves greater than the horizontal field of regard will manifest themselves as late alarms to the driver. This still provides an opportunity for collision mitigation, but will lower system effectiveness.

Vertical Field of Regard -- The driver warning sensor should have a nominal ±2.5° to ±3.5° vertical field of regard. The vertical field of regard was selected to provide 97% sensitivity in system effectiveness to changes in roadway grade and vehicle pitch, see Figure 7. A vertical field of regard of ±3.5° may not be feasible due to the problem of detecting overhead objects as threats. This problem may be eliminated with the system having some vertical *resolution*, but this may not practical with certain sensor technologies. Current knowledge indicates that the forward-looking collision warning system must have some vertical field of regard to respond to changes in roadway grade and vehicle pitch. A minimum vertical field of regard should be ±1° to ±1.5°.

Collision Warning Methodology -- The forward-looking collision warning system must provide a suitable warning to the driver. At a minimum, an inattentive driver warning must be included. A following-too-closely warning would also be beneficial (refer to the Section on driver warning). Rear-end collisions occur for various reasons. The primary reason is due to the driver being inattentive to the driving task (approximately 77% to 84%). The secondary reason is due to the driver following too closely (approximately 7% to 18%). Other rear-end collisions (approximately 4%) are due to vehicle failure, seizure, etc., and are considered unavoidable. A forward-looking collision warning system must provide a suitable warning for the inattentive driver. If the forward-looking collision warning system only contains an inattentive driver warning then the corresponding system effectiveness is reduced (by 7% to 18%). By including a following-too-closely warning the system effectiveness can be improved. Therefore, these guidelines are structured to provide two types of warnings to the driver. First and most important is the inattentive driver warning (see the Section on driver warning). Second, a following-too-closely warning that provides situational awareness feedback to the driver is an attempt to modify driving behavior by creating safer following distances.

Driver Warning Time -- Refer to the Section on driver warning for a discussion of the driver warning time. The warning should be based on range, relative speed and host vehicle absolute speed. It is difficult to provide more than basic guidance to driver warning time at this point in the research. Human factors testing in conjunction with forward-looking collision warning systems must still answer a number of questions prior to providing a sound basis for driver warning time. Testing performed to date in conjunction with this

guideline has shown significant increase in benefit based on **when** the driver is warned prior to collision. The more conservative the warning, the more drivers will find the warning to be a nuisance, and thus lower overall user acceptance and system effectiveness. Additionally, drivers react differently to different types of scenarios which may require different warning times for lead vehicle decelerating, lead vehicle stopped, etc.

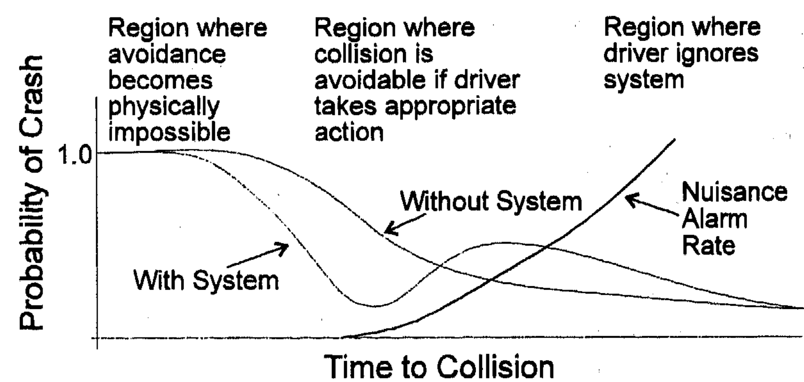


Figure 2 Diminishing Likelihood Curve

The diminishing likelihood curve shown in Figure 2 represents the probability of crash in comparison with the time to collision. In any circumstance there is a region where the collision is physically impossible. In this region, the probability of a crash approaches 1.0. Backing away (in time) from the collision there exists a region where the crash is avoidable if the driver takes appropriate action. The idea of the forward-looking collision warning system is to reduce the probability of a collision through driver notification as shown in the figure. However, if the warning to the driver occurs too far from the potential collision the driver will consider the warning a nuisance and the driver's overall reliance on the system will diminish and/or the probability of a crash will increase with the system. On-going research is attempting to quantify this curve for various collision scenarios.

System Delay Time -- The recommended system delay time is 300 milliseconds or less. System delay time is defined as the time required for the system to determine an object is present in the forward path (acquisition) and warn the driver by changing the status of the warning indicator. The system delay time contains the processing delay time (acquisition and track initiation, state estimation, warning algorithm) and the time required to change the status of the warning indicator. For scenarios where the lead vehicle is acquired at long range (either straight line or curves) longer system delay times can be compensated by longer acquisition ranges or wider horizontal fields of regard (for curves). The area where the system delay impacts overall effectiveness is in lane change and merging situations where lead vehicles are in relative close proximity and warning the driver of collision threats must occur quickly. System delay times that are greater than 300 milliseconds may cause a reduction in overall system effectiveness.

Atmospheric Conditions -- The driver warning system should function properly during varying environments such as rain, road spray, snow, fog, etc. As drivers begin to rely on the system, it is necessary that the system (either directly or through secondary systems) be able to detect when its sensing capability has degraded and inform the driver that the system is not operating properly.

Atmospheric Conditions are highly correlated to roadway conditions. It is anticipated that as roadway conditions degrade that the overall system effectiveness will degrade due to the associated longer stopping distances, independent of the forward-looking collision warning system performance degradation. Unless the system has the ability to detect and compensate for the roadway condition, system effectiveness will likely degrade.

Electromagnetic Compatibility / Electromagnetic Interference -- Other unrelated, non-safety, systems must not emit electromagnetic interference that would interfere with the operation of the forward-looking collision avoidance system. The forward-looking collision warning system must be compatible with other forward-looking collision warning systems developed by different manufacturers (FCC part 15 rules). Other specifications may apply. The system shall comply with DHHS Rules 21 CFR Subchapter J. Other specifications may apply such as SAE J1113. Due to the safety nature of these systems, it may be necessary to require that other emitting devices, that are not safety related, not be allowed to produce harmful interference that would interfere with a forward-looking collision warning system. Not much emphasis is placed on EMI / EMC for the prototype systems being developed by various manufacturers. However, once these systems have been developed it is possible that EMI issues could prevent deployment. This may be especially true of optical systems that are not restricted to a particular bandwidth for the collision avoidance task.

Safety -- The system must pose no threat to human safety and must meet or exceed all regulations for the provided technology.

IEEE *Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 KHz to 300 GHz*, IEEE, C95.1-1991

Department of Health and Human Services, Food and Drug Administration, *Part 1040: Performance Standard for Laser Radiation Emitting Products*, 21 CFR

DRIVER WARNING

Standard driver displays are desirable to prevent confusion by the driver when using systems from different manufacturers. The methodology used to communicate with the driver is of paramount importance. The quicker the system can direct the driver's attention, and the faster the system can get the driver to respond the better the overall system effectiveness. All of the benefits of implementing a forward-looking collision warning system could be lost on a poorly designed driver interface.

It is important to understand that rear-end collision have two primary types of causal factors: inattention and following too closely. Primary focus is on the inattentive driver since they constitute a larger overall problem (approximately 77% to 84%), but ignoring the following too closely

(tailgating) driver also ignores a portion of the rear-end problem (approximately 7% to 18%). Therefore, the system described and recommended within this guideline consists of two separate warning systems, one for the inattentive driver and one for the following-too-closely driver.

Inattentive Driver Warning -- The recommended inattentive driver warning is haptic (tactile or proprioceptive). This imminent warning should utilize braking initiation with a high jerk rate limit and a moderate deceleration level. An imminent collision warning is defined as requiring an immediate vehicle control response or modification of a planned response in order to avoid a collision. Because of the inattentive nature of the driver that is typical at the point of the warning, it is necessary to design the warning to grab the driver's attention quickly. Haptic or auditory alerts are the only sensory inputs capable of communicating with an inattentive driver. The haptic alert is recommended because it can grab the driver's attention and begin to slow the longitudinal progress of the vehicle with a single warning. Any alert modality will cause a delay in driver response due to driver perception and reaction. By using braking initiation, some of this delay can be overcome with the deceleration component of the warning. Auditory alerts are not recommended. However, if the warning is auditory it should emanate from the front of the vehicle to draw the drivers' attention toward the hazard. It should also be unique and salient and separate from all other displays especially other collision warning displays. An auditory display shall override the background noise level and be uniquely identifiable from other sound sources. User acceptance for auditory warnings is considered low, and may not be uniquely identifiable when other collision warning systems are present on the vehicle. A visual display can provide no warning benefit to the inattentive driver. A visual display can be used to provide feedback to the driver that the warning is not a malfunction of the system or vehicle. Jerk rate limit equates to a delay in actuation and therefore a high value is specified. Work is progressing on quantifying driver response to jerk rate limit and deceleration value. For this preliminary research, the following limits are defined:

Deceleration Level	Deceleration Value (g's)	Jerk Rate Limit (m/s³)
Low	$\alpha \leq 0.15$	$\beta \leq 35$
Moderate	$0.15 < \alpha \leq 0.3$	$35 < \beta \leq 75$
High	$\alpha > 0.3$	$\beta > 75$

A cautionary collision warning is defined as requiring immediate attention from the driver and which may require a vehicle maneuver, but which does not meet the definition of an imminent collision avoidance situation. A "following-too-closely" warning can be considered a cautionary warning for the inattentive driver. If a following-too-closely warning is not available then the system should have a cautionary warning. For this cautionary warning display it is recommended that a push-back accelerator pedal be utilized. A "push-back" accelerator pedal resists the

force placed by the driver on the accelerator pedal. This force informs the driver that he/she is driving too close to the vehicle in front, but does not prevent the driver from overriding.

Future vehicles may include a number of collision warning systems such as: rear-end, backing, side-looking, etc. It will be necessary for the driver to be able to automatically distinguish the rear-end crash warning from another warning type. For this reason, it is desirable for the haptic queues to be restricted to longitudinal haptic queues. The imminent warning using braking initiation is desirable since it tends to increase overall response to the collision by initiating vehicle braking. Braking initiation not only draws the driver's attention, but provides the driver with guidance to a desirable response. The braking initiation should be restricted to approximately moderate values (justification currently being researched). Higher values of braking initiation are possible but this begins to get into the realm of collision avoidance. The intent of the braking initiation for these guidelines is to provide the driver a warning and a suggested course of action. Haptic warning queues are recommended over auditory warning queues based on studies of user acceptance because haptic warning queues do not alert passengers in the vehicle to errors in judgment by the driver. The "push-back" accelerator pedal provides good system feedback to the driver with the ability for the driver to ignore and override the warning. Auditory displays must be easily separable from background noise for the driver. It may be difficult for the auditory warning to be distinguished over background noise without startling the driver.

If the collision warning system employs active deceleration (braking initiation) as a haptic warning to the driver, then the system should illuminate the brake light when the deceleration system is active.

Inattentive Driver Warning Time Model - The driver of the system should be warned when an imminent collision threat exists. The warning time should be based on range to vehicles (objects) in the forward path, relative speed (range rate) and host vehicle absolute speed. The driver should be warned in response to the vehicle that presents the highest threat of collision. Refer to the Section on driver warning for a discussion of the driver warning time. Without a warning, a driver is in jeopardy of becoming inattentive and having a collision. Studies have shown that any warning will tend to increase the driver's perception time. A warning must be processed by the driver before the driver can react. This adds a finite amount of delay to the driver's reaction. A properly timed warning can actually decrease driver reaction even with the resulting increase in driver perception time.

Following-too-Closely Display -- A Following-too-Closely display should provide the driver situational awareness for tailgating scenarios. It is not the intent of this warning to provide the driver with an optimum following distance but rather make the driver aware of unsafe following (tailgating) situations. It is expected that the following-too-closely warning will be relatively conservative in comparison to the inattentive driver warning. This may lead to an increase in nuisance alarms to the driver and may justify allowing the driver to turn off the following-too-closely warning. However, there is a corresponding reduction in system effectiveness if this feature is

allowed. This includes the ability to "dim" the display (assuming a visual display) along with the vehicle dash lights. These guidelines do not restrict the following-too-closely display from being adjustable by the driver. Again, this adjustability may cause lower overall system effectiveness.

The following-too-closely display can be visual, auditory or haptic. If the following-too-closely display is visual, then the following color codes should be utilized: Green shall be used for safe following distances, Orange, Amber, or Yellow for cautionary following distances and Red for unsafe following distances. The display may flash or vary in intensity when a collision becomes imminent. The visual display should be mounted to protrude above the dashboard slightly above the speedometer and be positioned within 15° of the driver's line of sight (this provides about 12-20 inches of margin for placement of the display in the dashboard). The visual display should be separate from all other displays and must be large enough with sufficient contrast and intensity for the driver to perceive the display in his or her peripheral vision. An example visual display is shown in Figure 3. If the following-too-closely display is haptic then it is recommended that a "push-back" accelerator pedal be used.

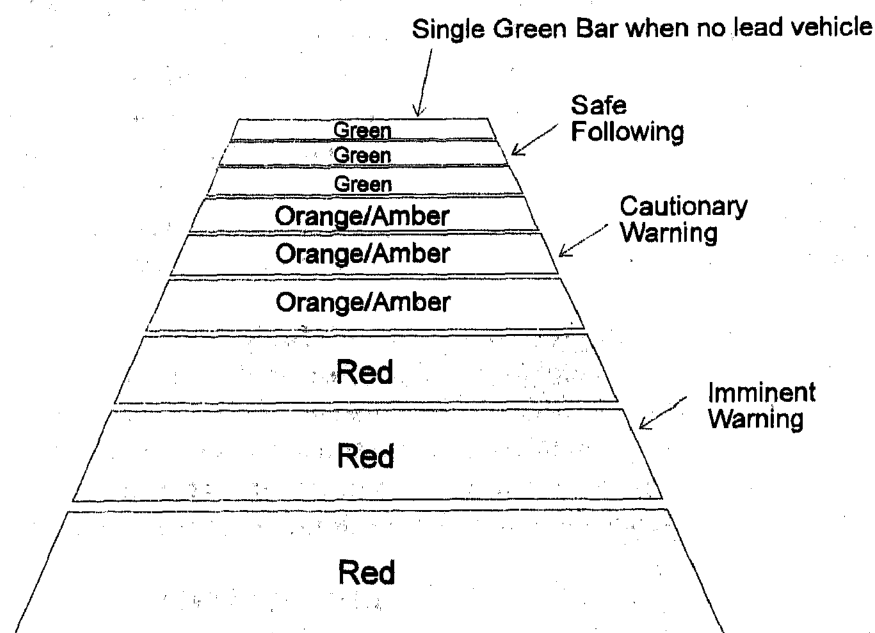


Figure 3 Example Following-to-Closely Visual Display

Following-too-Closely Warning Time Model -- The driver of the system should be made aware of following-too-closely situations. The warning time should be based on range and host vehicle absolute speed. Refer to the Section on driver warning for a discussion of the driver warning time. This warning should provide situation awareness to the driver and should be visual in nature.

Nuisance Alarms -- Nuisance alarms must be controlled to the greatest extent possible. Additional research, in relation to this contract, is being performed to quantify this requirement. Nuisance alarms are defined by efficacy rate. Efficacy rate is the number of times the system provides a true positive to the number of times that a signal is required according to the following table. Efficacy rate is also known as the hit rate and the complement is known as the "miss" rate.

System Response	Situations requiring a signal	Situations in which a signal is not required
Signal	True Positive (hit)	False Positive (false alarm)
No Signal	False Negative (miss)	True Negative

Nuisance Alarms are defined as false positive indications that occur when the system functions as designed but when the situation does not constitute a true crash threat for the driver in question. Conversely, false alarms are defined as false positive indications that occur when the system does not function as designed. From the driver's perspective, nuisance alarms and false alarms may appear to be the same but the causes and remedies are different.

Minimizing nuisance alarms is the key to user acceptance for forward-looking collision warning systems. It is not expected that nuisance alarms will ever be completely quantified because by definition, nuisance alarms are based on an individual drivers' driving behavior.

Additional Display(s) -- The forward-looking collision warning system may contain additional displays. If only haptic warnings are used, it is recommended that a visual status display be used to provide the driver feedback that the system is functioning properly. Additional displays may contain fault and/or status information as long as this information is separated from the warning display. It may be beneficial to warn the driver when driving beyond the sensor limit of the system by using a status display.

Icon -- Vehicles equipped with a forward-looking collision warning systems should display a unique and standardized symbol to inform the driver that the vehicle is so equipped. This icon should be different from ACC icons that may be present on the vehicle, or other types of collision warning system icons.

Fault Indication -- A fault indication must be presented to the driver when the system fails a self test, operational diagnostics test, or the sensor is unable to reliably detect vehicles in the forward path making the system inoperable.

DRIVER CONTROL

Standard driver controls should be used to avoid confusing the driver when using systems from different manufacturers.

Power-On Activation -- The driver warning system should power-on with application of ignition power. All other settings should be automatic. Under no circumstances should the driver have the ability to turn-off or disengage the imminent inattentive driver warning system. Drivers cannot be relied on to turn

on their collision warning devices each time they turn on the vehicle ignition. Therefore, the system should turn on automatically each time the vehicle is started. To prevent rear-end collisions, the inattentive driver warning must always be active. It is possible to allow the cautionary warning or the following-too-closely warning to be disabled, but again the result will be a lower overall effectiveness. The driver may forget that the system has been turned off and have a false sense of security about his or her driving behavior.

Warning Time Adjustment -- No adjustment of the inattentive warning time should be included with the system. By allowing the driver to adjust the warning time, it places the driver in the role of expert in knowing when the warning should occur. Current human factors testing of forward-looking collision warning systems shows that the inattentive driver warning should be based on an aggressive algorithm to improve driver response and reduce nuisance alarms. A warning time adjustment feature may be included as part of the cautionary warning or the following-too-closely warning. Under no circumstances should the driver have the ability to adjust the warning time below the minimum imminent warning time. If the driver was allowed to adjust the imminent warning time, below the minimum, the overall system effectiveness may be significantly reduced. Also,

DYNAMIC SITUATIONS

Dynamic situations describe the kinematics associated with longitudinal encounters with vehicles. Dynamic situations also provide a introduction and basis for the qualification testing. Two types of dynamic situations exist, one based primarily on the kinematics involved and the interaction of the two vehicles. Table 1 shows a matrix of lead and following vehicle dynamic situations. The fifteen dynamic situations consist of all combinations of five lead vehicle and three following vehicle kinematic situations. The system must meet the requirements of these guidelines under all kinematics of the host and target vehicle(s). The percentages shown are the actual occurrence of rear-end collisions as taken from an analysis of the 1992 National Accident Sampling System (NASS) Crashworthiness Data System (CDS).

Table 1 Percent of Rear-End Collisions versus Dynamic Situations

Lead Vehicle	Following Vehicle			Total
	Accelerating	Constant Velocity	Decelerating	
Stopped	1%	18%	1%	20%
Constant Velocity	2%	7%	0%	9%
Decelerating	0%	14%	3%	17%
Accelerating	0%	2%	0%	2%
Decelerating & Stopped	1%	50%	1%	52%
Total	4%	91%	5%	100%

The other type of dynamic situation is based on typical scenarios that lead to rear-end collisions. With each situation, an explanation of the system response is included.

No vehicles in the forward path -- The forward-looking collision warning system continually monitors the forward path of the host vehicle. No warnings are issued to the driver. The driver may be warned if he or she is driving beyond the sensor limit of the system, that is the absolute speed causes the warning distance to be beyond the acquisition range. Minimization of nuisance alarms in this situation is critical especially in relation to close proximity roadway objects and curves.

Vehicle in forward path, stopped -- If the range, relative speed and host vehicle absolute speed meet the warning criteria, then the driver is warned. This is a driver warning situation with a high relative speed between vehicles. Therefore, the warning system must warn the driver, the following-too-closely display may or may not be activated in this circumstance.

Vehicle in forward path, moving at the same speed -- The relative speed is assumed to be near zero and the following range short (tailgating). No driver warning is issued unless the relative speed causes an imminent collision warning. The cautionary collision warning may be used to provide the driver a warning for this circumstance. The following-too-closely display should be illuminated to inform the driver of the tailgating situation.

Vehicle in forward path, decelerating or decelerating to a stop -- If the range, relative speed and host vehicle absolute speed meet the warning criteria, then the driver is warned. This is a driver warning situation with a high deceleration rate possible between vehicles. Therefore, the warning system must warn the driver, the following-too-closely display should be activated if the driver was following too closely.

Changing Lanes -- If the lead vehicle speed is greater than the host vehicle's absolute speed, no warning is issued. If the range, relative speed and host vehicle absolute speed meets the warning criteria then the driver is warned. The following-too-closely display is illuminated if the criteria is met. In this scenario the driver's view may be blocked and the sensor's view may be blocked to the threat vehicle. This scenario may not allow sufficient time for an attentive driver to react and stop, or allow the system to warn the driver to avoid the collision. Certain kinematic situations exist with this scenario that may make the collision unavoidable, see the section on driver warning.

Vehicle merging into forward path -- If the lead vehicle speed is greater than the host vehicle's absolute speed, no warning is issued. If the range, relative speed and host vehicle absolute speed meets the warning criteria then the driver is warned. The following-too-closely display is illuminated if the criteria is met. This scenario may not allow sufficient time for an attentive driver to react and stop, or allow the system to warn the driver to avoid the collision. The sensor beam must be formed so as to allow for the earliest warning possible for this type of scenario. Certain kinematic situations exist with this scenario that may make the collision unavoidable.

QUALIFICATION TESTS

The following tests have been outlined to verify driver warning system operation in relation to the guidelines presented herein. A "black box" approach has been used to develop the qualification tests. This is to allow testing of "off-the-shelf" units instead of requiring the manufacturer to deliver a special unit for testing. Details of the qualification tests will be left to a separate document. Tests may be executed a number of times to judge system repeatability. Other requirements for testing are that the tests be relatively easy to perform and easily repeatable.

INTERFERENCE

The driver warning system shall comply with applicable standards for Electromagnetic Compatibility (EMC) and Electromagnetic Interference (EMI).

SAFETY

The system shall pose no threat to human safety and must meet or exceed all regulations for the provided technology. The directed energy from the sensor shall be measured and compared with the applicable standard(s).

STANDARD TARGETS

The guidelines are intended to be as technology independent as possible. As a result, test vehicles will be used as standard targets for the system under test. The following types of vehicular targets may be included: cars, light trucks, buses, heavy trucks, towed trailers, and motorcycles. These vehicle types will be detected and tracked without the requirement for special vehicle reflectors or an unreasonably high minimum cross sectional area. Additionally, the system may detect other objects in the forward path.

LONGITUDINAL PERFORMANCE

To execute the longitudinal performance tests, a measure of the absolute velocity of both vehicles as well as an independent measure of the range and relative velocity between vehicles will be required.

Inattentive Driver Warning -- The inattentive driver warning will be measured using a straight line approach to both stopped and moving lead vehicles. The host vehicle with the forward-looking collision warning system will proceed toward a lead vehicle traveling on a straight section of roadway. A measure of the range (distance in meters) at which the system warns the driver (both cautionary and imminent) will be obtained along with measures of absolute and relative speed for the two vehicles.

Following-too-Closely Warning -- The following-too-closely warning will be measured using a moving lead vehicle in a tailgating scenario. The host vehicle with the collision warning system will tailgate the lead vehicle traveling at a constant velocity. A measure of the range (distance in meters) at which the following-too-closely warning system warns the driver (both cautionary and imminent) will be obtained along with measures of absolute and relative speed for the two vehicles.

Range Resolution -- The ability of the forward-looking collision warning system to warn the driver on a smaller vehicle positioned behind a larger vehicle will be measured. This test will be performed in a manner similar to the inattentive driver warning test using a large lead vehicle (heavy truck trailer) stopped with a small vehicle (motorcycle) positioned behind the large lead vehicle. The tests will be executed without the small vehicle and then with the

small vehicle and the warning difference measured.

Angular Resolution -- The ability of the forward-looking collision warning system to not warn the driver while passing other vehicles will be measured. Using a straight section of three lane roadway, with the host vehicle occupying the center lane. Stopped and moving lead vehicles will be positioned one lane width to the right and one lane width to the left. The host vehicle will pass the lead vehicles and warning indications from the system will be measured.

CURVE PERFORMANCE

Horizontal Field-of-Regard Testing -- The performance of the forward-looking collision warning system will be measured on curves. Using a lead vehicle stopped, and a curve of known radius the inattentive driver warning will be measured.

Vertical Field-of-Regard Testing -- The performance of the forward-looking collision warning system will be measured on grades. Using a lead vehicle stopped, and a known change in grade the inattentive driver warning will be measured.

MERGE PERFORMANCE

The performance of the forward-looking collision warning system will be measured for merging scenarios.

System Delay -- The system delay of the forward-looking collision warning system will be measured. Using a blocking vehicle, a merging scenario will be scripted to allow the lead vehicle to "pop-up" in the warning range. The delay in warning from this scenario to the straight line scenario will be measured.

OTHER PERFORMANCE

Atmospheric Testing -- The performance of the forward-looking collision warning system will be assessed over varying atmospheric conditions. It is not possible or cost effective to measure system performance over varying atmospheric conditions. Therefore, no direct atmospheric testing will be performed. The manufacturer must provide evidence that the system performs well in varying atmospheric conditions.

Driver Interface -- The driver interface and the driver display to the system will be compared to this guideline for compliance. The intent is to keep a record of the system interface functions to determine what features may cause a detriment or benefit to overall safety. It is overly restrictive for this guideline to specify the exact nature of the driver interface. Many variations exist that should provide acceptable performance. The intent of this testing is to keep a record of interface characteristics to note any potential trends in safety benefit or dis-benefit.

BENEFITS ESTIMATES

This section presents system effectiveness versus system design parameters and a cumulative estimate of benefits. Prior to widespread deployment of forward-looking collision avoidance systems, an estimate of potential system benefit is desirable. This estimate will help quantify the need for such systems and allows top-level parameters to be optimized. In order to perform this, extensive mathematical modeling and simulation of a forward-looking collision warning system has been performed. The mathematical model is based on a 6-Degree-of-Freedom model with various random variables and constant quantities. The simulation is a Monte Carlo simulation

using a set of 159 accident cases that is representative of the population of rear-end collisions.

As previously stated, Rear-End Crashes represent 23% of all collisions, 1.45 million per year out of 6.26 million police reported crashes per year. This represents the scope of rear-end collisions over which the following statistics apply.

Research into rear-end causal factor analysis has identified the primary causes of rear-end collisions to be driver inattention and/or following-too-closely. Additionally, the causal factor analysis did identify four percent of the rear-end collisions that were unavoidable due to brake failure, seizure, etc.

The mathematical model is based on a modified warning distance algorithm that takes lead and host vehicle absolute speed and calculates a warning distance (range, relative speed and absolute speed). This warning distance is compared with the range to the vehicle in the forward path and if exceeded, the driver is warned.

$$WD = \frac{V_f^{\,2}}{2*\alpha_f} - \frac{V_l^{\,2}}{2*\alpha_l} + V_f * T_d + R$$

where,

V_f	following (host) vehicle speed
α_f	following (host) vehicle acceleration 0.75 g's
V_l	lead vehicle speed
α_l	lead vehicle acceleration 0.85 g's
T_d	warning time 2.4 seconds
R	confidence interval 2 meters, this is provides a stopping gap between the vehicles

Additionally,

D_{RT}	driver reaction time log-normal distributed between 0.9 and 2.4 seconds with a 0.3 second standard deviation and a mean that is a linear function of headway time.
T_{PROC}	processing delay time 0.300 seconds, is the delay between request for the warning and the actual warning initiation
R_a	acquisition range $\geq$ 130 meters

Noise is added to range and range rate measurements as in a real-world system. A optimized Kalman filter is used to perform range and range rate estimation.

Warning times are not solely a function of the warning time (T_d) as specified in the mathematical model. It is a result of a combination of parameters including the assumed deceleration, acquisition range, processing delay, etc.

System effectiveness is defined as the number of crashes prevented divided by the total number of runs. The mitigation rate is meant to give a broader measure of benefit than system effectiveness. The mitigation rate takes into account crash severity, while the effectiveness counts a 1 kph crash the same as a 55 kph crash, it is assumed that damage is proportional to kinetic energy which is proportional to velocity squared. The total damage is then the sum of the squares of the damages from all of the runs. It is then normalized by dividing by the total maximum damage which could result if no action were taken. So the mitigation is defined as:

$$Mitigation = 1 - \frac{\sum_{i=1}^{N} (\text{Final Lead vehicle velocity}(i) - \text{Final Following vehicle velocity}(i))^2}{\sum_{i=1}^{N} (\text{Initial Following vehicle velocity}(i))^2}$$

Where the velocity in the denominator is the initial following vehicle velocity and the velocities in the numerator are velocities measured at the time of impact. Mitigation is always the upper curve on the plots.

Warning Times -- Sensitivity of system effectiveness versus driver warning time is shown in Figure 4.

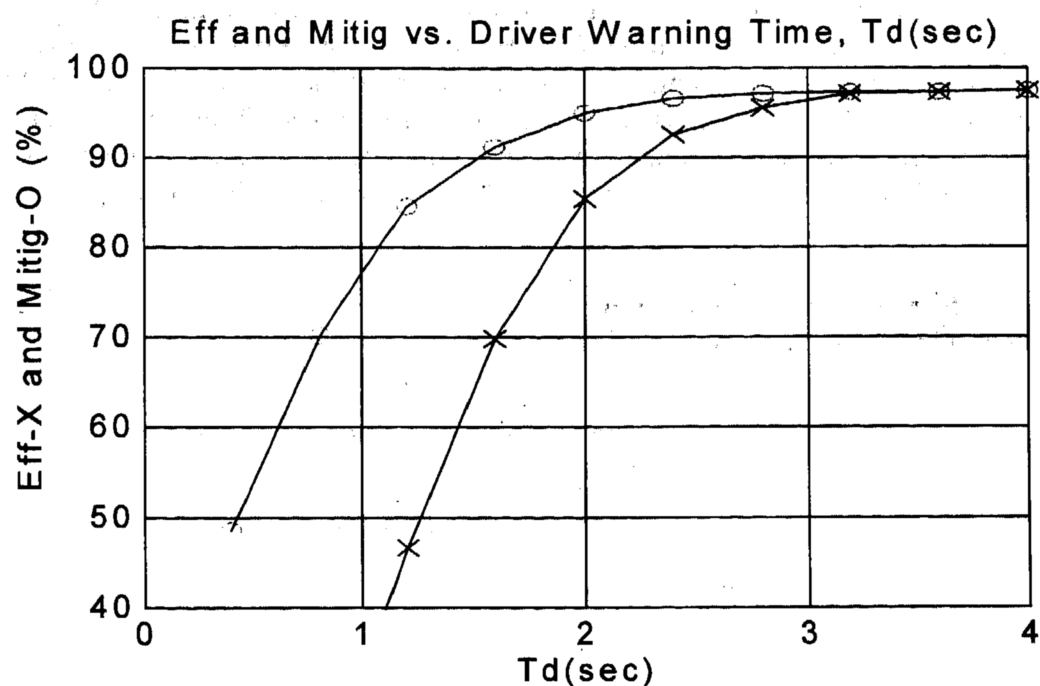


Figure 4 Effectiveness and Mitigation to Warning Time

Acquisition Range -- Sensitivity of system effectiveness versus acquisition range is shown in Figure 5.

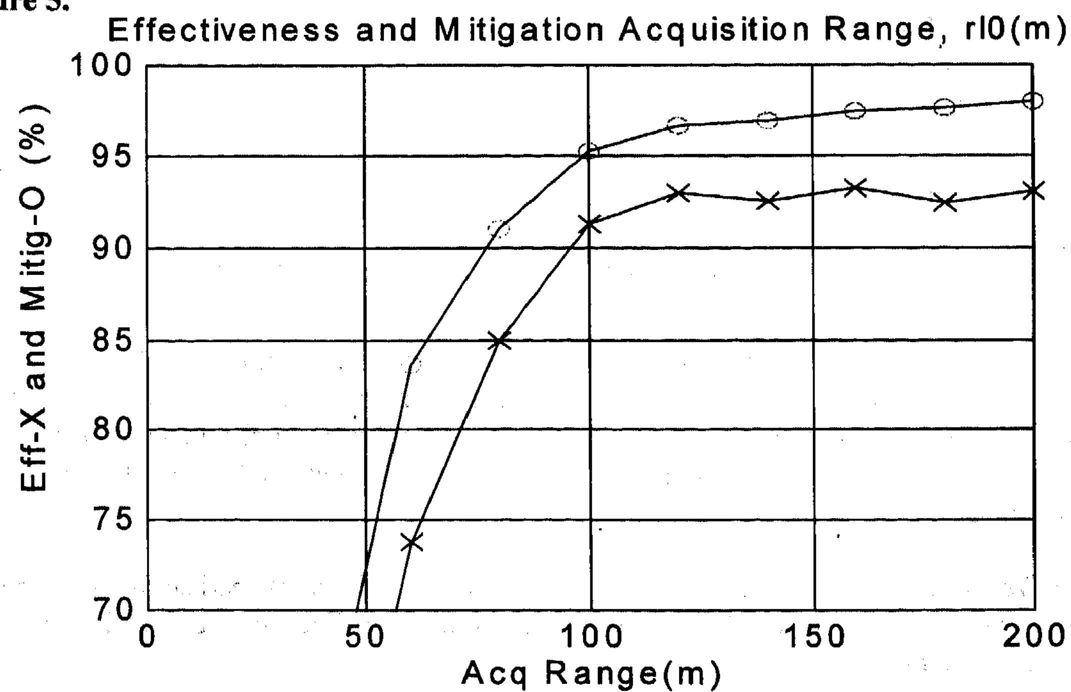


Figure 5 Effectiveness and Mitigation to Acquisition Range

The acquisition range plot includes both stationary and moving vehicles. For the purpose of these guidelines an acquisition range of ≥ 130 meters was selected. Because of the statistical nature of these systems, it was felt that a sensitivity of 93% is sufficient to eliminate or mitigate most collisions. A minimum acquisition range of 2 meters is a guideline. It should be noted that even at an acquisition range of ≥ 130 meters that the effectiveness drops off rapidly for high closure rate scenarios, such as lead vehicle stopped. At 65 mph the effectiveness drops to 64% for a lead vehicle stopped scenario and to 15% at 75 mph. This is a problem with any acquisition range chosen.

Horizontal Field-of-Regard -- Sensitivity of system effectiveness to horizontal field of regard is shown in Figure 6.

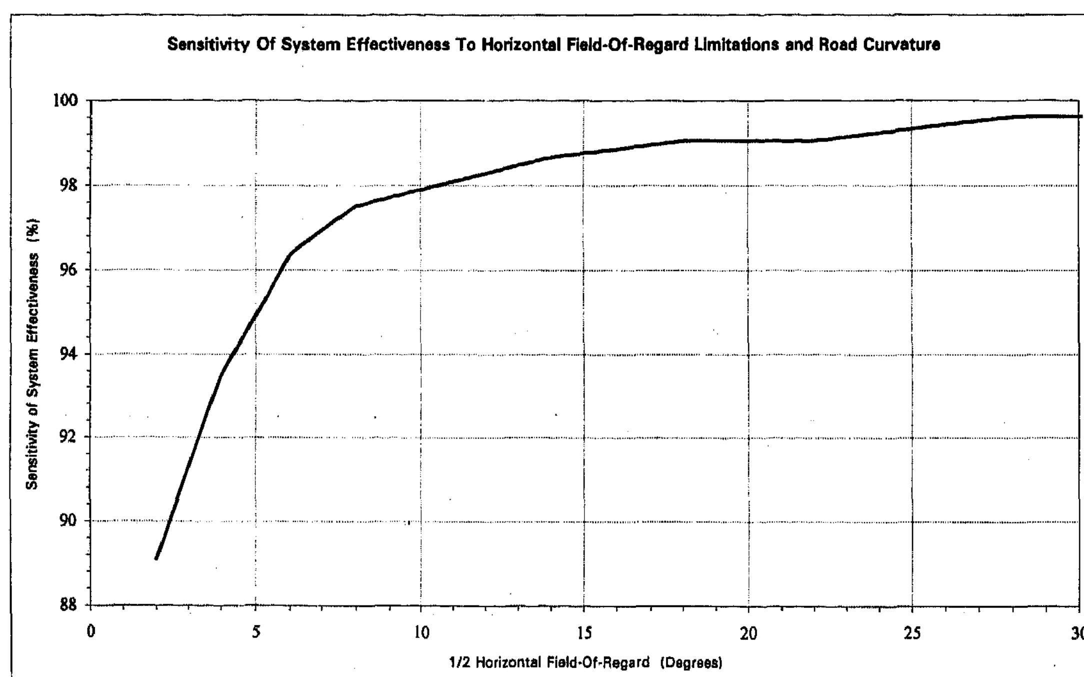


Figure 6 Effectiveness to Horizontal Field-of-Regard

The horizontal field of regard refers to the angle of view in the horizontal direction referenced to the longitudinal axis of the host vehicle. Because of the statistical nature of these systems, it was felt that a sensitivity to system effectiveness of 97% is sufficient to eliminate or mitigate most collisions. The sensitivity approaches 100% but never reaches 100% due to limitation of acquisition range.

To minimize nuisance alarms from vehicles in adjacent traffic lanes, parked vehicles, roadway signs, etc., the horizontal resolution must be fine enough to discriminate vehicles in adjacent lanes out to the acquisition range of the system.

Vertical Field-of-Regard -- Sensitivity of system effectiveness to vertical field of regard is shown in Figure 7.

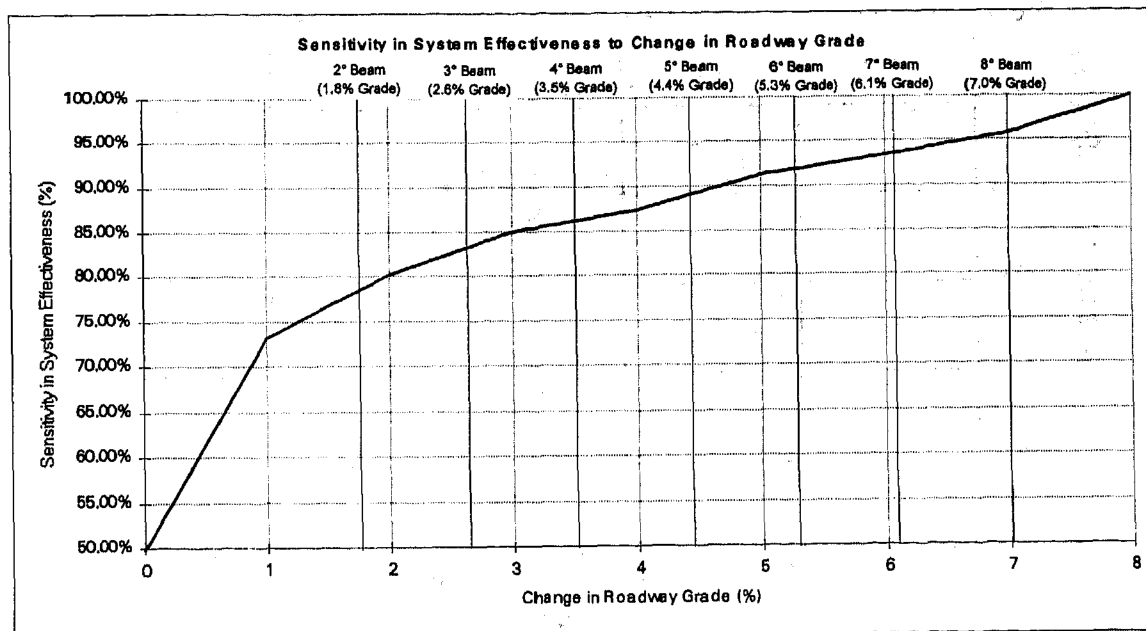


Figure 7 Effectiveness to Vertical Field-of-Regard

The vertical field of regard refers to the angle of view in the vertical direction referenced to the longitudinal axis of the host vehicle. Vertical field of regard relates system performance to roadway changes in grade and host vehicle pitch. The main issue for a forward looking system is the change in grade from the host to the target vehicle. This requirement affects the vertical field of regard required as well as the vertical angular resolution. The vertical field of regard should be large enough to overcome problems with the host and lead vehicle being on different roadway grades but not so large that overhead roadway objects present nuisance alarms.

System Delay -- Sensitivity of system effectiveness to system delay is shown in Figure 8.

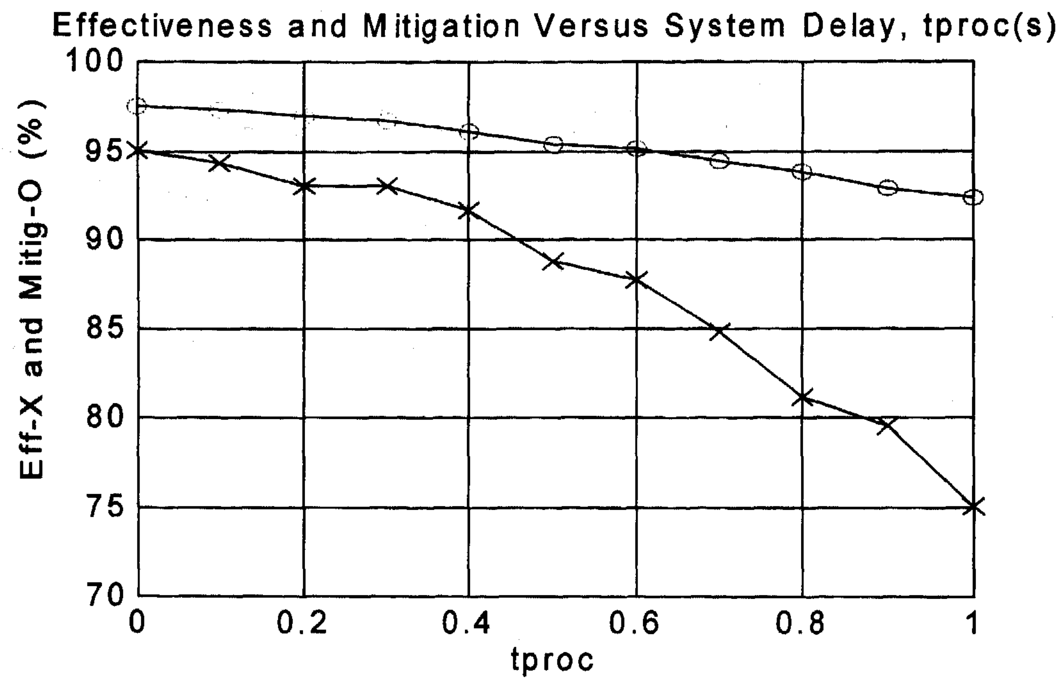


Figure 8 Effectiveness and Mitigation to System Time Delay

System delay is defined as the time required for the system to determine an object is present in the forward path and warn the driver by changing the status of the display. The system delay time contains the target acquisition delay time, the time to process through the warning algorithm and the time required to change the status of the display. A value of ≤ 300 milliseconds was selected. Because of the statistical nature of these systems, it was felt that a sensitivity to system effectiveness of 94% is sufficient to eliminate or mitigate most collisions.

Atmospheric Conditions -- Sensitivity of system effectiveness to atmospheric condition is shown in Figure 9.

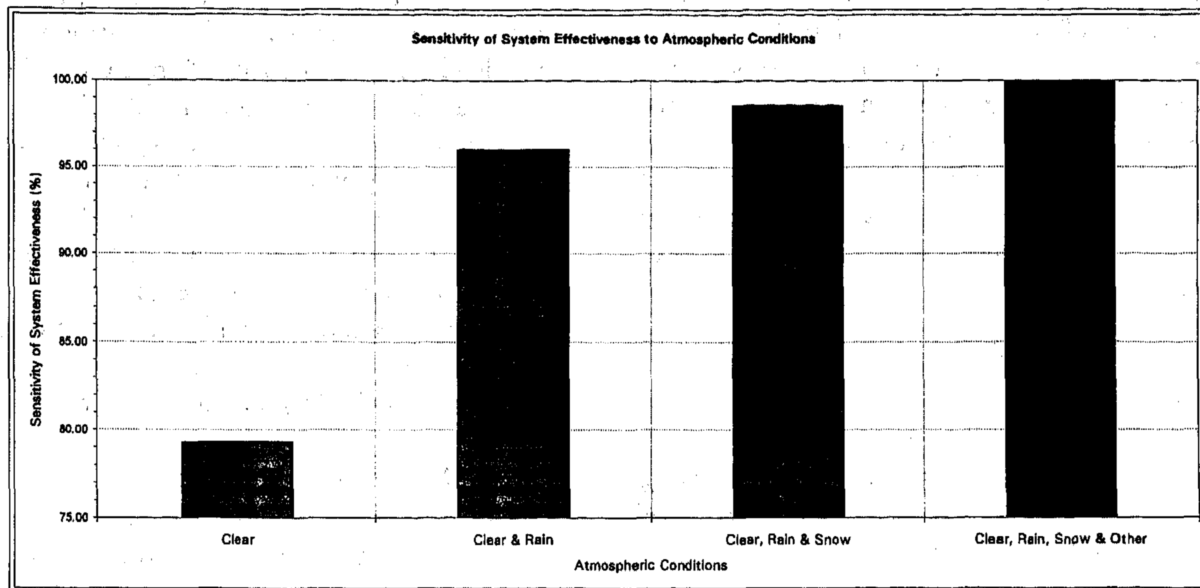


Figure 9 Atmospheric Condition vs. System Effectiveness

Atmospheric conditions relate to the system performance during various conditions: rain, snow, fog, dust, etc. If the system operates in clear and rainy conditions then the system effectiveness is approximately 96% instead of 79% for a system that only operates in clear atmospheric conditions. No quantitative measure of the amount or density of the atmospheric condition is available, but for a forward-looking collision warning system to be effective, it must operate correctly under degraded atmospheric conditions, or notify the driver that it is non-functional.

Table 2 Forward-Looking Collision Warning System Recommendations

Effectiveness	Description	Recommendation
96%	Driver Display Type	Collision Warning Display, Following-Too-Closely Display. Warning display only 77%.
96%	Moving and Stationary Threats	Both
93%	Warning Time	2.4 seconds
93%	Acquisition Range	≥130 meters
96%	Horizontal Field of Regard	± 8 degrees
92%	Roadway Profile, Vertical Field of Regard	5.3% Grade, 6 degrees
98%	Atmospheric Condition	Clear, Rain, Snow
94%	System Delay Time	≤ 300 milliseconds

ESTIMATION OF BENEFITS

Table 2 shows recommendations and their associated system effectiveness estimates. Overall effectiveness is calculated by the product of individual components (with assumed independence and equal weight). With these assumptions, the overall effectiveness is 65%. True effectiveness is expected to be considerably higher since some of the factors will have a greater weight.

LIMITATIONS

This section presents those items known at this time that may be beyond the capability of a forward looking collision warning system. System users need to be aware of these system limitations.

Head-On -- The forward-looking collision warning system is not required to respond to Head-On collision scenarios. This information may be eliminated from processing in order to reduce nuisance alarms.

Crossing-Path -- The forward-looking collision warning system is not required to respond to crossing path collision scenarios.

Operation beyond sensor limit -- The forward-looking collision warning system is not required to respond when the absolute speed of the host vehicle causes the warning distance to be beyond the sensor range. This problem is compounded on curved roadways and may necessitate a cautionary warning to the driver when the system is operated beyond the sensor limit.

Curves -- The forward-looking collision warning system is not required to respond on very short radius curves (corners), or the warning may be late dependent on the amount of roadway curvature.

Obscured Vehicle -- The forward-looking collision warning system is not required to respond to obscured vehicle scenarios where the lead vehicle suddenly "appears" within the warning distance and inside the finite system delay time. Some obscured vehicle scenarios may not allow the driver enough time to avoid a collision, let alone the system enough time to warn the driver to avoid the collision. An example of an obscured vehicle scenario is when the system and driver's view of the lead (target) vehicle is blocked by another vehicle.

CONCLUSION

Presented are preliminary performance guidelines for a forward-looking collision warning system. On-going work will further define and demonstrate the guidelines through on-road and simulator testing. Current work allows for development of a test bed system to meet the performance guidelines as well as extensive on-road testing and human factors development.

Additionally, this work is examining systems that automatically take control of the vehicle from the driver to avoid an impending collision as well as systems that perform Adaptive (Intelligent) Cruise Control.

These preliminary performance guidelines have been augmented with qualification tests to verify system performance versus these guidelines.

For additional information, or for a complete copy of the preliminary performance guidelines contact Terry Wilson at Frontier Engineering Sciences.

ACKNOWLEDGMENTS

This research is being sponsored by the National Highway Traffic Safety Administration (NHTSA) Office of Crash Avoidance Research (NRD-51) under contract number DTNH22-93-C-07326. The prime contractor is Frontier Engineering Sciences, Advanced Programs Division, 7655 E. Redfield Road, Suite 10, Scottsdale, Arizona 85260. (602) 483-1997, (602) 483-2011 Fax, FEIAPD@AOL.COM Email. Human factors testing is being performed under subcontract by The University of Iowa using the Iowa Driving Simulator. The authors wish to acknowledge: Joe Kanianthra, August Burgett and Art Carter of NHTSA; Steven Jahns, Raj Manikkal, Michael Mollenhauer, Stephen Reinach, Eric Sager and Jasic Lee from the University of Iowa; Barbara Hennessy of Corporate Development Services, Dick Sanneman, Val Prevallet, Jimmy Terry, Steve Ware, Dana Smith, Chet Shupe, Randy Moore, Mike Burns, Jim Cacciatore, Bill Reading, Mike Twitchell, Steve Miller, Joe Petrocelli, Chuck Chase, Paul Bagnasco, Doug Taylor, John Bjornholt and Dwight Brown of Frontier Engineering for their valuable contribution to this research.

APPLICABLE DOCUMENTS

The following list of documents contain closely related information that is critical for development of these guidelines.

Automatic Control System Performance Specifications, "IVHS Countermeasures for Rear-End Collisions", Frontier Engineering Inc., NHTSA Project No. DTNH22-93-C-07326, September 1995.

Task 4 Interim Report for IVHS Countermeasures for Rear-End Collisions, Frontier Engineering Inc., NHTSA Project No. DTNH22-93-C-07326, March 1995.

Preliminary Human Factors Guidelines for Crash Avoidance Warning Devices, NHTSA DOT HS 808 342 January 1996.

New Rules to Increase the Amount of Spectrum Available for Commercial Use, Federal Communications Commission, ET DOCKET 94-124, Adopted: October 1994.

SAE J1879 *General Qualification and Production Acceptance Criteria for Integrated Circuits in Automotive Applications*, October 1988.

SAE J1113 *Electromagnetic Susceptibility Measurement Procedures for Vehicle Components*, August 1987.

SAE J1211 *Recommended Environmental Practices for Electronic Equipment Design*, November 1978.

"Assessment of IVHS Countermeasures for Collision Avoidance: REAR-END CRASHES", NHTSA May 1993

National ITS Program Plan, First Edition March 1995

"Motor Vehicle Crash Involvement: a multi-dimensional problem size assessment", draft report, DOT NHTSA NRD-53 1995

"Rear-End Crashes: Problem Size Assessment and Statistical Description" USDOT, NHTSA, OCAR, May 1993.

A. Burgett "DOT's Approach to ITS Safety Evaluations", Presented at the ITS America Workshop on ITS Safety Evaluations, May 1, 1995.

970457

Safety Evaluation Methodology for the Intelligent Cruise Control Field Operational Test

Joseph S. Koziol
Volpe National Transportation Systems Center

Vaughan W. Inman
Science Applications International Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper describes the evaluation methodology for the Intelligent Cruise Control (ICC) Field Operational Test (FOT). The primary purpose of the evaluation is to assess safety impacts of the ICC system. Other benefits, such as convenience and comfort, as well as impacts of the system, e.g., fuel consumption and emissions, are also being assessed.

The ICC system incorporates a forward looking sensor and a headway controller with a conventional cruise control system, to automatically maintain a headway (with accelerator and downshift control inputs) between the ICC-equipped vehicle and a vehicle that precedes the equipped vehicle. The FOT will collect daily usage experiences from up to 162 lay drivers, each of whom will drive one of 10 ICC equipped vehicles for periods of 2 or 5 weeks. Data collection is scheduled to last 12 months. Each of the 10 vehicles has a data collection system that will support the evaluation. The data collection system will collect video data and a variety of vehicle performance measures.

Six evaluation goals have been identified. These include the evaluation of system benefits to users, system performance, institutional and legal issues, user acceptance, system costs, and ICC impacts on the transportation system. This paper focuses on the safety evaluation process, describing the safety objectives, hypotheses to be tested, measures of effectiveness, measures of performance, data sources and analysis methods.

INTRODUCTION

PURPOSE - This paper presents the evaluation methodology for the Intelligent Cruise Control (ICC) Field Operational Test. The purpose of the evaluation is to assess, based on data collected from the field operational test, actual and potential safety impacts of the ICC system. Other impacts of the system will be evaluated if they affect safety impacts indirectly (e.g., comfort, convenience, willingness to purchase, and willingness to use the system), or if those impacts are easily assessed as corollaries to the safety findings (e.g., fuel consumption, emissions). The following safety questions will be addressed:

1. Do drivers drive more, or less, safely with the system than without it, in ways related to the system?
2. Do vehicles equipped with the system have fewer, or more, collisions than vehicles without the system?
3. If all vehicles were equipped with the system, would there be a decrease, or increase, in the total number of collisions and collision-related injuries?

BACKGROUND - Recently, NHTSA signed a cooperative agreement with the University of Michigan Transportation Research Institute (UMTRI) to conduct a field operational test of an ICC system.[1] Other parties contributing to the field operational test are Leica AG, the Michigan Department of Transportation, and Haugen Associates.

With ICC, a forward looking sensor and intelligent controller are incorporated with normal cruise control to automatically maintain a headway between the ICC equipped vehicle and a vehicle that precedes the equipped vehicle. When other traffic are encountered, ICC equipped drivers are provided some relief from engaging, disengaging, or manually resetting speed, as might more often be necessary with conventional cruise

control. When not in traffic, the ICC system functions like conventional cruise control.

ICC was not developed as a safety device, and its potential effects on safety are being studied now. Plausible safety benefits include reduced driver fatigue and a reduction in rear-end collisions that may result from both (a) better maintenance of safe following distances, and (b) the alerting function that ICC's automatic response to closing headway may provide. However, by automating headway maintenance, ICC may also change demands on the driver in ways that adversely affect safety. Plausible adverse effects include: (a) delayed driver response to traffic because of over-reliance on the ICC system, and (b) driver disorientation due to sudden accelerations or decelerations as a result of the system losing lock on a preceding vehicle, or incorrectly locking on to a vehicle not in the user's lane of travel.

ICC FIELD OPERATIONAL TEST[2]

For the ICC Operational Field Test, ten 1996 Chrysler Concordes have been equipped with identical ICC systems by UMTRI. The ICC system consists of the standard Chrysler cruise control, Leica sensors, intelligent (adaptive) controller, and a data acquisition system. The data acquisition system will collect video data and a variety of system performance measures. Among the system performance measures that will be sampled 10 times per second (10 Hz) are:

- Range to the preceding vehicle
- Rate of change in range to preceding vehicle
- Vehicle velocity
- The headway time selected by the driver
- Curve radius
- Throttle setting

The video data will capture a 60 degree field of view scene in front of the vehicle. Video episodes of 2 to 30 seconds will be stored when predetermined criteria are met for certain driving events such as braking and near encounters. The stored video episodes will include a portion of time before the triggering event. The video data will later be used to assist in evaluating the effects of the ICC system on driving performance and safety.

Up to one hundred sixty-two volunteers will be recruited to drive the ICC equipped Chrysler Concordes each for 2 or 5 week durations in a naturalistic setting. No experimenters will be on board. One hundred eight drivers will receive vehicles for 2 weeks and the remaining 54 drivers will receive vehicles for 5 weeks.

Both groups of drivers will experience one week of driving the Concorde in a standard configuration (no ICC) before ICC functions are made available. The one week without ICC is intended to provide baseline data for driving performance without ICC.

One and four week exposures to ICC are included in the research design to enable assessment of longer term changes in behavior as a function of experience with ICC usage. Other subject variables in the research design are:

- Driver age
- Previous cruise control usage

Drivers in three age groups will be recruited: 20 to 30; 40 to 50 ; and 60 to 70 years of age.

Among the 108 drivers who participate for 2 weeks, half will be from among drivers who say that they regularly use conventional cruise control. The other half of this group will be from among drivers who say they do not use cruise control. All 54 drivers who participate for 5 weeks will be selected from among those who say that they regularly use conventional cruise control.

ICC EVALUATION

The ICC evaluation is being conducted by US DOT's Volpe National Transportation Systems Center (Volpe Center) with support from Science Applications International Corporation (SAIC).

ICC EVALUATION GOALS - In accordance with DOT FOT evaluation guidelines[3], six ICC evaluation goals have been identified:

1. Evaluate System Benefits to Users
2. Evaluate System Performance
3. Evaluate Institutional and Legal Issues
4. Evaluate User Acceptance
5. Evaluate System Costs
6. Evaluate ICC Impacts on the Transportation System.

ICC EVALUATION APPROACH - Separate evaluation studies will be conducted to address each of the goals.

Benefits Study - The *Benefits Study* will evaluate three classes of benefits that may derive from the use of the ICC system:

- Changes in driving performance that are potentially safety-related
- Changes in accident probability, accident severity, and in the frequency and severity of "near accidents"
- Non-safety related changes in driving performance.

The study will utilize digital vehicle data; video data; data from questionnaires, focus groups, and debriefings; and ICC accident investigation data if there are any crashes.

System Performance Study - In addition to providing a complete documentation of the ICC system design, the *System Performance Study* will evaluate the ICC system from three perspectives:

- IR sensor characterization
- Driver interface
- Overall integrated ICC system characterization.

Institutional and Legal Issues Study - The objective of this study is to identify the non-technical impediments, commonly called institutional and legal issues, that are likely to arise during the ICC field operational test. Identification of non-technical impediments during the field operational test may serve to minimize the ramifications of these issues on the commercial deployment of ICC and on the implementation of other ITS operational tests.

The *Institutional and Legal Issues Study* will attempt to answer four questions:

1. What institutional and legal impediments were encountered in establishing partnerships, deploying ITS services and products during the operational test, and developing and deploying ICC technologies?
2. Where in the life cycle of the operational test did these impediments occur?
3. What were the causes of these impediments and how were they overcome?
4. What lessons can be learned in dealing with these impediments that can be applied to the full deployment of ICC and to other operational tests?

Interviews and questionnaires will be the data sources for the study. Subjects of the interviews and questionnaires will include partners, evaluators, and other parties from government and industry who are involved in the operational test, as well as parties who are not partners in the field operational test, but may be key players in the deployment of ICC systems.

User Acceptance Study - The *User Acceptance Study* will evaluate the volunteer drivers' acceptance of the ICC system from three perspectives:

1. Actual Usage — how much, when, and where the drivers use the system
2. Stated Preference — what the drivers say they like and dislike about the system
3. Value — what the drivers think the system is worth.

Questionnaire, focus group, and debriefing data from ICC driver participants will serve as the data sources for addressing these questions.

Cost Study - The purpose of this study is to identify the projected costs to deploy the ICC systems. The cost estimate will include the initial costs to procure, install and calibrate a typical system (non-recurring costs) as well as the annual operations and maintenance (O&M) costs (recurring costs).

System Impacts Study - The purpose of the *System Impacts Study* is to assess the potential impacts of ICC at various levels of market penetration. The study will assess impacts on:

- Safety
- Traffic flow and/or capacity
- Fuel consumption
- Emissions.

The Impacts Study will use the following data sources:

- In-vehicle data logs
- User acceptance study data
- System performance study data
- Other ICC vehicle performance measures
- Traffic data logs
- Loop-detector data
- National Accident Statistics
- EPA/MOBILE5A.

Data analysis will include a number of modeling techniques and include the use of the INTEGRATION traffic model.

SAFETY EVALUATION - The remainder of this paper will focus on the safety evaluation process for the ICC system.

Table 1 summarizes the technical approach to the evaluation of the safety benefit objectives of 1) determining whether drivers drive more or less safely with the ICC system than without it in ways related to the system, and 2) assessing driver perceptions of safety and performance.

Several hypotheses will be tested under these objectives as indicated in the table. To the extent that data are available, each of the hypotheses will be examined for specific driving scenarios. The driving scenarios that are being considered for the ICC FOT are shown in the appendix.

Use of ICC reduces driver response time relative to driving without ICC - The hypothesis that driver response times will be reduced with ICC compared to response times driving without ICC (i.e., both without cruise control and with conventional cruise control) will be tested for events that cause drivers to brake rapidly or to make evasive lateral maneuvers. Because the ICC will cause noticeable deceleration when there is a high rate of closure between preceding and host vehicles, the ICC system has the potential to alert drivers to the need to brake earlier than they might detect this requirement on their own.

For rapid braking and lateral maneuvers, when captured on video, the evaluator will examine the video record for the event that precipitated the response. It is hypothesized that the response to the event will occur earlier with ICC than with conventional cruise control or without cruise control. Several assumptions underlie this hypothesis. One assumption is that for any type of event, rapid reactions sufficient to trigger video capture will occur regardless of the lag between event start and driver response. Where rapid response is required because the driver responded late, this assumption could be wrong. If there is evidence for a large and disproportionate number of late responses in conjunction with a cruise control mode (i.e., ICC, conventional cruise control, or no cruise) then the reaction time data used to test this hypothesis should be viewed with caution. Nonetheless, a disproportionate number of late responses in a particular mode will in itself be an important finding.

Because of limitations in the data recording, driver reaction time differences between cruise control conditions that differ by less than ¼ sec will be difficult, if not impossible, to identify. The difficulty in detecting small differences results because the video may be captured at a rate of 10 frames per second. Thus for any video frame, times can not be resolved below 100 milliseconds. Other temporal ambiguities, such as vertical and horizontal video sweep, and computer interrupt priority, compound the resulting imprecision. Therefore, this test is for relatively large response time differences, not for subtle differences. Because range data, when the sensor is tracking a lead vehicle, will be recorded at a rate of 10 hertz, it may be possible to resolve reaction times to about 100 milliseconds. Even 100 milliseconds is a relatively large difference relative to operator response capabilities.

Headway varies as a function of cruise control mode, road class, traffic density, and road geometry - It is hypothesized that, because use of automated headway control is convenient, drivers will use ICC and, as a result, have longer average headways than they otherwise would maintain. In addition to road characteristics, driver age is included as a factor. Although the direct relationship between any given headway time and safety is not well understood, there is an implicit assumption that within the normal range of headways, longer headways are safer.

ICC decreases driver use of brake pedal relative to driving without ICC - Because the ICC automatically controls speed with accelerator and downshift control inputs, ICC users should be able to moderate headway with less brake use. With conventional cruise control, the driver must either disengage the system with the brake, or may manually decrease the selected speed. Manual decreases in selected speed makes inputs to the accelerator but not the transmission and therefore could be expected to require brake intervention more frequently than would the ICC which, in addition to not requiring manual speed setting adjustments, can request downshifts for greater deceleration authority.

Brake pedal usage will be correlated with driver perceptions of effects on performance and with close call frequency and severity.

ICC decreases braking force required - Like conventional cruise control, the ICC system requires the driver to intervene by braking when headway closure rates are greater than the driver will tolerate. Unlike conventional cruise control, the ICC system automatically attempts to compensate for decreasing headway by coasting or downshifting. Therefore the rate of closure should, on average, be less when the driver intervenes. Given that the discrepancy between desired headway and actual headway should be less on average with ICC, driver interventions with the brake should require less force and less duration. Driver age will be a factor in the analysis.

Table 1. Approach to Evaluation Safety Benefits to Individual Users.

Objective	Hypothesis	Measure of Effectiveness	Measure of Performance	Data Source	Method of Analysis
Determine if drivers drive more or less safely with the ICC system than without it, in ways related to the system	Use of ICC reduces driver response time relative to driving without ICC	Reaction time to driving events to which drivers respond by braking	Latency to initiate braking	• Forward video record • Range Rate • Range • Brake Pedal • CONCERN button • Accelerometer	For all braking events when deceleration is high, determine latency of brake pedal change following beginning of rapid change in range rate
		Reaction time to driving events to which drivers respond with lateral avoidance maneuvers	Latency to initiate avoidance maneuver	• Forward video record • YawRate • CONCERN button	For all high YawRates that are the result of an avoidance maneuver, determine latency from initial stimulus to driver avoidance response
	Headway varies as a function of cruise control mode (ICC, CCC, no cruise control), road class, traffic density, and road geometry.	• Headway	• Range/velocity	• Range • Velocity • Tracking target	• Analysis of variance.
	ICC decreases driver use of brake pedal relative to driving without ICC	Frequency of Brake Use	Brake pedal depressions per kilometer by road class	• Road class • Time of day • Brake Pedal	Brake pedal usage analyzed as a function of road class and traffic density. Loglinear analysis.
	ICC decreases braking force required and this varies as a function of road class and traffic density	Brake pedal force	RMS G force while brake applied	• Road class • Time of day • Velocity • Accelerometer	Analysis of Variance
	Driver workload is reduced when ICC is used relative to driving without ICC	Subjective workload	Subjective workload rating	• Questionnaire • Observer probes	Descriptive statistics and analysis of variance
	ICC users show fewer indications of inattention to the roadway	• Yaw rate variability[1]	• Frequency of large yaw rate variance	• Yaw rate sensor	• Analysis of variance • Log-linear analysis
Assess driver perceptions of performance	Users perceive driving performance and safety to be enhanced by ICC	• Perception of driving performance • Perception of driving safety	Perception of: • Speed maintenance. • Awareness of other vehicles • Safety	• Questionnaire • Focus groups • Debriefing • Real-time observations	• Descriptive statistics • Analysis of variance

[1] It would be desirable to filter out lane changes and turns.

Driver workload is reduced when ICC is used relative to driving without ICC - Mental workload is a hypothetical construct in which it is posited that drivers have limited mental resources, and that drivers can perceive when the amount of resources required approach or exceed available resources. Perceived mental workload will be assessed from user questionnaires and focus group responses.

ICC users show fewer indications of inattention to the roadway than drivers not using ICC - Whenever a manual, or perceptual-motor task is automated, it is prudent to examine whether the system operator will continue to maintain a level of awareness of system operation so that operator intervention occurs in a timely and appropriate fashion. Therefore, the ICC evaluation will assess whether driving behaviors consistent with inadequate operator supervision are evident. With the ICC system, very late or inappropriate responses to roadway events would indicate the possibility that the driver was not adequately attentive to the driving task. The evaluator will examine the video data for signs of behavior consistent with inattention: hard braking that follows the initiating event after an excessive delay, or evasive maneuvers that follow an initiating event after an excessive delay. This analysis requires that the event that precipitated the driver response be identifiable in the video record.

Users perceive driving behavior to be enhanced by ICC - At the conclusion of their ICC experience, drivers will be debriefed and given a questionnaire. Later, the same drivers will return to participate in focus group sessions that address their ICC experience. Data from all three sources will be used to address whether the drivers perceived enhancements or detriments to safety that they attribute to use of the ICC system. In addition to perceptions of safety effects per se, users will be asked if they perceived their driving behaviors to change in other ways. Some behavior changes that the drivers may not perceive to be related to safety may have important safety implications. Although drivers may not think of road class and speed selection choices in terms of safety, these factors may have implications for accident exposure and severity of potential injuries.

Table 2 presents the approach to determining whether any accidents involving ICC vehicles are related to ICC use. The evaluation will also examine the frequency of close calls recorded in the video data. Three hypotheses will be tested:

Table 2. Approach to the Evaluation of Reduction in Accident Occurrence and Accident Severity Attributable to the User of ICC.

Objective	Hypothesis	Measure of Effectiveness	Measure of Performance	Data Source	Method of Analysis
Determine if any accidents of ICC vehicles are related to the use of the ICC system	• Accidents (if any) involving ICC equipped vehicles are not related to ICC use • ICC use results in fewer close calls • Close calls encountered when using ICC are qualitatively different from close calls encountered without ICC	• Cause of accident • Frequency of close calls	• Factors contributing to accident • Close call frequency • Close call severity	• Accident investigation • Video analysis • Observer recordings	• Accident analysis • Chi-square tests. • Close calls categorized and analyzed as a function of type of close call • Qualitative analysis of severity, and situational factors • Frequency of close calls related to overall exposure

Accidents (if any) involving ICC equipped vehicles are not related to ICC use - Because there will be only 10 ICC vehicles, and because across all participants in this study it is unlikely that more than a million miles will be logged, it is highly unlikely that a sufficient number of accidents will occur to support statistical analysis, either for testing whether accidents are more or less likely with ICC than without it, or whether ICC is associated with more or less severe accidents. However, the *Benefits Study* detailed study plan will provide a procedure for documenting accidents. The documentation will specifically address whether ICC was being used at the time of the accident, and whether ICC could have been a contributing factor. Procedures will be detailed for rating the severity of accidents, with special attention to the severity of rear end collisions.

ICC use results in fewer close calls. Close calls encountered when using ICC are qualitatively different from close calls encountered without ICC - The evaluator will examine the video data associated with extreme braking, extreme lateral accelerations, and CONCERN button pressings to document occurrences of close calls, or near accidents. Procedures used for the TravTek Camera Car Study and documented in the test plan for the Safety Evaluation of the ADVANCE Project will be used to quantify close calls.[4,5] Because the video analysis will only be done for close calls that resulted in extreme driver reactions or pressings of a CONCERN button that is provided in the vehicle for this purpose, several kinds of close calls may be missed entirely, or be underrepresented in the sample. Missed or underrepresented close calls may include: close calls that do not require extreme responses; close calls which the driver detects but then fails to press the CONCERN button; and close calls that the driver does not detect. Whereas the close call analysis may be suggestive of increased or decreased accident risk with ICC, given the dependence of the analysis on triggers for video data collection, this analysis, by itself, will not be definitive. However, close call findings, when considered with the body of evidence from tests of other safety hypotheses, are expected to provide a robust body of evidence regarding the safety effects of the ICC system.

DATA SOURCES AND DATA FLOWS - Much of the evaluation data will be collected onboard the ten ICC vehicles. These and other sources of data are indicated in Figure 1. Traffic data will be determined from vehicle-based video records and digital records derived from this video. In addition, a separate video will be captured by placing a camera on overpasses overlooking various freeway locations in the Michigan test area. GIS software will be used to identify roadways and road classes from GPS coordinates recorded in the time history and event log files. The data referred to as "digital" includes only the GPS, time history, and event files from the vehicles. This digital data will be used to characterize driving performance as a function of road class, road geometry, previous cruise control use, and ICC experience. The subjective data refers to the data generated from recruiting, questionnaire, debriefing and focus group activities. Comparable systems data will be obtained to project the costs to deploy a typical ICC system. Interviews of the partners will provide the information for the Institutional Issues Studies.

SUMMARY

This paper introduces the concepts involved in the safety evaluation process for the ICC Field Operational Test. The ICC system is first briefly described. Then the evaluation goals are delineated. Specific objectives, hypotheses, measures of effectiveness, measures of performance, data sources and methods of analyses are described to evaluate safety benefits to individual users and to evaluate reduction in accident occurrence and accident severity attributable to the user. The safety evaluation tasks will be carried out as an integral part of the overall ICC evaluation. The ICC Field Operational Test will cover approximately a period of one year with completion targeted for August, 1997.

REFERENCES

1. National Highway Traffic Safety Administration (1995). Cooperative Agreement Number DTNH22-95-H-07428.

2. University of Michigan Transportation Research Institute (1996). *Test Definition and Project Plan.* Unpublished.

3. MITRE. (1993). *Generic IVHS Operational Test Evaluation Guidelines.* Working paper prepared for the Federal Highway Administration.

4. Dingus, T. A., McGehee, D., Hulse, M. C., Jahns, S., Manakkal, N., Mollenhauer, M., and Fleischman, R. N. (1995). *TravTek Evalaution Task C_3 — Camera Car Study.* Federal Highway Administration, Technical Report No. FHWA-RD-94-076, Washington, DC.

5. McGehee, D. V., Dingus, T. A., Mollenhauer, M. A., and Inman, V. W. (1996). *ADVANCE Safety Evaluation Test Plan.* Unpublished Document: The Volpe National Transportation Systems Center, Cambridge, MA

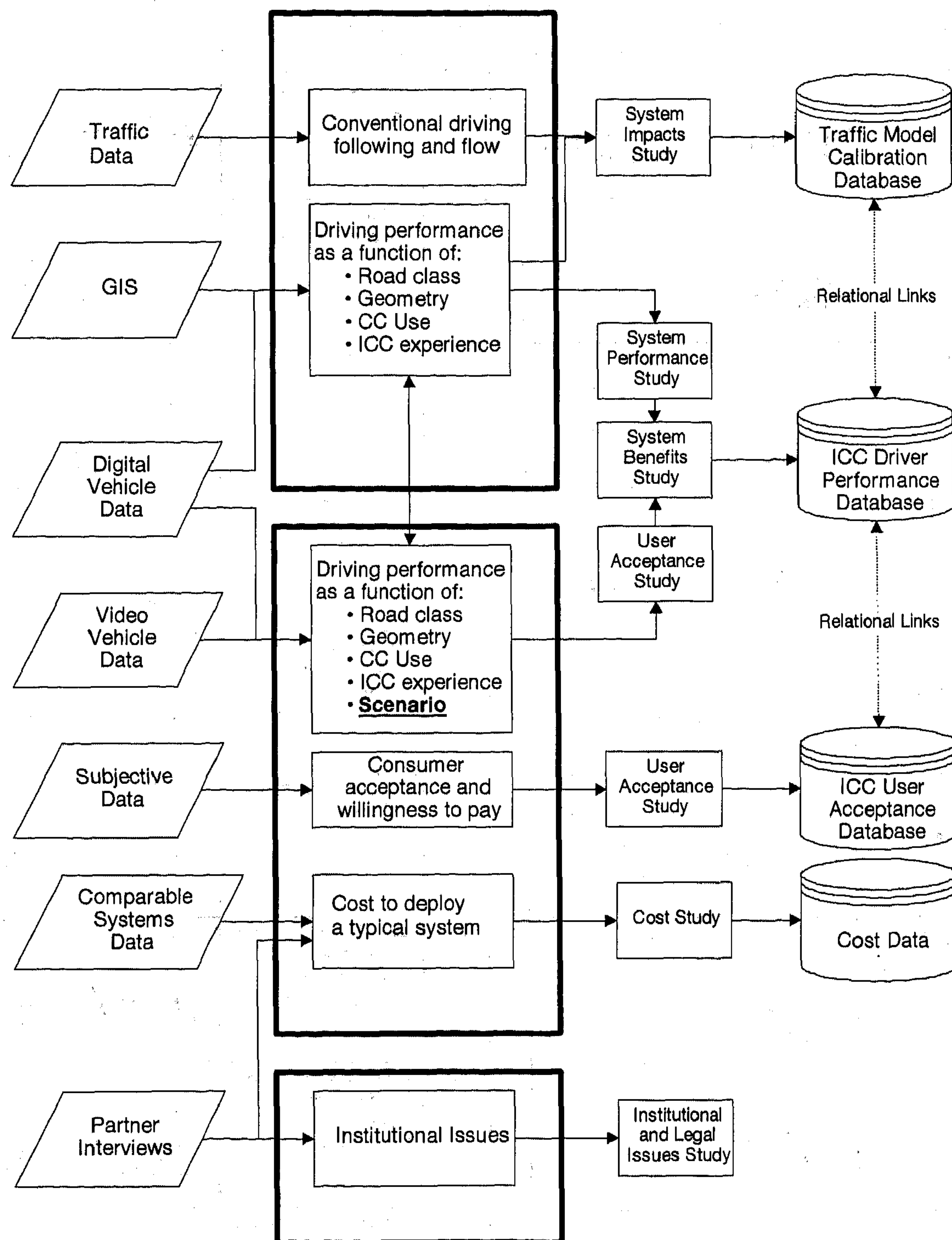


Figure 1. Evaluation Data Flows.

APPENDIX

ICC DRIVING SCENARIOS

No.	Driving Scenario	Schematics
1A	Approaching Lead Vehicle Going Straight	
1B	Approaching Lead Vehicle Turning Right	
1C	Approaching Lead Vehicle Turning Left	
2A	Vehicle or Object Stopped in Roadway, Going Straight	
2B	Vehicle or Object Stopped in Roadway, Negotiating a Curve	
3A	Following a Lead Vehicle in a Curve/Exit Ramp	
3B	Acquiring a Lead Vehicle in a Curve/Exit Ramp	
4A	Following a Lead Vehicle Over a Hill	
4B	Acquiring a Lead Vehicle Over a Hill	
5A	Following a Lead Vehicle in a Roadway Dip	

No.	Driving Scenario	Schematics
5B	Acquiring a Lead Vehicle in a Roadway Dip	
6A	Cut-In of Passing Vehicle	
6B	Cut-In of Blind Spot Vehicle	
7A	Merging with Slower Lead Vehicle in Merged Lane	
7B	Merging with Faster Following Vehicle in Merged Land	
7C	Overtaking Merging Vehicle	
8A	Lane Change to Pass Lead Vehicle Going Straight, Moving Out of Lane	
8B	Lane Change to Pass Lead Vehicle Going Straight, Moving Into Lane	
8C	Lane Change to Pass Lead Vehicle Turning Left	
8D	Lane Change to Pass Lead Vehicle Going Straight, Acquires Slower Moving or Stopped Vehicle in Adjacent Lane	

No.	Driving Scenario	Schematics
8E	Lane Change to Pass Lead Vehicle Going Straight, Overtaken by Faster Moving Vehicle in Adjacent Lane	
9A	Unplanned Lane Change, Other Vehicles Present	

No.	Driving Scenario	Schematics
9B	Unplanned Lane Change, No Other Vehicles Present	

LEGEND

LEAD VEHICLE

ICC EQUIPPED VEHICLE

GOING STRAIGHT, CONSTANT SPEED

GOING STRAIGHT, LOWER SPEED

GOING STRAIGHT, HIGHER SPEED

GOING STRAIGHT, SLOWING OR STOPPING IN TRAFFIC LANE

GOING STRAIGHT, STARTING IN TRAFFIC LANE

GOING STRAIGHT, STOPPED IN TRAFFIC LANE

PASSING

TURNING RIGHT, CONSTANT SPEED

TURNING RIGHT, SLOWING OR STOPPING IN TRAFFIC LANE

TURNING RIGHT, STARTING IN TRAFFIC LANE

TURNING LEFT, CONSTANT SPEED

TURNING LEFT, SLOWING OR STOPPING IN TRAFFIC LANE

TURNING LEFT, STARTING IN TRAFFIC LANE

NEGOTIATING A CURVE

CHANGING LANES

MERGING

970490

Facial, Periorbital and Ocular Injuries Related to Steering-Wheel Airbag Deployments

Donald F. Huelke, Lawrence W. Schneider, Matthew P. Reed, and Ryan J. Gilbert
University of Michigan Transportation Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

To determine the frequency of facial injuries from steering-wheel airbag deployments, 540 consecutive steering-wheel airbag deployments, investigated by the University of Michigan Transportation Research Institute (UMTRI) personnel, were reviewed. About 1 in 3 drivers sustain an injury to the face. Injuries to the area surrounding the eye (periorbital) or to the eyeball (ocular) rarely occur. The frequencies of facial or ocular injuries are the same for belted and unbelted drivers. Drivers of short stature had a higher frequency of facial injury. Females sustained ocular injuries more frequently than males. Untethered airbags were not overly involved in drivers with an ocular injury. No specific make or model car were overly represented in the ocular injury cases.

INTRODUCTION

The effectiveness of steering-wheel airbags has been well documented in the medical and engineering literature.[1-8] When properly used as a supplemental restraint with the lap-shoulder belt, the steering-wheel airbag reduces driver fatalities by about 23% in frontal and front angular crashes.[6] However, the airbag has also been implicated as the cause of certain injuries. For the most part, these injuries are minor, including abrasions, contusions, and lacerations of the upper extremity, face, and torso areas. More serious airbag-induced injuries have also been observed, including hand, finger, and forearm fractures or dislocations have been reported along with injuries to the torso.[9-16] Numerous articles have documented airbag-induced ocular injuries such as corneal abrasions, hyphema, angle recession, lens subluxation, traumatic iritis, viteous hemorrhage, retinal tears, commotio retinae, macular hole, chorodial rupture and alkali keratitis.[16-40]

The steering-wheel airbag module consists of the pyrotechnic inflator, the airbag, the outer module covering, and the module door or flap. When the airbag sensor system detects a crash for which the airbag should deploy, an electrical signal triggers the airbag pyrotechnic inflator. The pyrotechnic inflator produces gas which begins to inflate the folded airbag. As the gas pressure rapidly increases, the airbag punches through the module door. This occurs about 5 to 10 ms after the airbag is triggered. In order for the airbag to be fully deployed before any significant forward excursion of the driver's body occurs, airbags must be fully inflated in 40-50 ms. These high inflation rates can result in airbag initial fabric velocities of 320 km/h or more.[41,42] Vents on the back side of the airbag allow the gasses to flow out, so that the airbag deflates completely in about 2 seconds (The pressure inside the airbag returns to atmospheric within about 0.2 seconds).

During deployment, the trajectory of the airbag fabric is determined by the type of fold pattern and the presence or absence of tethering. In an untethered airbag, the filled airbag volume is about 60 liters and has a full and rounded convex surface (Fig. 1). The tethered airbag usually has a smaller filled airbag volume of about 50 liters and has several strips of material sewn to the center portion of the airbag surface (Fig 2). These tethers restrict the excursion of the airbag fabric during deployment and control the shape of the airbag during filling. When the airbag is completely filled, the tethered region of the airbag results in a depressed and

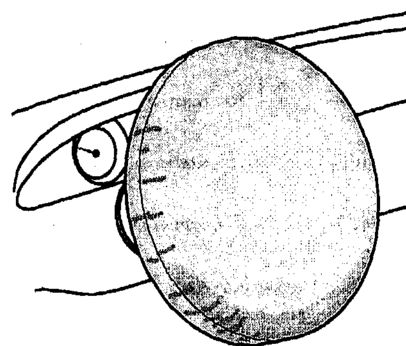


Figure 1. Steering wheel untethered airbag

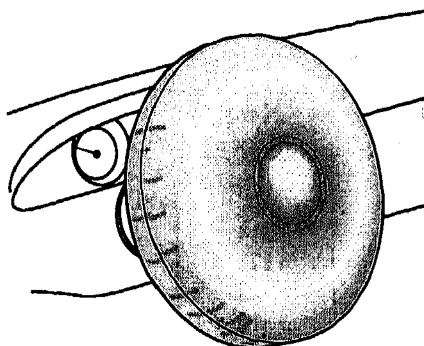


Figure 2. Steering wheel tethered airbag

flattened area about 6 inches in diameter in the center of the airbag surface facing the driver. Additionally, the type of bag folding pattern influences the airbag deployment kinematics and may aid in reduction of some injuries to the face or torso.[43] Both tethered and untethered airbags are used in various makes and models of vehicles.

This paper presents results from analysis of crashes involving steering-wheel airbag deployments with regard to the frequency of steering-wheel induced injuries to the face, periorbital area, and eyes. In the analysis that is presented, ocular (eye) injuries are a subset of periorbital injuries, which are a subset of facial injuries.

MATERIALS AND METHODS

Accident investigations are routinely conducted by the University of Michigan Transportation Research Institute (UMTRI) with the general goal of determining the causes of injuries and deaths in motor vehicle accidents. In all of the UMTRI investigations, the crush damage of the vehicle is measured and photographed, occupant contacts within the vehicle interior are noted and documented, and other vehicle and driver characteristics are obtained and recorded. Injury data are obtained for case-vehicle occupants from medical records of cooperating hospitals. A telephone interview is usually conduced with the driver to obtain additional crash or injury details. Since cases selected for investigation generally involve a tow-away crash with an injured occupant, the resulting data sets tend to be biased toward the injury-producing crash.

As more vehicles have been equipped with steering-wheel airbags, crashes involving steering-wheel airbag deployments have occurred with increasing frequency. A case database comprised of 540 UMTRI-investigated crashes involving steering-wheel airbag-deployments was available for analysis in this study. Each case was individually reviewed to determine if the facial injuries to the driver were caused by the airbag deployment. The data were analyzed and tabulated to establish the frequencies of facial, periorbital, and ocular injuries for different conditions of airbag tethering, belt restraint use, driver gender and stature, and vehicle make and model. The presence of an airbag tether was determined by reviewing the photographs of the steering-wheel or from inquiries to the vehicle manufacturers. In tabulating the data, each driver with an airbag induced facial injury was counted only once, even if there were multiple facial areas involved. Of the drivers with an airbag-induced facial injury, the periorbital injury cases were extracted as a subset, and ocular trauma cases were identified and separately reviewed as a subset of the periorbital injuries. The periorbital area includes the eyelids, eyebrow, adjacent nose, medial and lateral canthal areas and infraorbital cheek. "Ocular" refers to the eyeball only.

RESULTS

A. FREQUENCY OF INJURY--Table 1 displays the frequencies of driver injuries related to steering-wheel airbag deployments in the UMTRI database. An airbag injury to any body region was found in 63% of the drivers. Most of these injuries are at the AIS-1 level and include abrasions, contusions, small lacerations or errythemia. Approximately half of the drivers who were injured by the airbag sustained an airbag-induced facial injury, but only 27 of all drivers (5%) sustained a periorbital or ocular injury caused by contact with the airbag. Eleven drivers sustained an injury directly to the eyeball, and only two of these sustained an ocular injury without periorbital involvement.

B. INFLUENCE OF DRIVER STATURE ON THE FREQUENCY OF FACIAL INJURIES--Table 2 shows the frequency of facial, periorbital, and ocular injuries by driver stature. As indicated above, 32% of all drivers sustained a facial injury from the airbag. The percentage of drivers receiving facial injuries in each stature range, decreases with increasing stature (Table 2). For the short drivers (150-159 cm), approximately half sustained facial injuries. In contrast, for the tall drivers, only about 9% sustained facial injuries. Most of the facial injuries were sustained by drivers who were less than 189 cm in stature and drivers with ocular injuries, with or without periorbital injuries, were all less than 179 cm tall. However, seven of the 11 drivers with ocular injuries were of average stature (170 -179 cm).

C. INJURY FREQUENCY WITH SEAT BELT USE--Table 3 shows a breakdown of facial injuries by belt restraint use. While the number of restrained drivers sustaining facial injuries is significantly higher than the number of unrestrained drivers, the percentages of facial injuries are very similar for the two groups. This is also true for ocular injuries. Nine restrained drivers receiving ocular injuries and two others were unrestrained, but both represent only 2% of their respective group. Although few in occurrence, the frequency of periorbital injury, with or without an ocular injury, for the unrestrained drivers, was twice that of the restrained drivers (8% vs. 4%) even though restrained drivers outnumbered the unrestrained by 6 to 1.

Table 1

Driver Airbag and Deployment Injury Frequency

Injury	Total N	Total %	N for AIS = 1	N for AIS > 1
All Drivers with Airbag Deployments	540	100	--	--
No Injury	198	37	--	--
Airbag-induced injury (in any body region)	342	63	314	28
Facial Injuries	174	32	172	2
Periorbital + ocular injuries	27	5	27	0
Periorbital w/o ocular injuries	16	3	16	0
Periorbital with ocular injuries	11*	2	11	0

*Two of these drivers had an ocular injury without periorbital involvement

Table 2

Frequency of Facial, Periorbital, and Ocular Injuries by Stature (cm)

Injury	150 -159 N (%)	160 -169 N (%)	170 -179 N (%)	180 -189 N (%)	190 -199 N (%)	Total N (%)
None	24 (50)	92 (64)	111 (62)	106 (80)	20 (91)	353 (67)
Total facial	24 (50)	51 (36)	68 (39)	26 (20)	2 (9)	174 (32)
Periorbital only	1 (2)	8 (6)	5 (3)	2 (2)	-	16 (3)
Ocular w & w/o periorb.	1 (2)	3 (2)	7 (4)	-	-	11* (2)
Other facial	22 (46)	40 (28)	56 (32)	24 (18)	2 (9)	144** (27)
Total	48 (100)	143 (100)	179 (100)	132 (100)	22 (100)	524**(100)

*Three facial injury cases do not include stature data.
**16 cases did not have stature data.

Table 3

Driver Restraint Usage and the Occurrence of
Facial, Periorbital, and Ocular Injury

Facial Injury	Unrestrained N (%)	Restrained N (%)	Total N (%)
None	54 (66)	312 (68)	366 (68)
Total facial	28 (34)	146 (32)	174 (32)
Periorbital only	5 (6)	11 (2)	16 (3)
Ocular w/wo periorbital	2 (2)	9 (2)	11 (2)
Other facial injuries	21 (26)	126 (28)	147 (27)
Total	82 (100)	458 (100)	540 (100)

D. FREQUENCY OF INJURY BY GENDER--As expected from the results for stature, the incidence of facial, periorbital, and ocular injury is different for male and female drivers (Table 4). Of the 292 male drivers involved in airbag deployments, 24% sustained a facial injury from the airbag versus 42% of the 247 female drivers . Females also sustained a disproportionate number of injuries to the periorbital and ocular areas. Periorbital and/or ocular injury occurred to 7% of female drivers compared to 3% of the males, while ocular injury alone occurred to 3% of the female drivers compared to 1% of the male drivers.

Table 4

Occurrence of Facial, Periorbital and Ocular Injury
of Male and Female Drivers

Facial Injury	Males N (%)	Females N (%)	Total
None	222 (76)	144 (58)	366 (68)
Total facial	71 (24)	103 (42)	174 (32)
Periorbital only	6 (2)	10 (4)	16 (3)
Ocular w/wo periorbital	4 (1)	7 (3)	11 (2)
Other facial injuries	61 (21)	86 (35)	147 (27)
Total	293 (100)	247 (100)	540 (100)

Table 5

Occurrence of Facial, Periorbital and Ocular Injury
From Tethered and Untethered Airbags

Facial Injury	Tethered N (%)	Untethered N (%)	Total
None	287 (72)	79 (56)	366 (68)
Total facial	111 (27)	63 (44)	174 (32)
Periorbital only	13 (3)	3 (2)	16 (3)
Ocular w/wo periorbital	5 (1)	6 (4)	11 (2)
Other facial injuries	93 (23)	54 (38)	147 (27)
Totals	398 (100)	142 (100)	540 (100)

E. EFFECT OF AIRBAG TETHERING AND VEHICLE MODEL ON OCULAR INJURY--Table 5 shows the frequencies of facial, periorbital, and ocular injuries by airbag tethering. As indicated, most of the cases involved tethered airbags, but it is also clear that the occurrence of facial and ocular injuries differs for the two airbag conditions. Airbag-induced injury to any facial area, occurred to 27% of the drivers with tethered airbags and to 44% of the drivers exposed to untethered airbags. Of the drivers with ocular injuries, there is an equal number for untethered and tethered airbags, but the frequency of occurrence of the ocular injuries is lower with tethered steering-wheel airbags. The frequencies of periorbital injuries for the two types of airbags are similar (3% vs. 2%).

Table 6 indicates the presence of an airbag tether, and the make, model, and year of the vehicle for the cases with ocular injuries. The occurrence of these injuries is distributed over a wide range of vehicle and airbag types.

Table 6

Ocular Trauma Cases - Airbag Type and Vehicle Information

Tethering	Year	Make	Model
Tethered	1987	Ford	Tempo
Tethered	1991	Nissan	Infiniti Q-45
Tethered	1993	Saturn	SL1
Tethered	1993	Ford	Taurus
Tethered	1994	Mercury	Sable
Untethered	1988	Dodge	Daytona
Untethered	1989	Dodge	Daytona
Untethered	1989	Dodge	Daytona
Untethered	1994	Toyota	Corolla
Untethered	1994	Chevrolet	Corsica
Untethered	1996	Chevrolet	Cavalier

DISCUSSION

Almost all of the drivers in this study were involved in frontal-type crashes. This is expected, because the airbag crash sensors are designed to respond to longitudinal deceleration of the vehicle. Based on detailed crash investigations, all of the airbag-induced facial injuries appear to have resulted from the airbag and not from the airbag module cover.

In previous studies of frontal crashes of cars without airbags, a large percentage of facial injuries involve fracture of the facial bones, with or without overlying soft tissue damage, and/or brain injury, due to contact with the steering-wheel. Approximately one-third of all steering-wheel airbag deployments in the current study resulted in facial injury, but almost all of these injuries are minor-- including erythemia, abrasions and contusions. These injuries usually occur to the cheeks, chin, or tip of nose. Forehead involvement is infrequent. Occasionally small, deep abrasions are incorrectly reported as lacerations. Facial fractures due to the airbag are extremely rare in the UMTRI database and appear to be due to the hand or forearm being flung into the face by the airbag, resulting in an undisplaced nasal bone fracture. Periorbital and ocular injuries are less frequent than overall facial injuries. The frequency of airbag-induced periorbital injuries is low (5%) with only 2% of the drivers sustaining an ocular injury.

Facial injuries are just as likely to occur to belt-restrained drivers as to unrestrained drivers. This is probably due to the fact that the deploying airbag contacts the driver's face before the lap-shoulder belt has its full effect in limiting the driver's forward movement. Also many drivers, particularly those of short stature, are in the airbag deployment envelope before the crash.

In a previous study on cranial-cervical dislocations from driver interaction with the steering-wheel airbag, all injured drivers were 165 cm or shorter.[46] In the current study, there is a decrease in the frequency of airbag-induced facial, periorbital, and ocular injuries, with increasing driver stature, although these injuries occur throughout most of the range of stature. The UMTRI database also indicates that females are more likely to sustain both facial and ocular airbag-induced injuries, probably because of their shorter stature and closer proximity to the airbag module. Also, even though 74% of the 540 steering-wheel airbag deployments involved tethered airbags, the frequency of facial injuries was higher in the untethered airbag population.

Crash related ocular injuries are not new. Numerous ocular injuries in car crashes have been reported for unrestrained occupants in older model vehicles without airbags. In these cases, interior contacts producing ocular injuries include the windshield and pillars, sunvisors, side door glass, steering-wheel and instrument panel.[44] In England and Europe, Australia and Asia, unrestrained front-seat occupants of older vehicles with tempered-glass windshields often sustained ocular perforations and, during the 1970s and 1980s, approximately 1800 British occupants received serious eye injuries each year.[45]

Airbag-induced ocular injuries have been previously reported from direct airbag contact with the eye, with eyeglasses, and with a tobacco pipe stem, as well as from the driver's hand being flung into the driver's face causing fracture to the nasal bones and infraorbital rim. Of the eleven drivers in the UMTRI study with ocular injuries, all were less than 179 cm in stature, nine were restrained by a lap-shoulder belt, seven were females, and five were injured by tethered airbags.

From the literature and the UMTRI database, 60 cases of airbag-induced ocular injury have now been reported.[16-40] Of these drivers 27 are known to have sustained an ocular injury from a tethered airbag compared to 21 injured by an untethered airbag. The number of tethered airbags is probably higher in the total population and thus more ocular injuries are related to the tethered airbag.

REFERENCES

1. Werner JV, Sorenson WW: Survey of airbag involved accidents an analysis of collision characteristics, system effectiveness and injuries. SAE Paper No. 940802. *Soc of Auto Eng.* 1994; Warrendale, PA. 1994

2. Libertiny G: Airbag effectiveness - trading major injuries for minor ones. SAE Paper No. 950871, *Soc of Auto Eng.* 1995; Warrendale, PA

3. Padmanaban J, Ray R, Weiss J: Assessment of airbag effectiveness in reducing non-fatal injury risk. Safety Engineering and Risk Analysis, 1993; SERA-Vol. 1:185-188

4. Lund A, Ferguson S: Driver Fatalities in 1985-1993 cars with airbags. *J Trauma.* 1995; 38:469-475
5. Zador P, Ciccone M: Automobile driver fatalities in frontal impacts: airbags compared with manual belts. *Am J Public Health.* 1993; 33:661-666
6. Ferguson S, Lund A, Greene M: Driver fatalities in 1985-1994 airbag cars. *Insurance Institute for Highway Safety,* 1995; Washington, DC, 9 pp
7. Evans L. Restraint effectiveness, occupant ejection from cars, and fatality reduction. *Accid Anal Prev.* 1990; 22:167-175
8. Kahane C: Fatality reduction by air bags Analyses of accident data through early 1996. DOT HS 808 470, August 1996, 140 pp
9. Crandall JR, Kuhlmann TP, Martin PG, et al: Differing patterns of head and face injury with airbag and/or belt restrained drivers in frontal collisions. *Advances in Occupant Restraint Technologies: Proc of Jt AAAM-IRCOBI Special Session,* Lyon (FR), September 22, 1994
10. Huelke DF: An overview of airbag deployments and related injuries. Case studies and a review of the literature. SAE Paper No. 950866. *Soc of Auto Eng.* Warrendale, PA. 1995
11. Dalmotas D, Hurley R, German A: Airbag deployments involving restrained occupants. SAE Paper No 950868, *Soc of Auto Eng.*, 1995; Warrendale, PA
12. Huelke DF, Moore JL, Compton TW, et al: Upper extremity injuries related to airbag deployments. *J Trauma.* 1995; 38:482-488
13. Smock WS, Nichols GR: Airbag module cover injuries. *J Trauma.* 1995; 38:489-493
14. Augenstein JS, Perdeck EB, Murhta M, et al: Injuries sustained by drivers in airbag crashes. *15th ESV Conf.* Paper No. 96-S10-O-01, Melbourne, Australia, 1996
15. Freedman EL, Safran MR, Meals RA: Automotive airbag-related upper extremity injuries: a report of three cases. *J Trauma.* 1995; 38:577-581
16. Huelke DF, Moore JL, Ostrom M: Airbag injuries and occupant protection. *J Trauma.* 1991; 33:894-898
17. Gault JA, Vichnin MC, Jaeger EA, et al: Ocular injuries associated with eyeglass wear and airbag inflation. *J Trauma.* 1995; 38:494-497
18. Duma SM, Kress TA, Porta DJ: Airbag-induced eye injuries: a report of 25 cases. *J Trauma.* 1996; 41:114-119
19. Sastry SM, Copeland RA, Mezghebe H, et al: Retinal hemorrhage secondary airbag-related ocular trauma. *Trauma.* 1995; 38:582
20. Blair G, Larkin GL: Airbag-mediated facial trauma. *4th Intl Cong on Emerg Med.* 1992; Washington, DC
21. Rosenblatt M, Freilich B, Kirsch D: Airbag associated ocular injury. *Arch Opthalmol.* 1993; 3:1318
22. Goldberg MA, Valluri S, Pepose JS: Airbag-related corneal rupture after radial keratotomy. *Am J Ophthamol.* 1995; 120:800-802
23. Whitacre MM, Pilchard WA: Airbag injury producing retinal dialysis and detachment. *Arch Opthalmol.* 1993; 111:1320
24. Lesher MP, Durrie DS, Stiles MC, et al: Corneal edema, hyphema, and angle recession after airbag inflation. *Arch Opthalmol.* 1993; 111:1320-1322
25. White J, McClafferty KJ, Shkrum MJ, et al: Case study - alkali eye injury secondary to airbag deployment. *Proc Canadian Multidisiplinary Road Safety Conf IX,* Montreal, Canada, 1995; 387-394
26. Fukagawa K, Tsubota K, Kimura C, et al: Corneal endothelial cell loss induced by airbags. *Opthal.* 1993; 100:1819-1823
27. Driver PJ, Cashwell LF, Yeatts R: Airbag-associated bilateral hyphemas and angle recission. *Am J Opthalmol.* 1994; 118:250-251
28. Braude LS: Protective eyewear needed with driver's side airbag? *Arch Ophthalmol.* 1992; 110:1201
29. Onwuzuruigbo CJ, Gulda GJ, Larned D, et al: Traumatic blindness after airbag deployment: bilateral lenticular dislocation. *J Trauma.* 1996; 40:314-316
30. Vichnin MC, Jaeger EA, Gault JA, et al: Ocular injuries related to airbag inflation. *Ophthalmic Surgery and Lasers.* 1995; 26:542-548
31. Larkin G: Airbag-mediated corneal injury. *Am J Emerg Med.* 1991; 9:444-446
32. Rimmer S, Shuler JD: Severe ocular trauma from a driver's-side airbag. *Arch Opthalmol.* 991; 109:774
33. Mishler KE: Hyphema caused by airbag. *Arch Opthalmol.* 1991; 109:1635
34. Ingraham HJ, Perry HD, Donnenfeld ED: Air-bag keratitis. *New Eng J Med.* 1991; 324:1599-1600
35. Smally AJ, Binzer A, Dolin S, et al: Alkaline chemical keratitis: eye injury from airbags. *Ann Emer Med.* 1992; 21:1400-1402
36. Scott I, John G, Stark WJ: Airbag-associated ocular injury and periorbital fractures. *Arch Opthalmol.* 1993; 111:25
37. Han DP: Retinal detachment caused by airbag injury. *Arch Opthalmol.* 1993; 111: 1317-1318.
38. Campbell JK: Automobile airbag eye injuries. *Neb Med J.* 1993:306-307
39. Walz F, Mackey M, Gloor B: Airbag and eye perforation by tobacco pipe. *J Trauma.* 1995; 36:498-500
40. Baker RS, Flowers CW, Singh P, et al: Corneoscleral laceration caused by air-bag trauma. *Am J Opthalmol.* 1996, 32:135-137
41. Powell MR, Lund AK: Leading edge deployment speed of production airbags. SAE Paper No 950870. *Soc of Auto Eng.* 1995; Warrendale, PA
42. Sullivan LK, Kossar JM: Airbag deployment characteristics. *DOT HS 807 869,* Final Report, Feb. 1992; 84 pp
43. Reed M , Schneider LW, Burney RE: Investigation of airbag-induced skin abrasions. *36th Stapp Car Crash Conf Proc* Paper No. 922510, 1-12, 1992
44. Huelke DF, O'Day J, Barhydt WH: Ocular injuries in automobile crashes. *J Trauma.* 1982, 22:50-52
45. Mackay GM: Injuries for windscreens. A review of some recent literature. *A Report to the Commission of the E.E.C;* February, 1978
46. Huelke DF, Reed RT: Cranial-vertebral fractures and dislocations associated with steering-wheel airbag deployment. *15th ESV Conf.* Paper No. 96-S1-O-01. Melbourne, Australia, 1996

970491

Injury Risks in Cars with Different Air Bag Deployment Rates

John V. Werner and Steve F. Roberson
State Farm Mutual Automobile Insurance Co.

Susan A. Ferguson
Insurance Institute for Highway Safety

Kennerly H. Digges
George Washington University

ABSTRACT

Automobile insurance claims of two popular midsize cars with different air bag deployment frequencies -- the Dodge/Plymouth Neon and Honda Civic -- were examined to determine performance in higher severity crashes (the upper 30 percent of crashes ranked by adjusted repair cost). Previously, it was found that drivers sustained more, mainly minor, injuries in the Neon which had a higher deployment frequency in low speed crashes. This study examined, for these two cars, whether there was any trade-off associated with a higher deployment threshold. It was found that even at higher speeds, the Neon had a greater frequency of air bag deployments, which in turn resulted in a greater likelihood of driver injury. Once again upper extremity injuries were most prevalent for Neon drivers and were highest for female drivers. At the same time, there was little evidence that driver protection was compromised in the Civic in the more important high speed crashes. The limited evidence from this study suggests that efforts should be made to reduce the number of air bag deployments in low speed crashes.

INTRODUCTION

Air bags can be designed to deploy at various speeds, but the assumption is that deployments should not occur below the injury threshold. However, there is growing concern about the large numbers of air bag deployments in crashes where speeds are slow enough that air bags are not needed to provide protection, particularly for belted occupants (1, 2, 3). This is particularly troubling because air bags can cause injury even in very low severity crashes. Air bag-related injuries are mostly minor, but in rare instances they involve serious and fatal injury to the head and chest to people who are close to them when they deploy (2).

Air bags are designed to deploy in frontal crashes at barrier equivalent speeds of about 10-12 mph. However, many of them are occurring at speeds lower than this nominal threshold depending on sensor type and location as well as crash type. Analyses of data from the National Highway Traffic Safety Administration (NHTSA) National Accident Sampling System (NASS 1990-1994) reveals that about a third of all deployments occur in frontal crashes at longitudinal delta Vs of less than 10 mph (2). Forty-two percent occur between 10 and 15 mph. There is also evidence of tremendous variation in deployment frequencies and deployment thresholds among different makes and models (3,4). Moreover, a recent study found, in a comparison of low severity crashes in two popular midsize cars, that drivers in the vehicle with more low speed deployments were injured more often than drivers in the vehicle with fewer deployments; albeit these were mostly minor injuries (3).

It has been suggested that the solution to the high number of low speed deployments is to raise the deployment threshold. A recent petition to NHTSA (5) recommended that Federal Motor Vehicle Safety Standard 208 be amended to set a 12 mph threshold below which the air bag cannot deploy. There are, however, some concerns that need to be addressed. If deployment thresholds are higher there is the possibility that injury rates will increase for unbelted occupants who no longer will be protected by an air bag in some crashes. There is also the concern that higher deployment thresholds may result in later deployments. This could have the effect of increasing injuries because occupants have the potential to be closer to the air bag at the time of deployment.

The previous study using insurance claims data for the 1995 Dodge/Plymouth Neon and the 1992-1995 Honda Civic found higher injury rates for the vehicle with higher deployment frequencies in low speed crashes (3). However, these comparisons provided no information on the overall performance of these two vehicles. That is, there was no analysis of the protection offered at higher speeds. This paper provides a comparison of injury likelihood in the same make/model vehicles, but in crashes of higher severity -- the crashes that were not included in the previous paper.

METHOD

The following sources of information form the basis for the analysis and discussion in this paper:

1. State Farm Mutual Automobile Insurance Company's electronically prepared damage repair estimates for towed vehicles.

2. State Farm survey of claims involving frontal crashes for the 1995 Dodge/Plymouth Neon and the 1992-1995 Honda Civic.

3. The leading edge speed and excursion profile of driver air bag fabric in the 1995 Dodge/Plymouth Neon, the 1996 Neon, and the 1993 Honda Civic.

DEPLOYMENT FREQUENCY - A 1996 study escribed the use of State Farm damage repair estimates) identify towed vehicles with frontal damage (3). Deployment status was determined by analyzing the amage estimate for driver air bag module replacement. The Plymouth/Dodge Neon and the Honda Civic have imilar repair cost severity distributions, but a large ifference in their air bag deployment frequencies - the Honda Civic with an estimated deployment frequency of 0 per 100 frontal crashes, and the Plymouth/Dodge Neon rith an estimated 50 deployments per 100 frontal rashes.

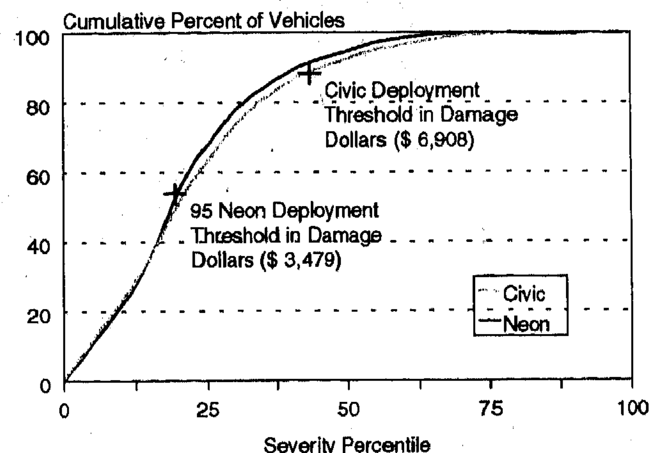
CLAIM SURVEY - Because insurance claim ıvestigators do not require crash severity or delta V to be etermined as part of the claim settlement process, the delta V" distribution of the vehicle crashes represented by ısurance data was approximated by fitting the rank rdered cumulative distribution of dollar damages for ıese two cars to the cumulative delta V distribution etermined from a analysis of midsized cars in the lational Accident Sampling System (NASS 1989-1994,). The repair costs were adjusted to remove air bag ystem related expenses. Claim files representing the pper 30 percent of the cumulative distribution of crashes, ank ordered by adjusted repair cost, were selected for ıis study.

In the absence of crush measurements for insurance ata, the correlation between the rank ordered repair cost istribution and the NASS delta V distribution is not exact; owever, a review of the photographs of vehicle damage nd a review of the damage estimates show a systematic ıcrease in the extent of unibody structural damage. The uthors are confident that the claim files selected for this tudy are representative of the moderate to serious rashes seen in NASS.

Claim files were surveyed to provide information egarding policyholder involved crashes. Information abulated included a crash description, age and gender of ıe insured driver, driver injury descriptions, and erification of air bag deployment status. The Abbreviated njury Scale (AIS) and Occupant Injury Classification OIC) coding were used to document the injuries to insured drivers that sought treatment resulting in claim payments.

RESULTS

REPAIR COST DISTRIBUTION - Figure 1 depicts the cumulative percentage of 1995 Neons and 1992-1995 Civics repaired by damage severity percentile for frontal crashes. Severity percentile was determined by dividing each car's repair cost range into eight equal repair dollar intervals. A Kolmogorov-Smirnov test was used to compare the equivalency of the two cumulative repair cost distributions. The test results support the assumption that the repair cost distributions are equal for 1995 Neons and 1992-1995 Civics (KSA [1020] = 0.697, $p = 0.716$).



Note: 100th severity percentile represents $16,198 for the Neon and $15,893 for the Civic.

Figure 1. Cumulative Neon and Civic collision claim severity distributions for frontal impacts with towing.

Cases included in these analyses were those in the upper 30 percent of the cumulative distribution of crashes rank ordered by adjusted repair costs. This portion of the distribution includes crashes judged to be of moderate to high severity. For 1995 Neons, these crashes represent damages in excess of $4,264 in adjusted repair costs. For Civics, crashes with more than $4,461 in damages were included.

$$\log(p/(1-p)) = b_0 + b_1x \quad (1)$$

$$x(p=.5) = -b_0/b_1 \quad (2)$$

A logistic regression model (Equation 1) was used to determine the damage severity at which the probability of an air bag deployment reached 50 percent. P is the probability of deployment and x is the damage severity (in dollars). For p = 0.5, the damage severity is given by Equation 2. The derived value is defined for this study as the air bag deployment threshold, that is, the dollar point in the distribution where the probability of deployment is 50 percent. The adjusted repair cost,

where the probability of deployment is 50 percent is $3,479 for Neons (95 percent CI = $3,223 to $3,757). Civics have a 50 percent probability of air bag deployment at a $6,908 adjusted repair cost (95 percent CI = $6,452 to $7,484).

DEPLOYMENT FREQUENCIES - Figure 2 shows air bag deployments per 100 frontal crashes by damage severity, for the comparison vehicles. The Civic has an estimated deployment frequency of 48 per 100 moderate to serious, frontal crashes, and the Neon has an estimated 89 deployments per 100 moderate to serious, frontal crashes.

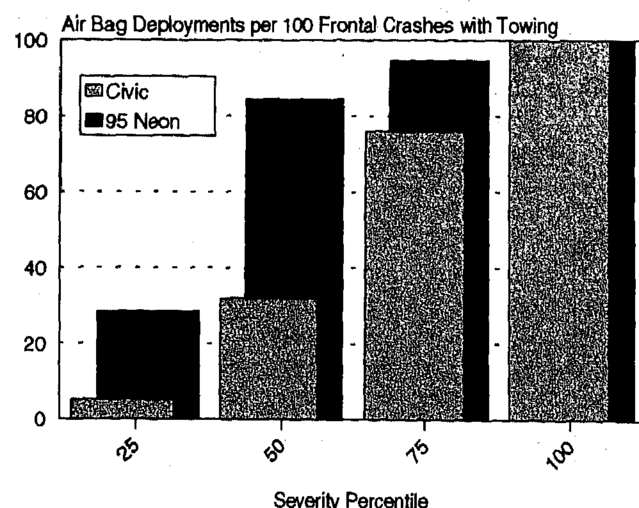


Figure 2. Air bag deployment rate by severity, frontal crashes with towing.

An analysis of the odds of deployment for moderate to serious crashes shows that there are significantly more deployments per 100 frontal crashes for Neons than for Civics (OR = 9.17; 95 percent CI = 4.72 to 17.54).

INJURIES - The following analyses compare injury rates in the two cars for moderate to serious crashes (the upper 30 percent of crashes ranked by adjusted repair amount). As shown in Table 1, 50 percent of Neon drivers report an injury in these crashes compared to 35 percent of Civic drivers. The relative risk of injury for Neon drivers compared to Civic drivers is 1.83 (95 percent CI = 1.16 to 2.96).

Table 1
Driver Injury by Car Model
for Moderate to Serious Crashes

	Driver Injury No	Driver Injury Yes	Total Crashes
Neon	57 (50.0%)	57 (50.0%)	114 (100%)
Civic	126 (64.6%)	69 (35.4%)	195 (100%)

Table 2 shows the maximum injury severity (MAIS) for the Neon and Civic drivers. Overall, injury severity was low for these drivers. Of the 309 Neon and Civic drivers, 292 (94.5 percent) had either minor injury or no injuries (MAIS-1); 15 (4.9 percent) had moderate injuries (MAIS-2); and 2 (0.6 percent) had a MAIS-3 injury. Very few MAIS-2 or greater injuries were associated with crashes in this crash severity range.

Appendix A shows the MAIS-2 or greater injuries, with driver age, gender, reported seatbelt use, and deployment status, rank ordered by the cumulative distribution percentile of crash severity.

Compared to the overall female-to-male driver ratio (2:1), females represented a higher proportion of the MAIS 2+ injuries (3.5:1). The odds of MAIS-2+ injury for female drivers was higher than for MAIS-2+ injured male drivers (OR = 1.81; 95 percent CI = 0.58 to 5.68). However, this relative risk was not significant.

Table 2
Drivers Injured by Maximum AIS
for Moderate to Serious Crashes

MAIS	Neon Count	Neon Percent	Civic Count	Civic Percent	Total Count	Total Percent
0	57	50.0	126	64.6	183	59.2
1	51	44.7	58	29.7	109	35.3
2	6	5.3	9	4.7	15	4.9
3	0	0.0	2	1.0	2	0.6
> 3	0	0.0	0	0.0	0	0.0
Total	114	100.0	195	100.0	309	100.0

Injuries by Gender - The injury data were also analyzed by gender of driver (Table 3). The proportion of male to female drivers involved in these crashes did not differ significantly for the two vehicles ($\chi^2[1] = .032$, $p = .857$), so overall injury risk can be compared without having to control for gender differences. Comparing across vehicles, the relative risk of injury to female drivers for Neons compared to Civics was 1.62. However, this relative risk was not significant (95 percent CI = 0.87 to 3.01). For male drivers, the relative risk of injury was 1.54 for Neons compared to Civics, again not significant (95 percent CI = 0.64 to 3.72).

Table 3
Driver Injury Status by Gender
for Moderate to Serious Crashes

Driver Gender	Driver Injury No	Driver Injury Yes	Total Crashes
Neon			
Female	26 (38.2%)	42 (61.8%)	68 (100%)
Male	19 (54.3%)	16 (45.7%)	35 (100%)
Civic			
Female	53 (50.5%)	52 (49.5%)	105 (100%)
Male	33 (64.7%)	18 (35.3%)	51 (100%)

Within a given vehicle, a comparison of injury risk differences by gender, shows that Neon female drivers

had higher odds of injury when compared to Neon male drivers (OR = 1.92; 95 percent CI = 0.84 to 4.39). In Civics, female drivers had a relative risk of injury equal to 1.83 when compared to male drivers (95 percent CI = 0.92 to 3.66) with neither result being statistically significant.

Injuries by Air Bag Deployment Status - Table 4 shows injury rates by car make and deployment status. In air bag deployment crashes Neon drivers had 1.45 times higher odds of injury than Civic drivers. However, this difference was not significant (95 percent CI = 0.83 to 2.54). For crashes in which the air bag did not deploy, the relative risk of injury for drivers was 0.87 for Neons, again not significant (95 percent CI = 0.22 to 3.45).

Table 4
Driver Injury Status by Deploy Status for Moderate to Serious Crashes

Deploy Status	Driver Injury No	Driver Injury Yes	Total Crashes
Neon			
No	9 (75.0%)	3 (25.0%)	12 (100%)
Yes	47 (46.1%)	55 (53.9%)	102 (100%)
Civic			
No	73 (72.3%)	28 (27.7%)	101 (100%)
Yes	52 (55.3%)	42 (44.7%)	94 (100%)

The odds of injury for Neon drivers in air bag deployment crashes was higher than for Neon drivers in non-deployment crashes (OR = 3.51; 95 percent CI = .898 to 13.73). Likewise, Civic drivers had higher odds of injury for air bag deployment crashes compared to non-deployment crashes (OR = 2.11; 95 percent CI = 1.16 to 3.82). However, it should be noted that crashes in which an air bag deployment occurs, typically are of higher severity and more injuries could be expected.

Location of Driver Injuries - Table 5 shows driver injury rates for "High" severity crashes (the upper 30 percent of crashes ranked by adjusted repair amount) and "Low" severity crashes (the lower 70 percent of crashes ranked by adjusted repair amount) by body region and car model. The injury rate (injuries per 1,000 frontal crashes with towing) includes multiple injuries per injured driver.

As might be expected, drivers in high severity crashes had higher rates of injury to all body regions, with the exception of head injuries to Neon drivers. Neon drivers had higher rates of injury to the upper extremities, including the upper arms, wrists, and hands for both high and low severity crashes. The rate of predominately minor facial injuries to Civic drivers is substantially higher than the Neon facial injury rate for the high crash severity category (103 compared to 35). The rate of facial injuries to Neon drivers is marginally higher than the Civic facial injury rate for the low crash severity category (20 compared to 13). The rate of lower extremity injuries for high severity crashes has a five-to-six fold increase over the rate of lower extremity injuries for low severity crashes. Neon and Civic drivers have similar lower extremity injury rates.

Table 6 shows driver injury rates by gender and car model for different body regions. Female drivers had generally higher injury rates compared with male drivers. Female Neon drivers had the highest overall injury rate in this comparison, with substantially higher injury rates to upper extremities compared with all other car model and gender categories. However, the rate of head, facial, and chest injuries to Civic female drivers is higher than for Neon female drivers.

Table 5
Driver Injuries per 1,000 Frontal Crashes with Towing by Car Model

Body Region	Neon Crash Severity High	Neon Crash Severity Low	Civic Crash Severity High	Civic Crash Severity Low
Head	26	38	41	22
Face	35	20	103	13
Upper Extremity				
Elbow	9	4	15	--
Upper arms	149	46	31	6
Wrist-Hand	228	81	56	17
Forearm	--	24	--	2
Total Upper Extremity	386	155	102	25
Chest	70	30	97	31
Neck/Back				
Neck/C. Spine	184	80	159	80
Shoulder	70	34	51	20
Back/T. Spine	105	53	138	71
Total Neck/Back	359	167	348	171
Lower Extremity				
Knee	88	19	113	22
Lower Leg	44	8	15	6
Ankle-Foot	61	8	36	--
Pelvic-Hip	9	4	26	2
Total Lower Extremity	202	39	190	30
Other Body Areas	--	--	10	--
Unknown	44	8	10	4
Total	1,122	457	901	296

Multiple injuries per injured driver possible.

Table 7 shows driver injuries by air bag deployment status and car model. There were very few claims to review for the Neon non-deployment classification, because of the high deployment frequency for Neons in these moderate to serious crashes. Neon drivers involved in air bag deployment crashes had higher rates

of injury to the upper extremities compared with Civic drivers involved in air bag deployment crashes.

Table 6
Driver Injuries per 1,000 Frontal Crashes with Towing by Gender and Car Model
Moderate to Serious Crashes

Body Region	Neon Male	Neon Female	Civic Male	Civic Female
Head	57	15	38	58
Face	57	29	58	163
Upper Extremity				
Elbow	--	15	19	19
Upper Arms	29	234	--	58
Wrist/Hand	143	309	58	77
Total Upper Extremity	172	558	77	154
Chest	114	59	58	154
Neck/Back				
Neck/C. Spine	200	206	154	221
Shoulder	86	74	58	67
Back/T. Spine	143	103	212	154
Total Neck/Back	429	383	424	442
Lower Extremity				
Knee	86	103	77	173
Lower Leg	--	74	--	29
Ankle-Foot	29	88	--	67
Pelvic-Hip	29	--	58	19
Total Lower Extremity	144	265	135	288
Other Body Areas	--	--	--	19
Unknown	57	44	19	10
Total	1,093	1,353	809	1,288

Multiple injuries per injured driver possible.

The rate of facial and chest injuries to Civic drivers involved in air bag deployment crashes was higher than the Neon facial and chest injury rate for air bag deployment crashes. Civic air bag deployments occur at the upper range of the crash severity distribution and likely reflect a more serious crash, on average, compared with Neon air bag deployment crashes.

Driver Air Bag Fabric Leading Edge Speeds and Excursions - As noted in the previous paper (3) both the Neon and the Civic air bags are tethered and constructed of relatively lightweight materials. The maximum excursion of the air bag fabric leading edge is also quite similar for the two bags (see Table 8). However, the maximum leading edge speed of the 1995 Neon air bag is much higher than that of the Civic (see also Figures 3 and 4). There is an increased potential for abrasive type injuries for higher speed air bags, particularly if they reach speeds in excess of 230 km/h (6).

Table 7
Driver Injuries per 1,000 Frontal Crashes with Towing by Air Bag Deploy Status and Car Model
Moderate to Serious Crashes

Body Region	Neon Deploy Status Yes	Neon Deploy Status No	Civic Deploy Status Yes	Civic Deploy Status No
Head	29	--	21	58
Face	39	--	149	58
Upper Extremity				
Elbow	10	--	11	19
Upper Arms	167	--	43	19
Wrist/Hand	255	--	106	10
Total Upper Extremity	432	--	160	48
Chest	69	83	138	58
Neck/Back				
Neck/C. Spine	186	167	149	163
Shoulder	78	--	64	38
Back/T. Spine	108	83	128	144
Total Neck/Back	372	250	341	345
Lower Extremity				
Knee	98	--	138	87
Lower Leg	49	--	32	--
Ankle-Foot	69	--	43	29
Pelvic-Hip	10	--	53	--
Total Lower Extremity	226	--	266	116
Other Body Areas	--	--	21	--
Unknown	49	--	21	--
Total	1,216	333	1,117	683

Multiple injuries per injured driver possible.

However, it should be noted that the Neon air bag underwent a design change for the 1996 model year and leading edge speeds in the 1996 Neon air bag now are lower than either the Honda Civic, or the 1995 Neon.

Table 8
Air Bag Maximum Leading Edge Speeds and Maximum Excursions

Model	Tethered	Maximum Excursions (mm)	Maximum Edge Speed (km/h)
Honda Civic	Yes	328	227
1995 Dodge Neon	Yes	334	325
1996 Dodge Neon	Yes	274	158

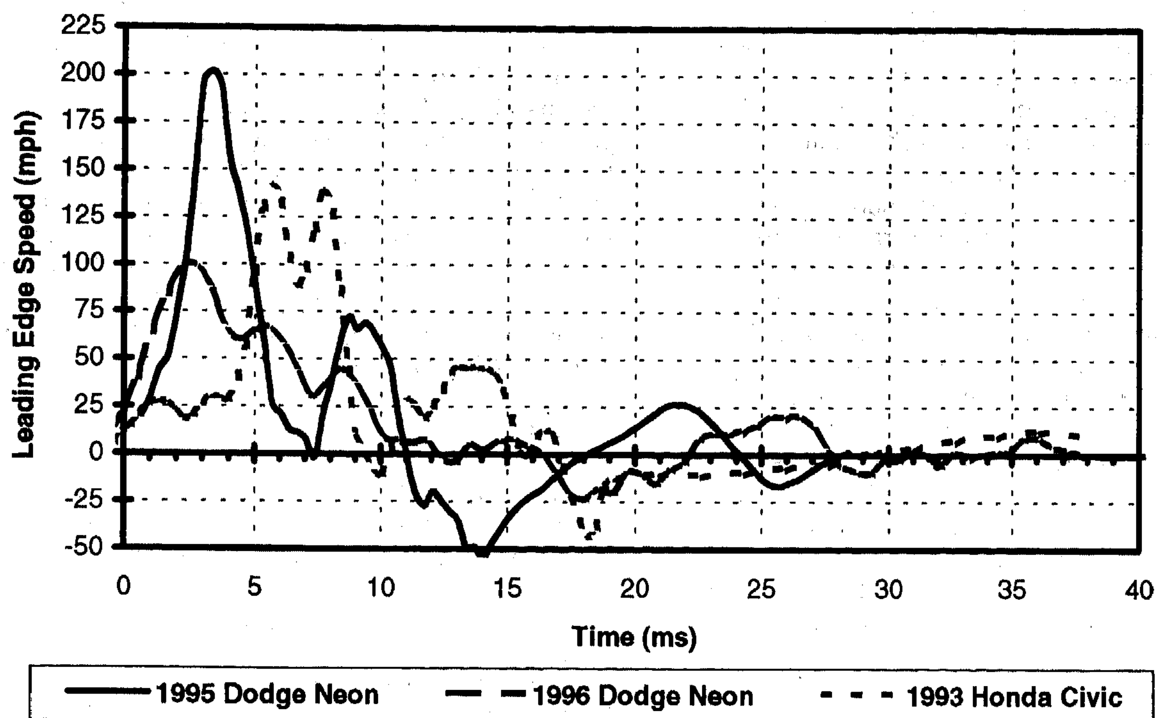


Figure 3. Air bag leading edge speeds by deployment time for the Neon and Civic air bags.

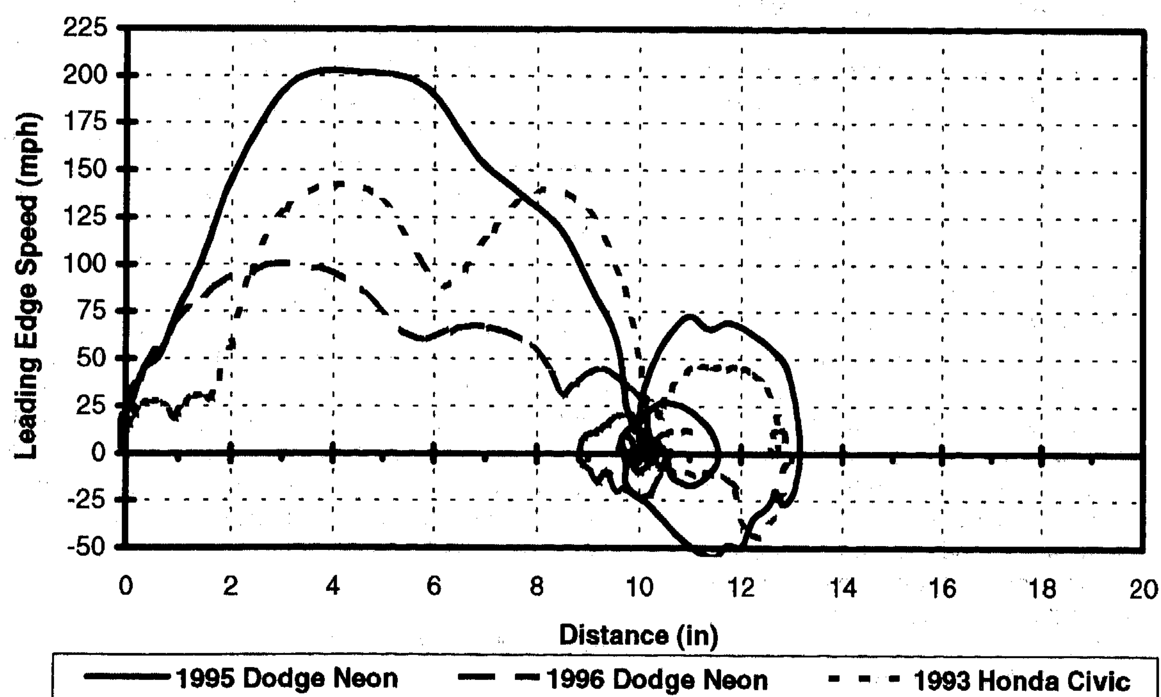


Figure 4. Air bag leading edge speeds by deployment distance for the Neon and Civic air bags.

DISCUSSION

The comparative performance of two similar mid-sized cars in moderate to serious crashes was considered in this paper to supplement results from a previous paper examining injury likelihood for the same two vehicles in low severity crashes. The previous study determined that drivers in the vehicle with more low speed deployments were injured more often than drivers in the vehicle with fewer deployments.

For the moderate to serious crashes surveyed in this study, one vehicle had substantially more deployments than the other vehicle. One might expect more moderate and serious injuries to drivers of the vehicle with fewer deployments in these moderate to serious crashes if air bags are failing to deploy when needed to prevent injury. This was not observed. There were very few moderate and greater injuries to the drivers of both vehicles and the proportion of these injuries was very similar; moreover, drivers in the vehicle with more deployments were injured more often overall than drivers in the vehicle with fewer deployments.

A comparison of the MAIS-2 and greater injury patterns, by body region, between vehicles, revealed no clear differences. Although drivers in both vehicles suffered head and chest injuries in crashes in which the air bag did not deploy, these injuries were predominantly minor. The absence of serious injuries to drivers of the vehicle with fewer deployments in moderate to serious crashes is encouraging. However, because of the limited number of cases available, and the absence of any very high speed crashes, further research is needed to confirm that in the Civic, and more generally, in other vehicles with fewer low speed deployments, that there is no injury trade-off when setting deployment thresholds higher.

The level of belt use is an important consideration for interpreting these results. Unbelted occupants are more vulnerable to injury. However, only self-reported belt use is available for drivers in this study and in the Civic and Neon, about 75 percent of the injured drivers reported that they used their belt. High belt use decreases the likelihood of observing serious injuries in non-deployment, moderate to serious crashes.

The results of this study indicate that the rate of upper extremity injury is higher, the higher the rate of deployment, regardless of crash severity. In addition, female drivers appear to be sustaining more injuries, particularly to the upper extremities, than males in these moderate to serious crashes. Proximity to the air bag at the time of deployment is clearly a risk factor. Less aggressive air bag designs and air bag systems currently under development that take into account occupant size, weight, and proximity to the air bag, may help alleviate this problem.

The authors' previous study outlined two factors that could be contributing to the higher rate of injury to Neon drivers compared with Civic drivers. Results of this study indicate that these same factors apply to the interpretation of the higher rate of Neon driver injury compared with Civic drivers in moderate to serious crashes. The 1995 Neon experiences many more air bag deployments in moderate to serious frontal crashes and the air bag fabric has a higher leading edge speed compared with the Civic. Both of these air bag performance characteristics appear to be changed for the 1996 Neon. The estimated deployment frequency for the 1996 Neon is 35 per 100 frontal crashes compared with 50 deployments per 100 frontal crashes for the 1995 Neon and 20 deployments per 100 frontal crashes for the Civic. Also, the leading edge speed in the 1996 Neon air bag is lower than either the Honda Civic, or the 1995 Neon. Results of this study and the previous related study suggest that these changes will result in lowering the injury rate to drivers of 1996 Neons involved in frontal crashes.

CONCLUSIONS

Auto manufacturers, the NHTSA, air bag system component suppliers, auto safety researchers, and auto safety advocates are currently challenged to find effective solutions to unintended injuries from air bag deployments. Smart air bag technology is expected to address the issues associated with the most serious of these unintended air bag induced injuries in the longer term. However, air bag system designs that optimize the air bag deployment threshold, using current and available technology, can partially address these concerns by reducing the number of unneeded deployments. For example, dual deployment thresholds that are set to deploy the air bag at different crash severities based on belt use could reduce unneeded deployments for belted occupants. Serious consideration should be given to raising the deployment threshold for unbelted occupants where the vehicle's air bag deployment threshold results in large numbers of low speed deployments. One concern in raising the unbelted threshold is that facial injuries, such as fractures to facial bones, could increase. This study, although limited in scope, found no evidence of such a trade-off.

The higher rate of mainly minor injuries for Neon drivers could be related both to the higher frequency of air bag deployment as well as the higher leading edge speed of the 1995 Neon air bag. Air bag design characteristics associated with the number of deployments and air bag leading edge speed were changed for the 1996 Neon. These changes will likely reduce the air bag induced injuries in low to moderate severity crashes, while maintaining air bag performance in higher speed crashes.

ACKNOWLEDGEMENTS

The authors acknowledge the contributions from Karon Bakel and Lisa Craver in the collection of the survey data.

REFERENCES

1. Dalmotas, D.J., Hurley, J., German, A., and Digges, K. D., 1996. Air bag deployment crashes in Canada. Proceedings of the 18th International Technical Conference on Experimental Safety Vehicles, Melbourne, Australia, Paper 96-S1-O-05.

2. Ferguson, S. A., Update on airbag performance in the United States: Benefits and Problems Proceedings of the Airbag 2000 Conference in Karlsruhe, Germany, November 26, 1996.

3. Werner, J.V., Roberson, S.F., Ferguson, S.A., Digges, K.D. Air bag deployment and injury risks. Paper 960664, SAE, Detroit, 1996.

4. Werner, J.S. and Sorenson, W.W., Survey of airbag involved accidents. An analysis of collision characteristics, system effectiveness and injuries. *Safety Technology*. SP-1041. Paper 940802. Detroit, SAE, 1994, pp 1-19.

5. Center for Auto Safety, Petition to the National Highway Traffic Safety Administration. November 8, 1996.

6. Powell, M.R. and Lund, A.K. Leading edge deployment speeds of production air bags. Detroit, SAE, No. 950870, 1995.

Appendix A
MAIS-2 or Greater Injuries
by Vehicle
Moderate to Serious Crashes

Vehicle	Distribution Percentile	MAIS	Deploy Status	Gender	Age	Belt Use	Injury
Civic	76	2	No	Female	19	No	Concussion
	82	3	No	Female	26	Yes	Ankle Fracture
	85	2	No	Female	35	Yes	Sternum Fracture
	86	2	No	Female	Unknown	Unknown	Scalp Laceration
	90	2	Yes	Female	Unknown	Unknown	Patella Fracture
	93	3	Yes	Male	Unknown	Unknown	Hip Dislocation
	93	2	Yes	Male	45	Yes	C-2 Fracture
	94	2	Yes	Female	52	Yes	Patella Fracture
	95	2	Yes	Female	48	Yes	Knee Sprain
	95	2	Yes	Female	34	Yes	Kidney Contusion
	98	2	Yes	Female	Unknown	Unknown	Face - Not Further Specified
Neon	75	2	No	Male	Elderly	Unknown	2 Rib Fractures
	80	2	Yes	Female	26	Unknown	Hand Laceration
	87	2	Yes	Female	Unknown	Unknown	Upper Extremity Fracture
	95	2	Yes	Male	71	Yes	Scalp Laceration
	97	2	Yes	Female	Unknown	Unknown	Lower Extremity Fracture
	100	2	Yes	Female	20	Yes	Multiple Lower Extremity Fracture

970522

Parametric Model of Elastomeric Bolt Isolators Under Large Deformation

Zhi-Quan Hu, T. Y. Chen, and Gary Novak
Fel-Pro Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

In an integrated noise isolation system, both gasket and bolt isolators play an important role in the resultant total load, contact stress, and system static and dynamic stiffness. One standard approach to design is based on using a shape factor in conjunction with a simple stress-strain relation. This approach can produce reasonable results under small deformation. But the deviation is so large that it is not appropriate to use it for reliable designs under large deformation. Experimental approaches and finite element methods are effective at characterizing the load-deflection relationship of bolt isolators under large compressive deformation. In this paper, a parameter study of a sequence of finite element analyses will be presented and a parametric model will be determined based on the FEA results to best represent the isolator load-deflection behavior under large compression. The resulting formula can be used to quickly and accurately determine the isolator static and dynamic stiffness. The parametric model was compared to experimental measurements with very good agreement.

INTRODUCTION

Bolt isolators and gasket are key components in an oil pan or rocker cover isolation system (illustrated in Figure 1). The function of a bolt isolator is to avoid direct contact between the bolt and the cover, and form an integrated isolation system with the gasket. The gasket isolation system has the dual function of sealing and isolation. Both the gasket and the isolator should be well designed to realize good performance in sealing and isolation[1]. To achieve the appropriate sealing stress, the elastomer gasket should be compressed to a level such that adequate sealing stress will always be maintained under a large range of temperature fluctuation, flange bending, and sealing surface out-of-flatness. The force required to compress the gasket to the compression level must be determined. Knowing the number of isolators required in the system, the amount of force required from each isolator is known. The isolators must also survive the harsh operating environment. Therefore, the load must be achieved at an appropriate level of compression. If the compression is too much, the isolator may break; if too little, it may lose load due to out-of-flatness and aging. The isolators must also fit in the space provided by the cover flange. Once the compression levels for the gasket and isolators have been determined, the grip length for the bolts are fixed based on the compressed rubber thickness of gasket and isolators, the flange thickness, gasket carrier thickness, and the isolator washer thickness. When assembled, the total compressed rubber thickness is a constant

$$l_g(1 - c_g) + l_i(1 - c_i) = C \qquad (1)$$

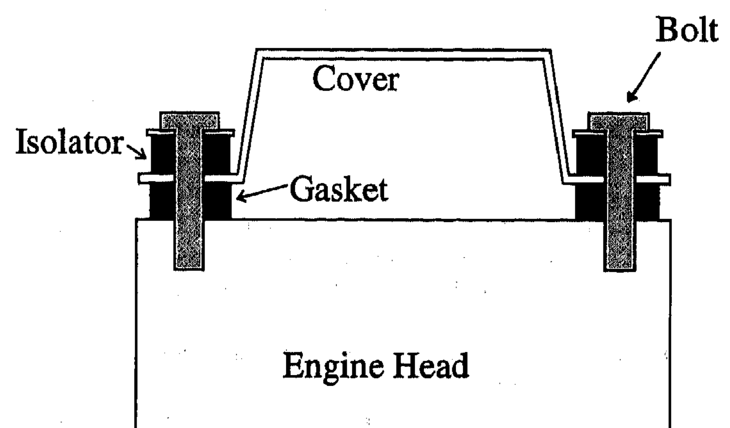


Figure 1: Illustration of a rocker cover to engine block isolation/sealing system

where l_g and l_i are the uncompressed thicknesses, and c_g and c_i are the compression levels, for the gasket and the isolators respectively, and C is a constant. Any change in the isolator compression will result in a corresponding change in the gasket compression in the opposite direction. The sealing and isolation performances will be affected. It is generally accepted that the material properties of elastomers change with many factors. But it is always advantageous to get as accurate a result as possible. Good analytical models are the key to a better system design.

The gasket design calculation can be done with the parametric modeling approach[2]. Analytical models for rubber mounts or bearings has been developed to predict the load–deflection or stress–strain relationship under special conditions. The linear elastic theory is commonly used with infinitesimal strains and incompressible materials. Under these assumptions, the following linear stress–strain equation[3]

$$\sigma = E(1 + 2s^2)\epsilon \tag{2}$$

was derived as an approximate solution to the fundamental equation:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} - \frac{12p}{t^2}\frac{G}{K} = -\frac{12G\epsilon}{t^2} \tag{3}$$

where E, G, and K are the Young's modulus, shear modulus, and bulk modulus of the material, t is the thickness of the bearing, $p(x, y)$ is the hydrostatic pressure as a function of the coordinate (x, y), ϵ is the average vertical strain, and s is the shape factor of the sample. The approximation is done over a very large range of shape factors. Bearings tested to high compressive strains display a nonlinear, stiffening load-deflection curve. The non-linear large strain models provide better approximation under finite strain. The following three theories have been proposed (see Stanton and Roeder[4] for a more detailed review);

1. Holownia[5]

$$\sigma = E_c \frac{\epsilon}{(1-\epsilon)^3} \tag{4}$$

2. Based on the kinetic theory of rubber-like solids, Gent[6] proposed

$$\sigma = -\frac{E_c}{3}(\lambda - \lambda^{-2}) \tag{5}$$

3. The third formula is a constant modulus, E, which relates incremental uniaxial stress and strain

$$\sigma = -E_c(1 - \frac{1}{\lambda}) \tag{6}$$

For small strains, all three equations reduce to

$$\sigma = E_c\epsilon \tag{7}$$

But, for large deformation, the computed stresses from the three equations could be quite different. The three non-linear models and the linear elastic model are compared in Figure 2. The three non–linear models are very close to the linear model at small strain. The difference between the four models are very large at large strain. Comparison with actual test data tell us that the linear model will under predict load, while the Holownia model will over predict load. The kinetic theory model and the incremental stress–strain model generally give better results. Still, the error can be significantly large when the shape factor is very small (less than 0.2). The apparent modulus E_c used in the models was obtained by multiplying the material Young's modulus by $1 + 2s^2$. But the constitutive theories do not describe the full complexity of the material behavior. Gent and Linley[5] found that the best fit with experimental results was obtained by introducing an empirical constant k to give

$$E_c = E * (1 + 2ks^2) \tag{8}$$

The constant k has the value 1.0 for a truly elastic material. Since the value of k is material dependent, it must be experimentally determined before it can be used. Also, the apparent modulus formula was only applicable to bearings with both surfaces bonded to rigid surfaces.

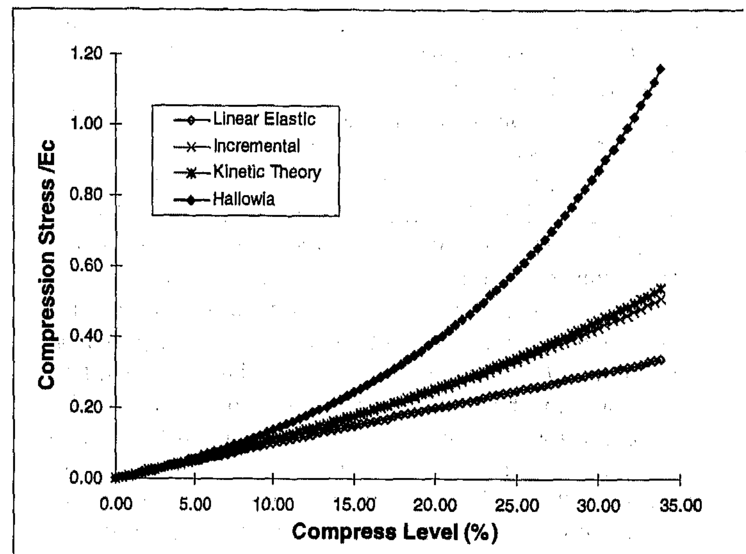


Figure 2: Comparison of the theoretical models

The bolt isolator is a special type of rubber bearing, which usually operates under large compressive strain(20% to 40%). The errors could be significant if the appropriate equation is not used. Also the bolt isolator is only bonded on one surface.

Finite Element Analysis (FEA) is a numerical method which results in an efficient, accurate and comprehensive technique for solving complex design problems. Several constitutive theories, based on strain energy density functions, have been developed for elastomeric materials. The two most widely used models for elastomer analysis are the Mooney–Rivlin model and the Ogden model. In this

paper, the axisymmetric element and the Mooney-Rivlin model are used for the FEA simulation to calculate load, deflection and the stress field. Pure shear, uni-axial tensile, and compressive tests of the elastomeric specimens are performed to provide the material properties for FEA.

FEA can deal with complex shapes and boundary conditions and eliminate the need for actual prototypes. The load–deflection curves created by FEA simulation match experimentally generated data. Figure 3 shows an empirical and an FEA simulated load-deflection curve. There is close correlation between the two methods when strain is less than 10% There are some errors when the strain is larger than 10%, but the difference is well within 15%. Since FEA result is very close to the test result, it is used as a virtual test to perform simulation on samples of different shapes.

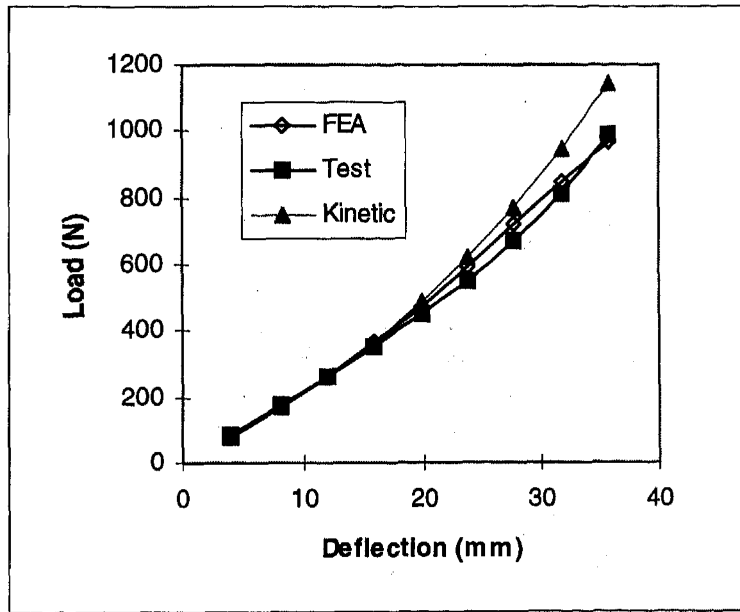


Figure 3: Comparison of FEA result with test result

PARAMETRIC MODEL

A typical bolt isolator which is illustrated in Figure 4 can be described geometrically by three parameters, the bolt hole diameter, d, the outer diameter, D, and the height, h. When a uniaxial load, F, is applied vertically at a temperature, T, the isolator will deform to a height of h_1. The load, F, is a function of the geometric parameters and material property. Mathematically, it can be written as

$$F = F(D, d, h, h_1, T, E) = \frac{\pi(D^2 - d^2)}{4}\sigma \qquad (9)$$

Since the kinetic theory model and the incremental stress–strain model are closer to the test result, and the incremental stress–strain model is simpler, the parametric model will be derived based on the incremental stress–strain model. Substituting $\lambda = 1 - \epsilon$ into equation 6, it becomes

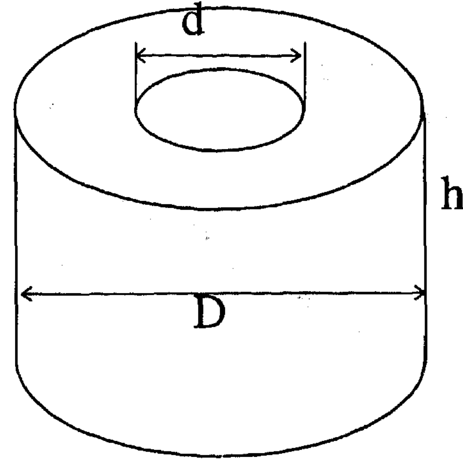


Figure 4: Illustration of a typical isolator

$$\sigma = E_c \frac{\epsilon}{1 - \epsilon} = E(1 + 2ks^2)\frac{\epsilon}{1 - \epsilon} \qquad (10)$$

or

$$\frac{\sigma}{E} = (1 + 2ks^2)\frac{\epsilon}{1 - \epsilon} \qquad (11)$$

The above equation says that for a specific material, the stress–strain relationship is only a function of shape factor s. Empirical data and FEA simulation suggest that isolators of the same shape factor, with different geometries, may not have the same stress–strain relationship Figure 5 shows the stress–strain plot of two isolators of the same shape-factor of 0.25. The difference between the stresses at 30% strain is approximately 13%. Apparently, the shape factor does not truly reflect the effect of geometry. To fix this problem, the term $(1 + 2ks^2)$ is replaced with a coefficient $\alpha(D, d, h)$, which is a function of geometric parameter. Equation 11 becomes

$$\frac{\sigma}{E} = \frac{\alpha(D, d, h)\epsilon}{1 - \epsilon} \qquad (12)$$

Fitting the FEA simulated data to the above equation resulted in a poor fit. To achieve a better curve fit, a new coefficient, $\beta(D, d, h)$, has been introduced.

$$\frac{\sigma}{E} = \frac{\alpha(D, d, h)\epsilon}{1 - \beta(D, d, h)\epsilon} \qquad (13)$$

or

$$\sigma = E\epsilon \frac{\alpha(D, d, h)}{1 - \beta(D, d, h)\epsilon} \qquad (14)$$

This is the general model for a bolt isolator. The material property is represented by E, and the effect of geometry by the two coefficients α and β as functions of D, d, and h. The temperature effect can be represented by a temperature dependent modulus E.

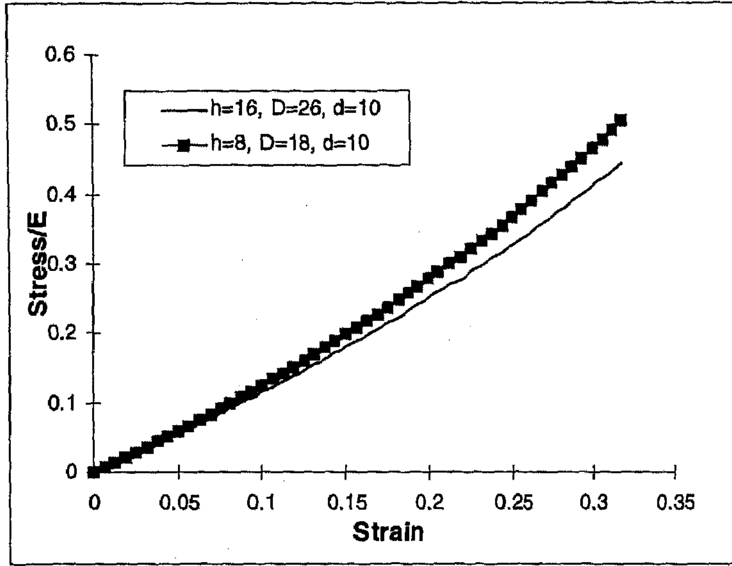


Figure 5: Stress–strain plot of two isolators of the same shape factor

This model is expected to work in a practical range of $D_{min} < D < D_{max}$, $d_{min} < d < d_{max}$, and $h_{min} < h < h_{max}$. The limit values D_{min},D_{max}, d_{min},d_{max}, h_{min} and h_{max} are determined by the type of application. In the cover and oil pan application, e.g., the outer diameter is rarely smaller than $15mm$ or larger than $30mm$. The bolt hole diameter should be between $10mm$ to $15mm$, and the height between $8mm$ to $16mm$.

The recommended modeling procedures are as follows:

1. For a given material, and any given set of geometric parameters D, d, and h, obtain the engineering stress-strain data through FEA simulation.

2. Fit the stress-strain data to the function

$$\sigma = E\epsilon\frac{\alpha}{1-\beta\epsilon} \tag{15}$$

   and obtain the two coefficients α and β.

3. Fit the coefficient α and β to the parameters D, d, and h to get the equation

$$\alpha = \alpha(D, d, h) \tag{16}$$

$$\beta = \beta(D, d, h) \tag{17}$$

A MODELLING EXAMPLE

Let the bolt hole diameter $d = 10mm$, $D_{min} = 16mm$, $D_{max} = 26mm$, $h_{min} = 8mm$, and $h_{max} = 16mm$. There are 30 combinations of parameters if ΔD and Δh are both $2mm$. For isolators of each size, perform FEA simulation and obtain the stress–strain curves. The coefficients α and β were determined by fitting the stress–strain curves to equation 15. Table 1 listed some selected coefficients.

d	h	D	α	β
10	16	16	0.9441	0.2487
10	8	18	1.1450	0.8776
10	8	26	1.3635	1.4426
10	12	16	1.0050	0.2207
10	12	18	1.0351	0.4783
10	12	20	1.0657	0.6437
10	12	22	1.0970	0.7869
10	12	24	1.1299	0.8914
10	12	26	1.1737	0.9401
10	10	20	1.1212	0.7991
10	14	24	1.0809	0.7416
10	16	26	1.0655	0.7328
10	8	20	1.1845	1.0727
10	8	22	1.2433	1.1837
10	8	24	1.3124	1.2673
10	10	22	1.1551	0.9524
10	10	24	1.2036	1.0454
10	10	26	1.2580	1.1229

Table 1: Selected coefficients α and β for the stress–strain curves for different isolators

The coefficients α and β for all the different isolators will be fitted to the following equations

$$\alpha = a_0 + a_1\frac{D}{h} + a_2\frac{d}{h} + a_3\frac{d}{D} + a_{-1}\frac{h}{D} + a_{-2}\frac{h}{d} + a_{-3}\frac{D}{d} \tag{18}$$

$$\beta = b_0 + b_1\frac{D}{h} + b_2\frac{d}{h} + b_3\frac{d}{D} + b_{-1}\frac{h}{D} + b_{-2}\frac{h}{d} + b_{-3}\frac{D}{d} \tag{19}$$

The fitted results are shown in table 2 and Figure 6.

a_0	a_1	a_2	a_3	a_{-1}	a_{-2}	a_{-3}
0.6152	0.2991	-0.3678	0.6491	-0.2029	0.0219	0.0134
b_0	b_1	b_2	b_3	b_{-1}	b_{-2}	b_{-3}
2.7503	-0.5071	2.4078	-8.0999	3.3498	-1.2688	0.1476

Table 2: Fitted coefficients for parameters α and β of the stress–strain curves for isolators with both sides bonded

The modeling results are compared with other methods in Figure 7 to Figure 9. The model agrees with the FEA results for every geometry. The above approach was also applied to isolators with one side bonded and the other side thoroughly lubricated. The coefficient for α and β are shown in table 3. Figure 10 shows the results of

a_0	a_1	a_2	a_3	a_{-1}	a_{-2}	a_{-3}
0.8351	0.1258	-0.0961	0.1800	-0.1328	0.0377	-0.0478
b_0	b_1	b_2	b_3	b_{-1}	b_{-2}	b_{-3}
0.8460	1.0422	-1.7652	0.6792	-1.0666	1.0623	-0.7818

Table 3: Fitted coefficients for parameters α and β of the stress–strain curves for isolators with one side bonded and the other side lubricated

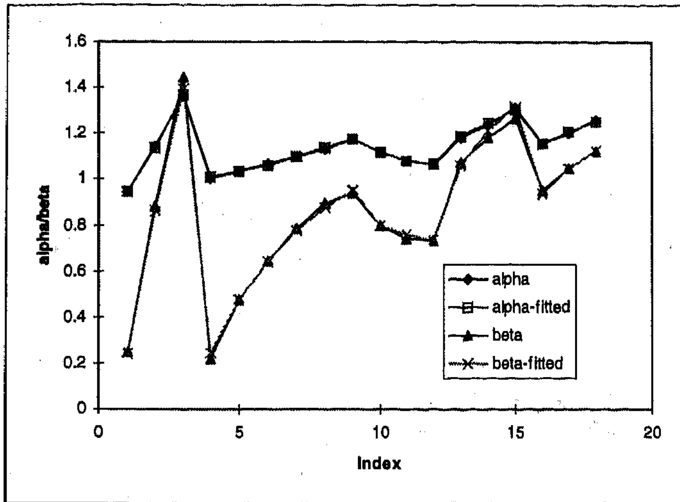


Figure 6: Fitted results of α and β

two isolators, one with both sides bonded and the other with one side bonded and one side lubricated. For the two conditions, the structure of the models are the same (equation 15). But the coefficients that relate α and β to the geometric parameters D, h, and d are different. Both parametric models fit the FEA results very closely, while the other models differ significantly.

It has been demonstrated that FEA simulation results closely agree with test results, and the parametric model derived from FEA simulated load–deflection curves matches FEA results. Therefore, it is expected that the parametric model will match the actual test result. Two one–side bonded isolators of dimensions ($D = 26mm$, $d = 11.68mm$, $h = 8.8mm$) and ($D = 22.6mm$, $d = 9.8mm$, $h = 10.03mm$) and Young's modulus of $5.85MPa$ were tested and compared with the parametric model. As demonstrated in Figure 11, the same model closely matches the test results in both test cases. The maximum error is within 10% of the test results. The error is the accumulation of errors from FEA simulation to test and parametric model to FEA simulation. If the parametric model were derived from test data instead of FEA simulated data, the error would be smaller. But the advantage of simulating many different isolators without actually making the parts would be lost.

APPLICATION

The FEA simulation can produce a result which is very close to the test result, and the parametric model is in close agreement with with the FEA simulation. Therefore, it is safe to conclude that the parametric model also agrees with the test result. Once a load-deflection model is determined, the stiffness can be determined mathematically from the load-deflection model. A design engineer can use

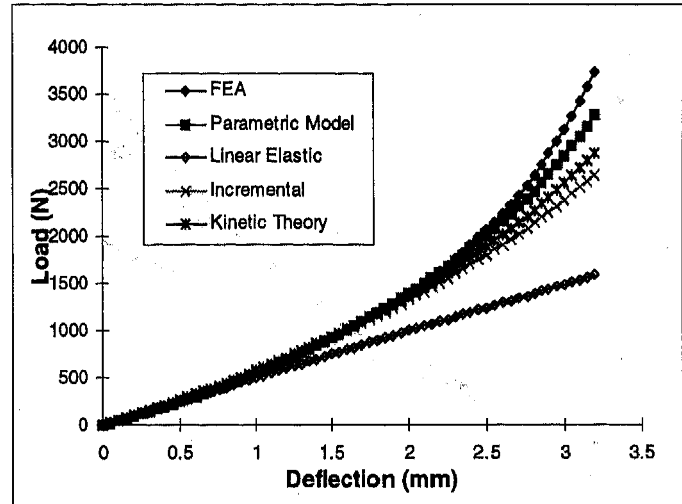


Figure 7: Comparison of Parametric Model with Linear and Non-Linear Models (D=26mm, d=10mm, h=8mm, E=5.85MPa)

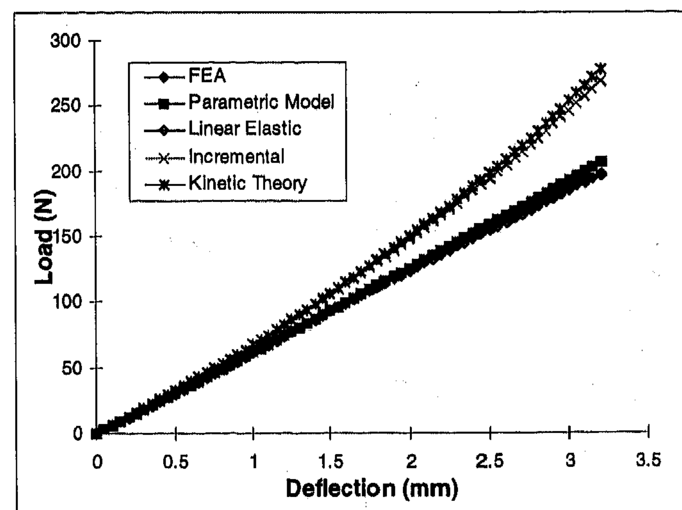


Figure 8: Comparison of Parametric Model with Linear and Non-Linear Models (D=16mm, d=10mm, h=12mm, E=5.85MPa)

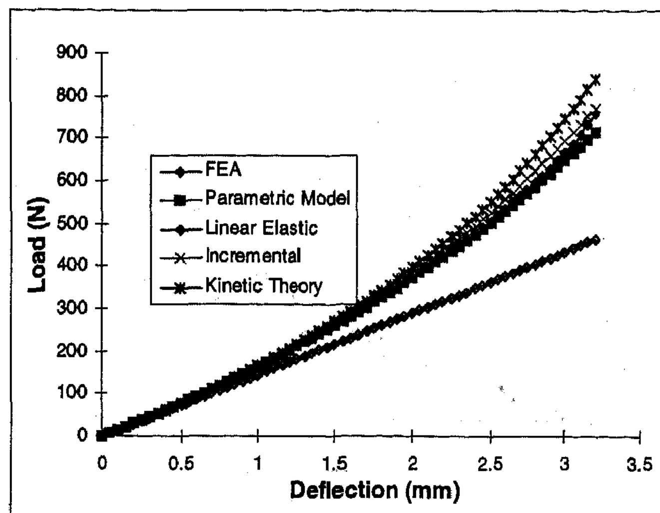


Figure 9: Comparison of Parametric Model with Linear and Non-Linear Models (D=18mm, d=10mm, h=8mm, E=5.85MPa).

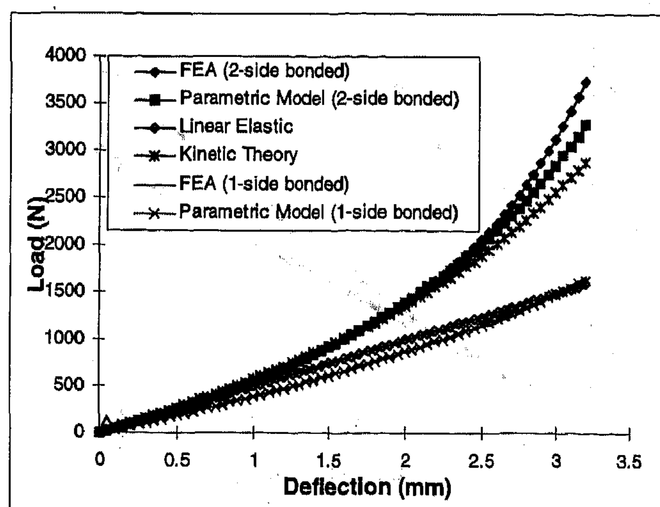


Figure 10: Comparison of Parametric Model with Linear and Non-Linear Models (D=26mm, d=10mm, h=8mm, E=5.85MPa).

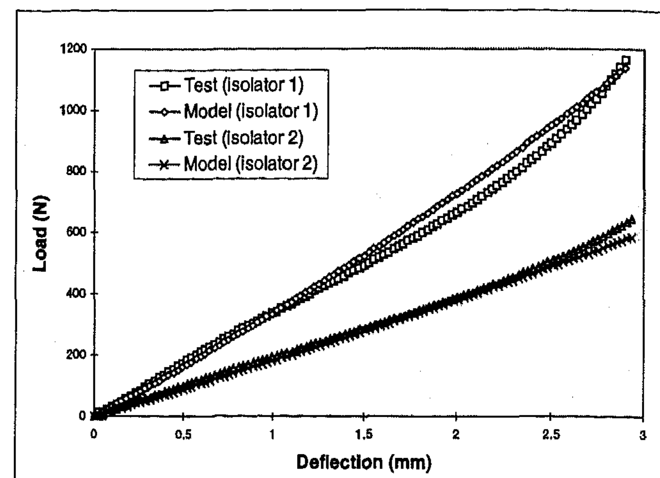


Figure 11: Comparison of Parametric model with test results

the model to determine the desired geometry and select the appropriate material based on the load required at a specific deflection. The one-time FEA modeling of a few samples of one material is extended to many different geometries and materials. The resultant parametric model can be used repeatedly in design and analysis.

CONCLUSION

FEA model can be used to generate load-deflection data which is very close to the actual test data from a load frame. Based on the large strain non-linear stress-strain model, a modified stress-strain relationship was established to better represent the test and FEA simulated data. The modified stress-strain equation determines the basic structure of the parametric model. The relationship between the coefficient of the stress-strain equation and the geometric parameters were determined through curve fitting. The resultant model agrees with the FEA and test results. The parametric model can be used in engineering design without repeated employment of FEA and can be used as an engineering tool for design, analysis, and optimization.

REFERENCE

1. James R. Zwick, Zhi-Quan Hu, and Gary Novak, "A design/analysis method of gasketed bolted joints for noise/vibration control", *SAE Paper 952087.*

2. Zhi-Quan Hu, T. Y. Chen, and Gary Novak, "A Method for Determining the Load-Deflection of an Elastomeric Gasket", *SAE Paper 960217.*

3. Gent, A. N., and Lindley, P. B., "The Compression of Bonded Rubber Blocks." *Instn. Mech Engres. Proc.*, Vol. 173 pp. 111–122, 1959.

4. Stanton, J. F., and Roeder, C. W., "Elastomeric Bearings Design, Construction, and Materials." National Cooperative Highway Research Program Report 248, 1982.

5. Holownia, B. P., "Compression of Bonded Rubber Blocks." *Journal of Strain Analysis*, Vol. 6, No. 2, 1971.

6. Gent, A. N., "Load–Deflection Relations and Surface Strain Distributions for Flat Rubber Pads." *Rubber Chemical Technology*, Vol. 31, pp. 395–414, 1958.

970524

The Effect of Flange Flexibility on the Response of Gasketed Bolted Joints Subjected to External Forces

Gary Novak, Zhi-Quan Hu, Michael Sadowski, and Edward Widder
Fel-Pro Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The response of a gasketed, bolted joint to an external load is understood to be effected by all components involved in the joint. The analysis involves the equilibrium of the gasket compressive force with the bolt tension force and the forces external to the bolted joint. Geometric compatibility is maintained when the change in the stretch of the bolt caused by an external force is equal to the change in compressed thickness of the gasket. When the flanges are treated as nondeformable, the classical joint diagram analysis indicates that externally applied loads, which unload the gasket, increase bolt tension. In this paper, the effect of flexible flanges is included in the analysis of simple gasketed bolted flanges. The results show that bolt tension response can deviate significantly from the rigid flange behavior. In certain situations where flanges have a relatively high level of flexibility, external joint forces that unload the gasket also unload the bolt.

INTRODUCTION

The need to predict gasket stresses is driven by goals to optimize design, maximize performance, reduce design iterations, and compress overall product development time and cost. The assembly condition and the load cases that are expected to cause extreme gasket loads need to be evaluated. Gasket stresses under these conditions can be used to evaluate the expected performance of the gasket as both a structural member and seal. Initial as assembled, in service operation, and long term performance can be estimated by using initial geometry and material properties, end of life geometries and material properties, failure models that look at stresses and deflections at a point in time, or failure criteria that include time effects such as Weibull or cumulative damage methods. Regardless of which analysis is being conducted, an accurate prediction of gasket forces or stresses is needed.

The gasket is one part of the bolted joint that also includes the fasteners and the flanges. The flanges can be very complex structural components such as cylinder blocks or heads. The performance of all of these components must be considered when evaluating the performance under load of any one component. Joint diagram analysis [1] and finite element analysis, FEA, [2] are two approaches that are commonly used to evaluate bolted joints subject to a variety of assembly and operational loads. Both approaches are used to predict fastener and gasket deformation and load distribution when a bolted joint is subjected to various loading conditions.

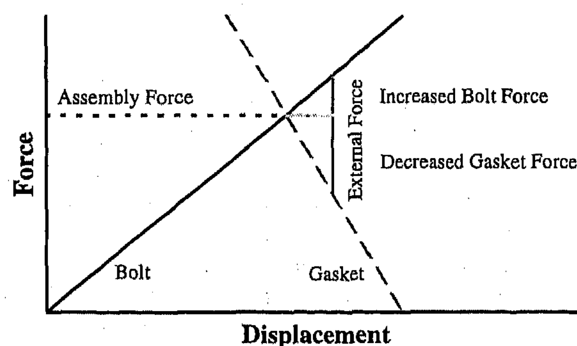
In this work, a combination of experiment and FEA was used to study the effect of flange flexibility on the response of gasketed bolted joints subjected to external forces. The work demonstrated significant differences between the flexible flange mechanics and the idealized behavior modeled by the joint diagram. Two examples were analyzed and experimentally verified to show how flange flexibility can cause bolt force levels to be less than values predicted by joint diagram analysis. Both analysis and experiment demonstrate a decrease in bolt tension in response to a joint separating force under certain conditions.

BACKGROUND

Joint diagram models are extremely effective at predicting the joint equilibrium forces and deformation of idealized joints. These idealized

conditions require that the flanges are rigid or that the flange deformation is uniform around the joint. Joint diagrams predict equilibrium assembly bolt stretch and gasket compression when total fastener tension, fastener stiffness, gasket stiffness, and idealized flange stiffness are specified. Nonlinear gasket material models with inelastic recovery can be included in joint diagram analysis [3]. Joint diagrams are used to redistribute external forces between the fastener and the gasket to determine increases and decreases in fastener and gasket total force (Figure 1). These external forces include both joint separating forces such as in-cylinder combustion forces and forces that tend to increase gasket compression. Joint diagrams have even been used to predict the effects of thermal expansion on the equilibrium of forces and geometric compatibility in these idealized bolted joints.

Figure 1
Joint Diagram



In many real world applications of gasketed bolted joints, flanged connections do not behave as the idealized bolted joints with rigid uniform flanges or uniform flange deflection. In these cases, the flange deflection is significant and varies along the sealing path of a gasket, flange thickness and stiffness varies, or there is a significant out of flatness that can cause the gasket loading to vary significantly around the sealing path. The joint diagram cannot accurately predict the gasket contact stress distribution or gasket and fastener equilibrium conditions. The joint diagram analysis does not include the effects of nonuniform flange deflection, flange flatness, bolt load nonuniformity, or gasket nonuniformity. Joint diagrams have been used to qualitatively describe the behavior of a bolted joint when flange deformation effects are included [4,5]. The response of a bolted joint to an external force is predicted by joint diagram analysis as being split between the gasket and fastener in proportion to their relative stiffness. An external force that separates the flanges increases the bolt force and decreases the gasket compressive force while maintaining force equilibrium; gasket compressive force plus external force equals bolt force. Conversely, an external force that tends to compress the flanges decreases the bolt force and increases the gasket compressive force in proportion to their relative stiffness while also maintaining force equilibrium. Joint diagram analysis indicates that separating forces increase total fastener tensile force while decreasing total gasket compression force.

In situations where flange deformation is significant and large variation in gasket contact stresses occur, joint diagram methods are not effective. Analytical methods involving the simulation of deformable flanges have been used in these situations through closed form solutions, approximate methods, and customized computer formulations[6,7,8]. Finite element analysis, FEA, is a numerical analysis technique that is widely used to analyze problems involving complicated geometries where material models, loading conditions, or nonlinearities make closed form solutions or custom programs difficult or impossible to use. FEA has been used to include the effects of flange flexibility in the analysis of gasketed bolted joints [1,9]. Nonlinear material properties, deformation due to complex flange geometry, and flange flatness have been included in these analyses. The gasket contact stresses and compression distributions are the primary outputs of an FEA; while the total gasket force and the gasket compression are determined from a joint diagram analysis of a rigid or uniform deformable flange system.

Gasket compressive deformation is in equilibrium with the bolt stretch and the flange deflection. Under the bolt, flange deflection is due to the through thickness flange compressibility. Between bolts, bending causes flange deformation. In the cases where flanges are relatively thin, under the bolt flange compression is negligible compared to the very significant bending effect that takes place away from the bolts. Less gasket compression develops at locations farther away from the bolts than occurs at locations closer to the fasteners; because thin flexible flanges cause the midspan flange deformation to be large. For uniform thickness gaskets, gasket compressive stresses are maximal near the fasteners where high levels of gasket deformation cause high compressive strain levels. Stresses are lower at those areas farther from the fasteners where gasket compressive strain levels are reduced compared to the levels under the bolt. Flange flexibility effects cause the total gasket and bolt forces to be less than the values that would be predicted by the rigid flange models when the joint is assembled to a specific level of interference (under the bolt gasket compression plus bolt stretch). When

external loads are applied to a bolted joint, flange flexibility affects the way that the unloading gasket clamp force is distributed. The force is not applied as a uniform stress acting over the total contact area. Changes in the total gasket and bolt forces can be significantly different from those predicted through joint diagram type analysis. Bolts force increases, in response to joint separating forces, are less than those predicted by joint diagram effects. In certain situations of flexible flanges with high stiffness fasteners, separating forces can decrease bolt tensions rather than increase them.

GASKETED JOINT SAMPLE PROBLEM

There are many applications where external loads are introduced into a gasketed bolted joint. The external load changes the joint conditions and affects the sealing performance of the joint. External loads are often introduced into a joint when the internal pressure within a bolted joint changes. The effect of the external load is predicted through traditional joint mechanics, where forces and displacements must balance. Joint diagrams are often used to represent these fundamental relationships between forces and displacements. In a rigid flange model, the interactions between the external force and the joint can be readily predicted and even depicted on a joint diagram. On the other hand, when the deflection of the flange becomes significant, a more complex joint model is necessary to predict the reactions in the joint. To demonstrate this complex interrelationship between the joint and the external load, a simple sample joint was designed. This sample joint will be used to demonstrate the impact an external load has on the joint equilibrium.

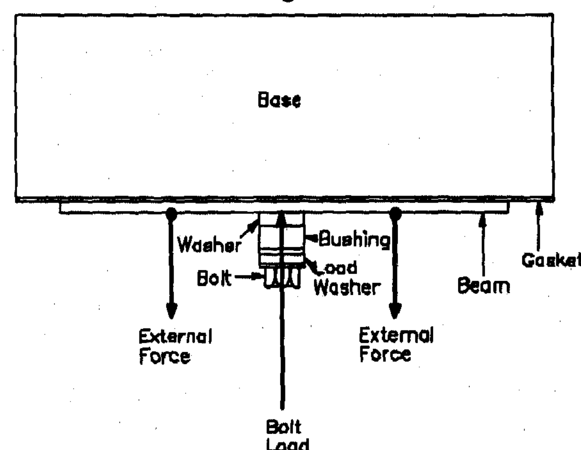
The sample joint consists of two flanges connected together with a single bolt at the center. Each flange is 25 mm (1 inch) wide and 260 mm (10.25 inches) long. One flange was 102 mm (4 inches) thick and represents a rigid foundation. A single layer of a gasket facing was placed between the flanges; the gasket facing was 0.76 mm (0.030 inches) thick. This joint was assembled with a 12.7 mm (0.5 inch) bolt to a load of 4448 N (1,000 lbs). Two 118 N (26.5 lb) external lifting loads were placed on the thinner flange on either side of the bolt. The loads were located 3 inches from the center of the bolt. The 118 N (26.5 pound) loads interact with the joint producing changes in bolt load, gasket load and flange deflection.

ANALYTICAL MODEL OF SAMPLE JOINT - The sample joint was analyzed by modeling the joint and applying the external loads. The analytical model includes flange deflections, non uniform load distributions on the gasket, non-linear gasket material properties, and a linear spring model for the bolt. The beam was divided into 100 nodes along its length for the analysis. The sample problem was analyzed with both a 50.8 mm (2 inch) thick flange and with a 6.35 mm (0.25 inch) thick flange.

In the analysis of the rigid 50.8 mm (2 inch) thick beam, the bolt load increased when the external load was applied. The bolt load increased from 4448 N (1,000 pounds) to 4461 N (1003 pounds) in response to the 236 N (53 pound) external load. In contrast, when the external load was applied to the flexible beam, the bolt load decreased. The initial load of 4448 N (1,000 pounds) dropped to 4186 N (941 pounds) with the addition of the 236 N (53 pound) external load.

EMPIRICAL RESULTS OF SAMPLE JOINT - To verify the gasketed bolted joint phenomena exhibited by the analysis, an experiment was devised to provide the application of an external load to a bolted joint. The bolted joint had the same configuration of the sample joint described earlier. The external loads were applied through a pair of eye bolts that were attached to the beam, equidistant from the bolt. The applied bolt load was measured using a bolt load washer transducer. A spacer bushing was placed between the beam and the load transducer to reduce the effects of non-uniform loading on the transducer due to beam bending. The external load was applied to the beam using a dead weight hung by a cable arrangement. This fixture insured that the load was distributed equally on either side of the bolt. This test arrangement is depicted schematically in Figure 2; the actual test fixture configuration is shown in Photo 1.

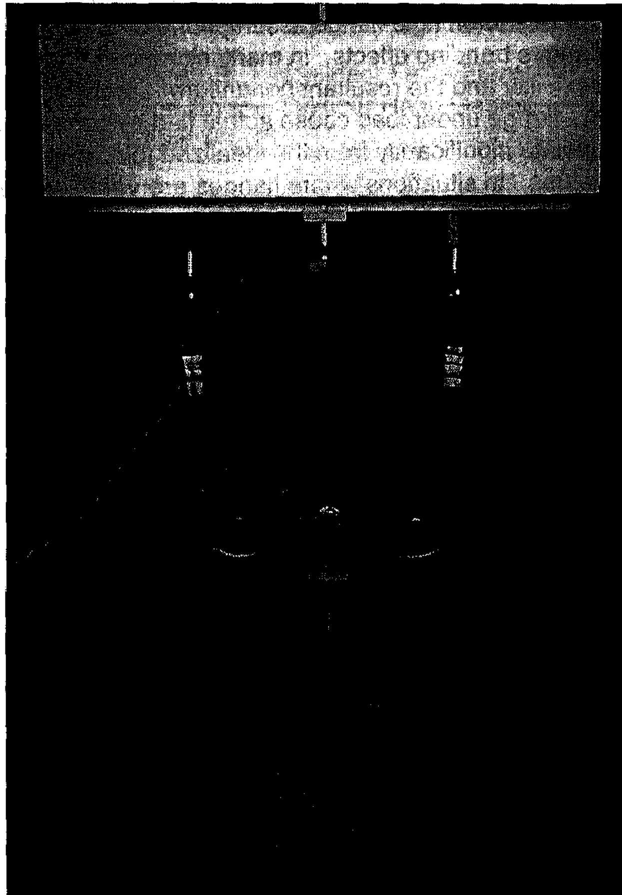
Figure 2



Following the tightening of the bolt, the applied load is observed to decrease with time. This behavior is typical of creep-relaxation characteristics

of gasketed, bolted joint systems. In order to reduce the effect of this time dependent load decay, the bolted assembly was allowed to stabilize for 2 hours, before the external load was introduced.

Photo 1



A bolt load of 4448 N is initially applied to the beam. When an external load of 236 N is applied in an opposite direction to the beam flange, the change in bolt load is measured, also using the bolt load washer transducer. This change in bolt load was recorded for two different flange thicknesses, 6 and 51 mm, for the same conditions of applied bolt load and external load. The thicker flange approaches the stiffness of an infinitely rigid beam, for the dimensional and loading conditions of this experiment. It can be observed that, with the thinner flange, the bolt load is seen to decrease when the external load is applied. The magnitude of this load loss decreases with increasing flange (beam) thickness. With a very thick flange, the observed load change is positive, but still less than the external load value, which is consistent with rigid flange joint mechanics theory.

DISCUSSION OF RESULTS

Both the analytical model and the empirical tests demonstrated the importance of including flange deflection into a joint analysis. The flange deflections significantly affect the load balance within the joint. A comparison between the analytical results and the empirical results for both a rigid flange and a flexible flange, are shown in Table 1. The rigid flange model represents the results that would be predicted through simple joint mechanics. The flexible flange model demonstrates the importance of including the flange deflection into the joint analysis.

Table 1
Change in Bolt Load After the Addition of a 236 N External Load

	Analytical Model	**Empirical Test**
Rigid Flange		
mean	+13 N (+3 lbs)	+18 N (+4 lbs)
SD		9 N (2 lbs)
Flexible Flange		
mean	-262 N (-59 lbs)	-156 N (-35 lbs)
SD		18 N (4 lbs)

In a rigid flange model, the impact of the external load can be predicted through joint mechanics or through a joint diagram. Prior to the addition of the external load, force equilibrium requires the total load on the gasket to equal the total load on the bolt. After the external force is added, force equilibrium requires that the total load on the gasket plus the external load must equal the bolt load. The redistribution of the loads within the rigid joint is a function of the stiffness ratio between the bolt and the gasket. Part of the 236 N (53 pound) external load will increase the bolt load and part of the 236 N (53 pound) external load will decrease the total gasket load. The rigid sample problem demonstrated the traditional response to an external load. When the external load was applied, the bolt load increased by a fraction of the applied load. The analytical model predicted an increase of 13 N (3 pounds). The empirical test showed an increase of 18 N (4 pounds). The balance of the 236 N (53 pound) external load appears as a decrease in load on the gasket.

In the more complex situation with a flexible flange, the redistribution of load is more complicated. When the flange is only 6.35 mm (0.25 inches) thick, it becomes very difficult to predict the effect of the external load through simple joint mechanics or a joint diagram. In the analytical model the bolt showed a decrease in load of 262 N (59 pounds) when the

external load was applied. Likewise, the empirical tests showed a decrease of 156 N (35 pounds). When the external load was applied, the flexible flange produced a decrease in bolt load rather than the increase in bolt load that was observed with the rigid flange. With the decrease in bolt load, the gasket load also drops. The gasket looses both the 262 N (59 pounds) that the bolt lost and the additional 236 N (53 pounds) from the external load for a total loss of 498 N (112 pounds).

As demonstrated, the reaction of a joint to an external load can vary significantly depending on the joint configuration. In the previous example, the effect of the flange stiffness on the joint interactions was demonstrated. Whenever there is a change in the deflection profile of the flange, there is a change in the joint balance. In the case of a rigid joint, the external loads do not significantly change the deflection profile or the distribution of load. Under these conditions, the effects of the external loads are easy to incorporate into the joint using convention joint mechanics or the joint diagram. However, when the flange thickness was reduced, the external loads caused large changes in the deflection profile of the flange. It was the change in the flange deflection profile that redistributed the loads in the joint and produced a significant change in the joint balance. In this example, the load was redistributed toward the bolt. The increase in gasket compression force under the bolt decreased the gasket thickness which produced a loss in bolt stretch.

To further demonstrate the impact of changing the deflection profile of the flange, a second comparison was modeled. In this comparison, the deflection profile was changed by moving the location of the external loads instead of changing the flange thickness. In both cases, a 6.35 mm (0.25 inch) thick flange was used with the rigid foundation. In the first case, two 118 N (26.5 pound) loads were placed at a distance of 114 mm (4.5 inches) from the bolt. The bolt load decreased from 4448 N (1,000 pounds) to 3861 N (868 pounds) and the gasket load decreased from 4448 N (1,000 pounds) to 3625 N (815 pounds). The external load had a significant impact at changing the deflection profile of the flange. In the second case, the external loads were applied at 12.7 mm (0.5 inches) from the bolt. In this case, the bolt load increased from 4448 N (1,000 pounds) to 4586 N (1031 lbs) and the gasket load decreased by 98 N (22 pounds). The application of the external loads at positions where they have little effect on the deflection profile of the flange caused the total gasket load to decrease 98 N (22 pounds) instead of 823 N (185 pounds).

CONCLUSION

The gasket and fastener forces predicted by joint diagram analysis are based on force equilibrium and geometric compatibility between the gasket, fasteners, and flanges. Joint diagram constructions are based on rigid flange assumptions or uniform flanges bending effects. In many situations flange flexibility and the resultant nonuniform deformation of the flange under load cause actual flange behavior to deviate significantly from the idealized joint diagram values. In situations where flanges are very flexible, the deviation between the rigid flange approximation and the actual behavior can be significant. In this study, FEA and experiment demonstrated that predicted increases in bolt tensions due to a separating external force were less than the increases predicted by the rigid flange analysis. It was further demonstrated both empirically and analytically that under conditions of low flange stiffness and high bolt stiffness, a separating external force can cause bolt tension to decrease. An analysis technique that includes the effects of flange flexibility should be used to simulate the as-assembled and loaded conditions of bolted joints.

ACKNOWLEDGMENTS

Thanks and appreciation to Don Bajner for the many hours of testing and his dedication to generating accurate results.

REFERENCES

1. Czernik, D., "Gasketing in the Internal Combustion Engine", SAE Paper 800073.

2. Robinson, M., Gavin, W., Tronel, T., Schulze, S., "Applications of Advanced Finite Element Techniques to the Design and Development of Automotive Gaskets", SAE Paper 960218.

3. Widder, E., Novak, G., "Gasketed Joint Analysis Using Computer Aided Engineering Techniques", SAE Paper 920131.

4. Widder, E., Sadowski, M., Novak, G.,"Gasketed Joint Analysis Including Flange Bending Effects", SAE Paper 930120.

5. Bickford, John, An Introduction to the Design and Behavior of Bolted Joints, Third Edition, Marcel Dekker, New York, 1995.

6. Oren, J., "Design Considerations of Deflection in a Rigid Flange Gasket Joint", SAE Paper 830216.

7. Sawa, T., Kumano, H., "On the Characteristics of Bolted Joints with Gaskets", Bulletin of JSME, Vol 27, No. 234, 1984, pp 1635-2642.

8. Nau, B., "Computer Modeling of the Sealing Behavior of Gaskets in Flanged Joints", Nuclear Engineering and Design 90, 1985, pp181-191.

9. Graves, S, Utley, T., "Nonlinear Finite Element Analysis of Diesel Engine Cylinder Head Gasket Joints", SAE Paper 932456.

970525

Evaluation of Ikon®-12 Refrigerant for Motor Vehicle Air Conditioning

James J. Jetter and N. Dean Smith
U. S. Environmental Protection Agency

Krich Ratanaphruks, Angelita S. Ng, Michael W. Tufts, and Francis R. Delafield
Acurex Environmental Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A proprietary refrigerant, called Ikon®-12, was evaluated as an alternative to HFC (hydrofluorocarbon)-134a for automotive air conditioning. The evaluation was motivated by concern over the relatively high global warming potential of HFC-134a. In preliminary tests, Ikon®-12 was found to be compatible with a polyolester lubricant and engineering materials. Refrigeration capacity and efficiency for Ikon®-12 compared favorably to those for HFC-134a. In a preliminary durability test, Ikon®-12 refrigerant showed no significant chemical breakdown after extended operation with an elevated compressor discharge temperature. Further testing would be required to determine if stability and materials compatibility are acceptable for long-term use.

INTRODUCTION

HFC-134a is currently used as the refrigerant in new and retrofitted motor vehicle air conditioners, replacing the stratospheric ozone depleting refrigerant, CFC (chlorofluorocarbon)-12. Although HFC-134a has no ozone depletion potential, it has a global warming potential approximately 3300 times that of CO_2 (carbon dioxide) over a 20 year time horizon, 1300 times that of CO_2 over a 100 year time horizon, or 420 times that of CO_2 over a 500 year time horizon.[1] By the year 2005, approximately 260 million vehicles with HFC-134a air conditioners will be in service in developed countries and will require an estimated 74,800 metric tons of refrigerant annually including 33,400 metric tons for new vehicles and 41,400 metric tons for service usage.[2] Refrigerant emissions due to service usage represent an annual, CO_2-equivalent of 53.8 million metric tons, based on a 100 year time horizon. In 1994, the estimated total U.S. greenhouse gas emissions were 6109 million metric tons CO_2-equivalent.[3] Increased concern over global climate change could eventually lead to restrictions on HFC-134a consumption. For this reason, the U.S. EPA (Environmental Protection Agency) has evaluated alternatives for HFC-134a in motor vehicle air conditioning.[4]

A proprietary, near-azeotropic blend of an FIC (fluoroiodocarbon) and an HFC, called Ikon®-12, has been proposed as a refrigerant for automotive air conditioning and other applications.[5] The global warming potential for Ikon®-12 has been reported to be 4 percent of that for HFC-134a.[5] Ikon®-12 has been reported to be nonflammable according to testing based on ASTM (American Society for Testing and Materials) methods E681-85 and E918-83.[6-7] Extensive toxicity testing has been reported for both components of the refrigerant blend. Reported results have been generally favorable, although toxicity may be somewhat higher than for HFC-134a. Under EPA's SNAP (Significant New Alternatives Policy) Program, Ikon®-12 has been approved for use in motor vehicles subject to use conditions including the use of unique fittings, the use of descriptive labels, a prohibition against topping off one refrigerant with another, and appropriate recovery or recycling equipment.[8]

LUBRICANT MISCIBILITY

Miscibility was tested for the Ikon®-12 refrigerant with a PAG (polyalkylene glycol) lubricant and a POE (polyolester) lubricant. The PAG lubricant was supplied by an OEM (original equipment manufacturer). Its viscosity was reported to be 135 cSt at 40°C and 25 cSt at 100°C. It was reported to contain proprietary additives. The POE lubricant was supplied by a chemical manufacturer. Its viscosity was reported to be 134 cSt at 40°C and 15 cSt at 100°C. It also contained proprietary additives. Sealed glass tubes were prepared with the Ikon®-12 refrigerant and lubricant at lubricant concentrations of 2, 5, 10, 20, and 30 weight percent.

Complete miscibility with the PAG lubricant was found at low temperatures with all sample concentrations remaining miscible down to -45°C. However, separation of the refrigerant and PAG lubricant occurred at approximately 95°C for the 2 and 5 weight percent samples. The 10, 20, and 30 weight percent samples remained miscible up to 125°C, at which point the test was terminated.

Miscibility with the POE lubricant was tested over the temperature range of -45 to +115°C. No phase separation was observed for all concentrations, indicating complete miscibility of the Ikon®-12 refrigerant and the POE lubricant at the conditions tested.

STABILITY AND MATERIALS COMPATIBILITY

As a preliminary evaluation of the thermal and hydrolytic stability of Ikon®-12 and its compatibility with common materials of construction used in automotive air conditioners, sets of sealed glass tube samples were prepared. Materials were selected in preparation for testing the refrigerant in an automotive air-conditioning system designed for HFC-134a. Materials compatibility and stability were tested in accordance with the methods described in ANSI/ASHRAE Standard 97-1989.[9] Stability tests were conducted by sustained heating of the Ikon®-12 refrigerant by itself and with copper, aluminum, and steel in sealed glass tubes at 175°C for 14 days. Compatibility tests were conducted by sustained heating of elastomers, polymers, desiccant, cast iron, and brass with the refrigerant at 125°C for 14 days. One set of tubes was prepared without any lubricant, one set was prepared with PAG lubricant, and one set was prepared with POE lubricant. Duplicate samples were prepared and tested for each combination.

Sealed tubes containing Ikon®-12 refrigerant by itself were visually inspected after sustained heating at 175°C for 14 days. The fluid changed from colorless to purple color, indicating the presence of iodine. Volatiles released upon breaking the tubes under vacuum were analyzed by gas chromatography, mass spectrometry, and infrared spectrometry. The resultant chromatograms and spectra were compared to those obtained for pure refrigerant. A small amount of breakdown of the FIC was found as evidenced by the formation of a minute quantity of CF_3H (trifluoromethane). The addition of copper, aluminum, and steel did not affect the results significantly. The addition of a small amount of water also did not affect the results.

Sealed tubes containing Ikon®-12 refrigerant in combination individually with buna-n, neoprene, 6/6 nylon, Teflon, cast iron, and brass were maintained at a temperature of 125°C for 14 days. Chemical analysis of the volatiles released from those tubes indicated the formation of a minute amount of CF_3H.

Visual inspection of sealed tubes containing Ikon®-12 and PAG lubricant showed substantial decomposition as evidenced by the presence of carbonized solids after the 14-day aging period. Vapor-phase contents of one tube containing only the Ikon®-12 and PAG lubricant were analyzed by gas chromatography and mass spectrometry. A host of degradation products were evident from the analyses indicating that the incompatibility of Ikon®-12 and PAG lubricant was independent of the presence of other materials. Vapor-phase components identified from the analyses suggest that the PAG lubricant and both components of the Ikon®-12 blend decomposed. Analysis of the contents of other tubes containing materials with the PAG lubricant was not continued because of the obvious instability. The PAG lubricant was considered unacceptable for use with the Ikon®-12 refrigerant.

Sealed tubes containing Ikon®-12 and the POE lubricant appeared from visual inspection to have better chemical stability than the tubes containing the PAG lubricant. Stability of the refrigerant and lubricant combination was tested by sustained heating at 175°C for 14 days. Volatiles released upon breaking the tubes were analyzed. The addition of the POE lubricant resulted in a greater decomposition of the FIC than without the lubricant, but only a relatively small amount of decomposition was found. Further work would be required to quantify the decomposition. For comparison, HFC-134a with PAG lubricant typically has no discernible breakdown at 175°C. However, mineral oil lubricant used with CFC-12 refrigerant typically has substantial breakdown at 175°C.

When copper, aluminum, and steel were added to the mixture of refrigerant and lubricant, greater decomposition of the FIC was found as evidenced by the formation of larger concentrations of CF_3H and the deposition of salts on the metal surfaces. The combination of Ikon®-12, POE lubricant, and metals also yielded trace amounts of CO (carbon monoxide) and monofluoroethene. The presence of the monofluoroethene indicated a probable degradation of the HFC component of the refrigerant blend.

Sealed tubes containing Ikon®-12 and the POE lubricant in combination individually with buna-n, neoprene, 6/6 nylon, Teflon, XH-7 desiccant, cast iron, and brass were maintained at a temperature of 125°C for 14 days. Chemical analysis of the volatiles released from those tubes indicated a small amount of decomposition of the refrigerant. In the presence of the desiccant, there was a very slight formation of monofluoroethene, again indicating a probable degradation of the HFC component of the refrigerant blend.

Data for materials tested with Ikon®-12 in sealed tubes with and without POE lubricant are summarized in Table 1. Average values for each pair of tubes are reported. Hardness values were not obtained for nylon

Table 1. Materials Compatibility Test Results

Material	Weight Change (Percent)	Volume Change (Percent)	Linear Swell (Percent)	Dia. Change (Percent)	Hard. Change (Percent)
Buna-N	+ 37.2	+ 8.4	+ 9.6	+ 4.1	+ 7.3
Buna-N (with POE lube)	+ 20.1	+ 10.0	+ 5.7	+ 4.9	+ 12.9
Neoprene	+12.5	+7.5	+2.8	+2.3	+13.7
Neoprene (with POE lube)	+7.3	+10.2	+2.7	+4.1	+2.1
Teflon	+ 4.8	- 8.9	+ 5.2	- 4.6	not tested
Teflon (with POE lube)	+ 2.7	- 6.4	+ 4.7	- 3.3	not tested
Nylon 6/6	+ 12.9	+ 5.8	+ 1.1	not tested	not tested
Nylon 6/6 (with POE lube)	- 9.4	- 10.1	- 4.4	not tested	not tested

and Teflon since these harder materials were not amenable to the test method. With the POE lubricant, both Teflon and nylon shrank in volume while buna-n swelled in volume. For Teflon and nylon, volume change was excessive compared to generally accepted criteria for volume change from -5 to +25 percent.

It was concluded that the stability and compatibility of the Ikon®-12 refrigerant and POE lubricant combination were adequate to proceed with performance testing at temperatures not exceeding 125°C. Further laboratory and field testing would be required to determine if stability and materials compatibility are acceptable for long-term use.

REFRIGERATION CAPACITY AND ENERGY EFFICIENCY

Refrigeration capacity and COP (coefficient of performance) were experimentally determined for Ikon®-12 and for HFC-134a with an instrumented, automotive air-conditioning system. Instrumentation was similar to that described in a previous paper.[10] The automotive air-conditioning system was an OEM unit designed for HFC-134a. Refrigeration capacity was measured with an uncertainty of approximately ±100W and a repeatability of approximately ±30W. COP was determined from the measured power input to the compressor with an uncertainty of approximately ±75W and a repeatability of ±50W.

The OEM air-conditioning system had a suction line accumulator and orifice tube expansion device. Refrigerant was added to the system until liquid was present in the accumulator as indicated by minimal superheat at the compressor inlet. The refrigerant charge size was 920 g for HFC-134a and 1520 g for Ikon®-12. The same orifice tube was used for both refrigerants. Approximately the same amount of subcooling was measured at the orifice tube inlet for both refrigerants at each test condition. Test results are shown in Figure 1. Refrigeration capacity and COP for Ikon®-12 and HFC-134a were measured at three compressor rotational speeds, three evaporating temperatures, and three condensing temperatures. Over the range of test conditions, refrigeration capacity for Ikon®-12 averaged 5 percent higher than for HFC-134a. COP for Ikon®-12 averaged 11 percent higher than for HFC-134a. Compressor discharge temperatures with Ikon®-12 averaged 2°C higher than with HFC-134a. Over the range of test conditions, the compressor discharge temperature with Ikon®-12 varied from 0.8°C lower to 6°C higher than with HFC-134a.

DURABILITY

Following the performance tests, the test system was run for an extended period of time with an elevated compressor discharge temperature to determine if refrigerant decomposition would occur. Test conditions were adjusted to obtain a compressor discharge temperature of 110°C. Compressor suction pressure was 338 kPa and discharge pressure was 2614 kPa. Compressor rotational speed was 2050 rpm, and compressor ambient temperature was 70°C. After 192 hours of continuous operation at those conditions, mechanical failure of a compressor suction reed valve occurred. The Ikon®-12 refrigerant was recovered from the system, and the compressor was disassembled and inspected. The broken reed valve damaged the cylinder surface, but there was no evidence of excessive wear or any other problem. Failure of the reed valve was apparently caused by metal fatigue due to the severe operating conditions. The lubricant remained clear, and no deposits were found on the internal surfaces of the compressor.

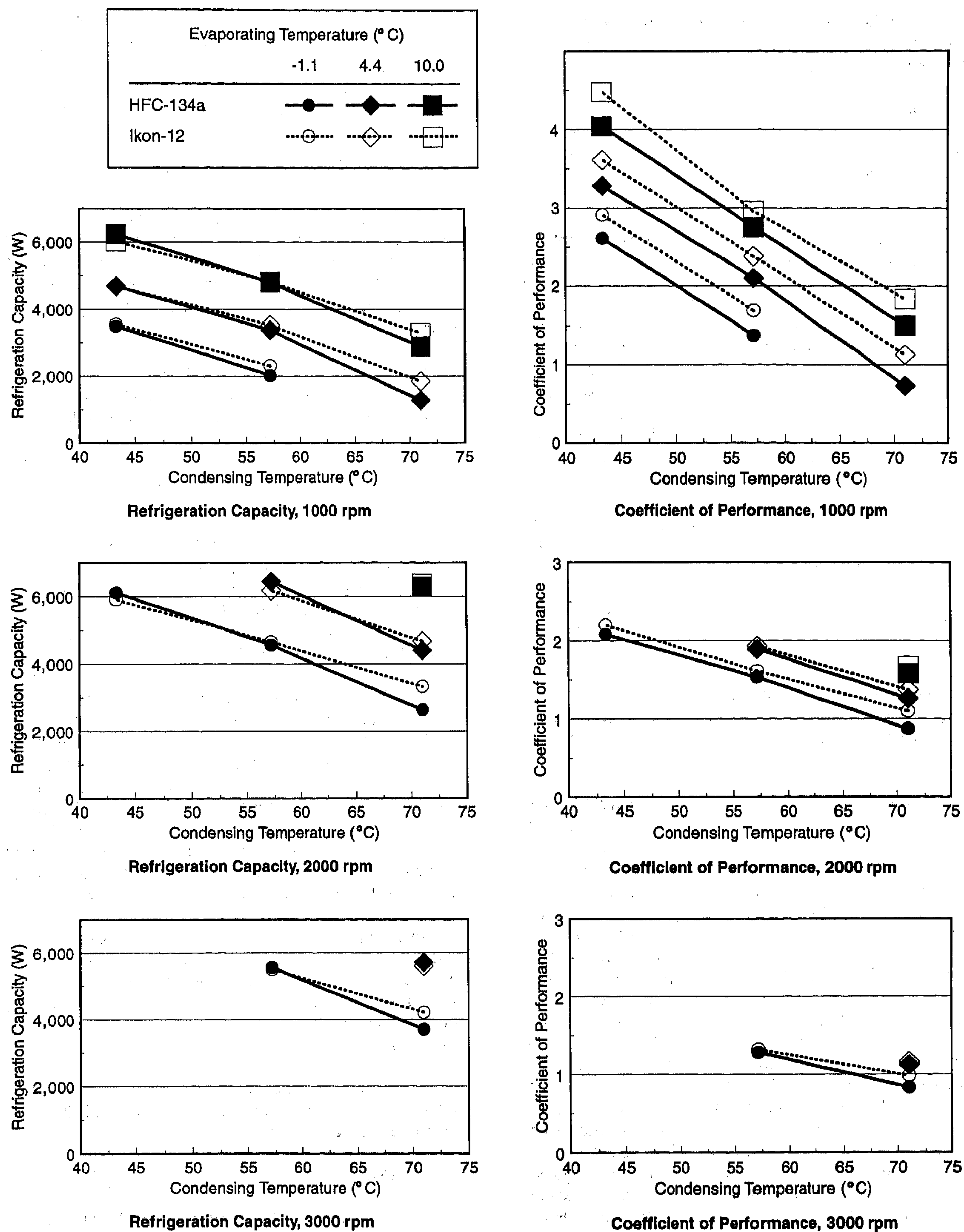


Figure 1. Refrigeration Capacity and Coefficient of Performance of Ikon-12 Compared to HFC-134a

Following the compressor failure, samples of refrigerant and lubricant were analyzed. Vapor- and liquid-phase materials were analyzed with gas chromatography, mass spectrometry, and infrared spectrometry. The resultant chromatograms and spectra were compared to those obtained for pure refrigerant and lubricant. There was no evidence of refrigerant breakdown in the vapor or liquid. Purity of the Ikon®-12 was greater than 99.5 percent.

The failed compressor was replaced with a similar, new compressor. Before installation, PAG lubricant was flushed from the new compressor with POE lubricant, and approximately 40 ml of the new POE lubricant was added to the new compressor to replace the amount removed with the failed compressor. The total amount of lubricant in the air-conditioning system was approximately 240 ml. The same Ikon®-12 refrigerant used for the previous testing was recharged into the system. Test conditions were adjusted to maintain the 110°C compressor discharge temperature, but the compressor rotational speed was reduced to prevent premature failure of the valves. Compressor rotational speed was 1000 rpm, and compressor ambient temperature was 90°C. Compressor suction pressure was 479 kPa and discharge pressure was 2867 kPa.

Refrigerant and lubricant samples were analyzed after 467 hours and again after 632 hours of operation, including the 192 hours of operation on the first compressor. Gas chromatography/mass spectrometry analyses of the refrigerant vapor and gas chromatography/Fourier transform infrared analyses of the lubricant showed no significant changes over the 632 hour period.

After 732 hours of operation, a leak in a refrigerant line resulted in the loss of the refrigerant charge. A lubricant sample was analyzed with gas chromatography/Fourier transform infrared, and no significant change was found. During inspection of the test system, it was found that the compressor shaft could not be rotated by hand. When torque was applied to the compressor shaft with a wrench, the moving parts loosened, and the shaft could then be turned by hand. The compressor was disassembled and inspected. A dark, viscous fluid was found in small quantities on some bearing surfaces and in some cavities inside the compressor. No excessive wear was found on any surfaces, and no indication of mechanical failure was observed. Lubricant found inside the compressor remained clear. Since no similar dark substance was found inside the first compressor after 192 hours of operation, and since Ikon®-12 refrigerant was known to be incompatible with PAG lubricant, it was suspected that the foreign substance resulted from a small amount of PAG lubricant remaining in the second compressor after flushing. An infrared spectral analysis of the dark, viscous fluid exhibited general features resembling both PAG and POE lubricants. However, the fluid could not be unambiguously traced to residual PAG lubricant left in the compressor. Further durability testing and chemical analyses were beyond the scope of the project. Further testing would be required to determine if stability and durability are acceptable for long-term use.

SUMMARY

1. A proprietary, near-azeotropic, low-GWP, refrigerant blend of an FIC and an HFC, called Ikon®-12, was evaluated as an alternative to HFC-134a for automotive air conditioning. The evaluation was motivated by concern over the relatively high global warming potential of HFC-134a.

2. Ikon®-12 was not compatible with a PAG lubricant, but was sufficiently compatible and stable with a POE lubricant and engineering materials to proceed with testing in an automotive air-conditioning system. Over the temperature range of -35 to +100°C, the Ikon®-12 refrigerant was completely miscible with the POE lubricant.

3. Refrigeration capacity and COP were determined experimentally with an instrumented, OEM, automotive air-conditioning system. Refrigeration capacity for Ikon®-12 averaged 5 percent higher than for HFC-134a over the range of test conditions. COP for Ikon®-12 averaged 11 percent higher than for HFC-134a.

4. Durability of the refrigerant and lubricant was tested by operating the test system with an elevated compressor discharge temperature. During the course of the testing, a compressor suction reed valve was broken. Failure of the reed was apparently caused by metal fatigue due to the severe operating conditions. After 632 hours of operation with a discharge temperature of 110°C, no significant breakdown of the refrigerant or lubricant was detected from chemical analyses by gas chromatography, mass spectroscopy, and infrared spectroscopy.

5. Following the durability test, the compressor was disassembled and inspected. A small quantity of dark, viscous fluid was found inside the compressor and may have been caused by residual PAG lubricant. However, the fluid could not be positively identified by chemical analyses.

6. Further laboratory and field testing would be required to determine if stability and materials compatibility are acceptable for long-term use.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of Jon Nimitz of The Ikon Corporation, who provided information on the Ikon®-12 refrigerant; and of Georgi L.

Kazachki of Acurex Environmental Corporation, who provided technical consultation for performance testing.

References

1. *Radiative Forcing of Climate Change, the 1994 Report of the Scientific Assessment Working Group of IPCC.* Intergovernmental Panel on Climate Change. 1994. Page 28.

2. *1994 Report of the Refrigeration, Air Conditioning and Heat Pumps Technical Options Committee for the 1995 Assessment of the Montreal Protocol on Substances that Deplete the Ozone Layer.* United Nations Environment Programme. 1994. Pages 163-183.

3. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-1994.* U.S. EPA Office of Policy, Planning and Evaluation. Washington, D.C. EPA-230-R-96-006 (NTIS PB96-175997). 1995. Page 4.

4. Jetter, J.J., Smith, N.D., Ratanaphruks, K., Ng, A.S., Tufts, M.W., Delafield, F.R. "Evaluation of Alternatives for HFC-134a Refrigerant in Motor Vehicle Air Conditioning." Proceedings of the 1996 International Conference on Ozone Protection Technologies (accepted). Washington, D.C. 1996.

5. *Significant New Alternatives Policy (SNAP) Program Submission to the U.S. Environmental Protection Agency, Ikon®-12 Refrigerant, Docket Copy, For Public Release.* Submitted March 6, 1995 by Jon Nimitz, The Ikon Corporation.

6. *Standard Test Method for Concentration Limits of Flammability of Chemicals, ASTM Designation: E681-85 (Reapproved 1991).* American Society for Testing and Materials. Philadelphia. 1991.

7. *Determining Limits of Flammability of Chemicals at Elevated Temperatures and Pressures, ASTM Designation: E918-83 (Reapproved 1988).* American Society for Testing and Materials. Philadelphia. 1988.

8. "Protection of Stratospheric Ozone: Listing of Substitutes for Ozone-Depleting Substances." *Federal Register.* Volume 61, Number 100. May 22, 1996. Pages 25585-25594.

9. *ANSI/ASHRAE Standard 97-1989, Sealed Glass Tube Method to Test the Chemical Stability of Material For Use Within Refrigerant Systems.* American Society of Heating, Refrigerating and Air-Conditioning Engineers. Atlanta, Georgia. 1989.

10. Jetter, J.J., Delafield, F.R. "Retrofitting an Automotive Air Conditioner with HFC-134a, Additive, and Mineral Oil." *Proceedings of the 1994 International CFC and Halon Alternatives Conference.* Washington, D.C. 1994. Pages 785-794.

970526

Total Environmental Warming Impact (TEWI) Calculations for Alternative Automotive Air-Conditioning Systems

Steven K. Fischer and James R. Sand
Oak Ridge National Lab.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The Montreal Protocol phase-out of chlorofluorocarbons (CFCs) has led manufacturers to develop refrigeration and air-conditioning systems that use refrigerants that do not damage stratospheric ozone. Most refrigeration industries have adapted their designs to use hydrochlorofluorocarbon (HCFC) or hydrofluorocarbon (HFC) refrigerants; new automobile air-conditioning systems use HFC-134a. These industries are now being affected by scientific investigations of greenhouse warming and questions about the effects of refrigerants on global warming. Automobile air-conditioning has three separate impacts on global warming; 1) the effects of refrigerant inadvertently released to the atmosphere from accidents, servicing, and leakage; 2) the efficiency of the cooling equipment (due to the emission of CO_2 from burning fuel to power the system); and 3) the emission of CO_2 from burning fuel to transport the system. The Total Equivalent Warming Impact (TEWI) is an index that should be used to compare the global warming effects of alternative air-conditioning systems because it includes these contributions from the refrigerant, cooling efficiency, and weight.

This paper compares the TEWI of current air-conditioning systems using HFC-134a with that of transcritical vapor compression system using carbon dioxide and systems using flammable refrigerants with secondary heat transfer loops. Results are found to depend on both climate and projected efficiency of CO_2 and flammable refrigerant systems. Performance data on manufacturing prototype systems are needed to verify the potential reductions in TEWI. Extensive field testing is also required to determine the performance, reliability, and "serviceability" of each alternative to HFC-134a to establish whether the potential reduction of TEWI can be achieved in a viable consumer product.

INTRODUCTION

The Montreal Protocol phase-out of chlorofluorocarbons (CFCs) has led manufacturers to develop refrigeration and air-conditioning systems that use refrigerants that do not damage stratospheric ozone. The automobile industry responded to these requirements by developing air-conditioning equipment based on HFC-134a that provided comparable or better performance than the preceding generation of air conditioners using CFC-12. Worldwide concerns are now being raised about the impact of using fluorocarbon refrigerants if they are released to the atmosphere because of their global warming potentials (GWPs). When considering the global warming impact of any refrigeration system, however, it is essential to remember that factors other than the GWP of the refrigerant are important, or even dominant for some refrigeration systems.

Automobile air-conditioning has three separate impacts on global warming; 1) the effects of refrigerant inadvertently released to the atmosphere from accidents, servicing, and leakage; 2) the efficiency of the cooling equipment (due to the emission of CO_2 from burning fuel to power the system); and 3) the emission of CO_2 from burning fuel to transport the system. The Total Equivalent Warming Impact (TEWI) is an index that can be used to compare the global warming effects of alternative air-conditioning systems.

ASSUMPTIONS

This paper examines the TEWI of hypothetical automobile air conditioners using HFC-134a, isobutane (HC-600a), propane (HC-290), and carbon dioxide (R-744). The system with HFC-134a is assumed to use the "conventional" components (e.g. belt driven open compressor, direct expansion evaporator, air-cooled condenser) with "representative" component efficiencies. The two air conditioners using hydrocarbons as refrigerants are assumed to be similar to the HFC-134a system with the notable exception of employing a secondary heat transfer loop to isolate the flammable refrigerant from

the passenger compartment. This addition results in a thermodynamic loss because of the additional ΔT imposed by the secondary loop, extra weight (due to the heat exchanger, hoses, and pump), and pumping energy. The air conditioner using CO_2 has been frequently in the literature [1,2,3] and resembles a conventional vapor compression system with the notable exception that the high side heat exchanger operates above the critical point and hence is a gas cooler; there is no condensation from vapor to liquid in the high side gas cooler (above the critical temperature, 31 °C).

Previous papers have compared TEWI of alternative technologies by computing energy use from system COP at a single design condition and a fixed number of equivalent full load hours [4,5]. This simplified approach has been criticized as being unrealistic because of the broad range of variables affecting actual system performance and energy use. The analysis described in this paper takes a first step toward expanding upon previous analysis to incorporate information on vehicle speed and regional climate, though there are still simplifying assumptions that could be improved.

Several basic assumptions affect the calculation of TEWI irrespective of system configuration or vehicle air-conditioning load. These include:

- 57 liters gasoline / 100 kg incremental weight / 10,000 km,
- 2.32 kg CO_2 released / liter fuel consumed,
- 0.243 kg CO_2 released / kWh heat energy into the engine,
- 25% engine efficiency, and
- 85% of vehicle operating time at engine speeds above 16.1 km/h (10 mph), 15% of time at idle conditions

Other assumptions are independent of the climate, but vary with the choice of refrigerant. They include the evaporating temperature, refrigerant GWP, system weight, and the high side refrigerant temperature. These are summarized in Table 1. The analysis performed in this paper assumes that there is a temperature difference, the approach ΔT, between the ambient air temperature and the high-side refrigerant temperature. Two sets of values have been chosen that depend only on compressor shaft speed; two ΔT's are used for the subcritical systems (HFC-134a, HC-600a, HC-290) and two smaller ΔT's for the transcritical system [2].

Other parameters, such as the design capacity and driving distance, vary between Europe and the U.S. These are summarized in Table 2. Typical meteorological binned weather data has been used for cities in four different geographic regions in the U.S. [6] and typical meteorological year binned data were averaged for several cities in each of five European countries for use in the energy input calculations. These data are shown in Figures 1 and 2.

Siewert published a distribution for vehicle speed during 50,000 miles (80,500 km) of use [7]. These data indicate 15% of vehicle use is at or below 16 kph (10 mph) and that 85% is at higher speeds; consequently it is assumed that 15% of compressor operation is at a low shaft speed and 85% is at a high speed. Siewert's data are also used to estimate the number of hours of vehicle use per year. These results are summarized in Table 3.

Figure 3 shows the dependence of air-conditioner operation on average daily temperature [8]. A compromise in accuracy was made by applying this correlation to binned hourly weather data instead of average daily temperature; future work should employ a correlation of on-time with ambient temperature instead of average daily temperature.

System COPs were computed for each of the refrigerants using a commercial software package designed for solving thermodynamic problems [9]. The subcritical systems were evaluated using calculated refrigerant properties at:

- the compressor inlet,
- the compressor exit,
- the condenser exit
- the evaporator inlet, and
- saturated vapor leaving the evaporator.

The analysis of the supercritical CO_2 system includes the use of intermediate heat exchange between high and low-side, although this does not provide useful cooling, and pressures that give the maximum COP. The calculated COPs are shown in Figs. 4 and 5.

METHODOLOGY

The global warming effects of the energy to operate and transport the air conditioner and from emissions of refrigerant are calculated separately. Energy input is computed from annual cooling provided divided by the temperature dependent COPs. The number of annual vehicle operating hours

in Table 3 are combined with the temperature distributions in Figs. 2 and 3 to estimate hours of vehicle operation as functions of the ambient temperature (the temperature distributions are first converted from hours/year at each temperature to percentage of year). The derived distribution of hours of vehicle operation as a function of temperature are then combined with the compressor on-time in Fig. 3 and design capacity to calculate air-conditioner cooling output as functions of ambient temperature. Highway speed and idle condition energy inputs are computed at each temperature by dividing the cooling output by the corresponding COP and summed over the year. The average annual energy input is computed as 85% of the highway speed a/c energy use and 15% of the idle condition energy use. Air conditioner energy input, in kWh, is converted to lifetime CO_2 emissions using the assumed engine efficiency (25%), 0.243 kg CO_2 /kWh input, and the assumed lifetime.

Fuel consumption for transporting the weight of the air conditioning system is computed using 57 liters/100 kg/10,000 km for incremental fuel use for weight increases. This value is combined with the assumed vehicle use for the U.S. and Europe and the air-conditioner lifetime. Lifetime fuel use for the weight of the air conditioner is converted to CO_2 emissions using 2.32 kg CO_2/liter of gasoline.

Refrigerant emissions are based on the assumed lifetimes of the air conditioners and average refrigerant losses of 55 grams/year [10,11] and the assumption that 95% of the refrigerant charge is recovered after the useful lifetime of the system. No practical experience is available for leakage from either the hydrocarbon or carbon dioxide air conditioners, but the effect on TEWI is insignificant because of the extremely low GWPs of these gases.

RESULTS

The results of the analyses are shown in Figures 6 and 7 for the four regions of the U.S. and the five European countries. There is a separate bar for each of the refrigerants in each of the countries and regions; each bar has a segment for CO_2 emissions resulting from power input to the compressor, belt/clutch assembly, and fan; a segment corresponding to fuel consumption for transporting the weight of the air conditioner assembly, and a segment for the direct global warming effect of refrigerant emissions. Each bar for the transcritical CO_2 system (R-744) also contains a segment resulting from unknowns in the operating efficiency. The first segment for energy use for CO_2 is determined using small approach temperatures for the gas cooler (i.e. 17°C idle and 3°C highway) which is believed achievable by at least one manufacturer. The second segment is additional energy required using the same approach temperatures assumed for the subcritical systems. The segments of each bar due to weight are nearly the same for all of the systems and direct effects of refrigerant emissions are only visible on the bars for HFC-134a. The direct effects are so small for the hydrocarbon refrigerants and transcritical system that they do not show up on these drawings.

In regions of the U.S. with high cooling demands, the hydrocarbon systems have TEWI comparable to HFC-134a systems. Recent information about prototype testing indicates that the transcritical system could have steady state COPs similar to or better than the current COPs for HFC-134a [12, 13]. If this is the case, the transcritical system would have a TEWI lower than any of the other systems; although efficiency of systems using HFC-134a can be improved if it is optimized to reduce energy use.

CONCLUSION

Several conclusions can be drawn from the results in Figs. 6 and 7. First, there may or may not be a lower TEWI using a flammable refrigerant depending on the system size and annual cooling load. There is no significant reduction in TEWI for the hydrocarbon refrigerants in the U.S. The CO_2 system has a comparable to higher TEWI than HFC-134a in the U.S. under the current assumptions for COPs and though higher than the hydrocarbons, it is not significantly higher using the low values of the approach temperatures.

References

1. J. Pettersen and G. Skaugen, *"Operation of Trans-critical CO_2 Vapor Compression Circuits in Vehicle Air Conditioning,"* Refrigeration Science and Technology Proceedings: New Applications of Natural Working Fluids in Refrigeration and Air Conditioning, Hannover, Germany, May 10-13, 1994, ISSN 0151-1637, pp. 495-506.
2. J. Pettersen, *"An Efficient New Automobile Air-Conditioning System Based on CO_2 Vapor Compression,"* ASHRAE Transactions (1994) Vol. 100, Pt. 2, pp. 657-665.
3. G. Lorentzen and J. Pettersen, *"A New, Efficient and Environmentally Benign System for Car Air Conditioning,"* International Journal of Refrigeration, Vol 16, No. 1, 1993, pp. 4-12.
4. S. Fischer, P. Hughes, and P. Fairchild, Energy and Global Warming Impacts of CFC Alternative Technologies, AFEAS Report, December 1991.
5. S. Fischer, J. Tomlinson, and P. Hughes, Energy and Global Warming Impacts of Not-In-Kind and Next Generation CFC and HCFC Alternatives, AFEAS Report, 1994.
6. Engineering Weather Data, AFM 88-29, TM 5-785, NAVFAC P-89, July 1978.
7. H. Siewert, *"Automotive Air-Conditioning Compressors,"* ASHRAE Transactions (1983), Vol. 89, Part 1B, pp. 622-629.
8. Personal communication from H. Fernqvist of Volvo Car Corporation to J. Sand of Oak Ridge National Laboratory, May 31, 1996.
9. EES: Engineering Equation Solver: Users Manual, F-Chart Software, Middleton, Wisconsin.
10. Personal communication from K. Hara, Japan Industrial Conference for Ozone Layer Protection, to J. Sand, Oak Ridge National Laboratory, June 17, 1996.
11. Personal communication from J. Köhler, IPEK, to V. Baxter, Oak Ridge National Laboratory, May 29, 1996.
12. Personal communication from H. Sugi, Nippondenso Co. Ltd., to S. Fischer, Oak Ridge National Laboratory, September 19, 1996.

13. Personal communication from Dwayne Taylor, Nippondenso America, to S. Fischer, Oak Ridge National Laboratory, May 16, 1996.

Table 1. Assumed Parameters for Alternative Refrigerants

System Parameter	HFC-134a	HC-600a	HC-290	R-744
Evaporator				
type	direct expansion	secondary loop	secondary loop	direct expansion
temperature	4.4°C	-1.1°C	-1.1°C	4.4°C
secondary fluid pump	N/A	47 W	47 W	N/A
Condenser/Gas Cooler				
approach ΔT (idle)	25°C	25°C	25°C	17°C
approach ΔT (highway)	15°C	15°C	15°C	6°C and 3°C
Compressor				
efficiency (idle)	65%	65%	65%	70%
efficiency (highway)	60%	60%	60%	65%
belt/clutch efficiency	95%	95%	95%	95%
Evaporator Blower				
power	250 W	250 W	250 W	250 W
alternator efficiency	55%	55%	55%	55%
Refrigerant GWP	1300	11	11	1
System Parameters				
capacity	7.0 kW	7.0 kW	7.0 kW	7.0 kW
weight	13.6 kg	16.3 kg	16.3 kg	13.6 kg

Table 2. Assumed Geographic Parameters

System Parameter	U.S.	Europe
Capacity	7.0 kW	7.0 kW
Charge (HFC-134a)	1100 g	700 g
Emissions (g HFC-134a/year)	55 g	55 g
Equipment Lifetime	11 years	10 years
Vehicle Use		
km / year	16,580	13,200
hours / year	249	198
Weather Data	Southeast (Miami) Northeast (Long Island) Midwest (Chicago) Southwest (Phoenix)	United Kingdom Germany Greece Italy Spain

Table 3. Vehicle Usage in the U.S. and Europe

Vehicle Speed		Vehicle Use[1]		Pro-Rated Annual U.S. Usage		Pro-Rated Annual European Usage		Fraction of Total
		miles	hours	kilometers	hours	kilometers	hours	
Idle	Idle	0	150	0	31	0	25	12.4%
10 mph	16 kph	350	35	116	7	70	6	2.9%
20 mph	32 kph	3,500	175	1,160	36	720	29	14.5%
40 mph	64 kph	14,000	340	4,640	70	2,880	56	28.1%
60 mph	97 kph	25,000	420	8,290	87	5,150	69	34.7%
80 mph	129 kph	7,150	90	2,370	19	1,480	15	7.4%
Total		50,000	1,210	16,580	249	13,200	200	100%

---

[1]H. G. Siewert, *"Automotive Air-Conditioning Compressors,"* ASHRAE Transactions, 1983, Part 1B.

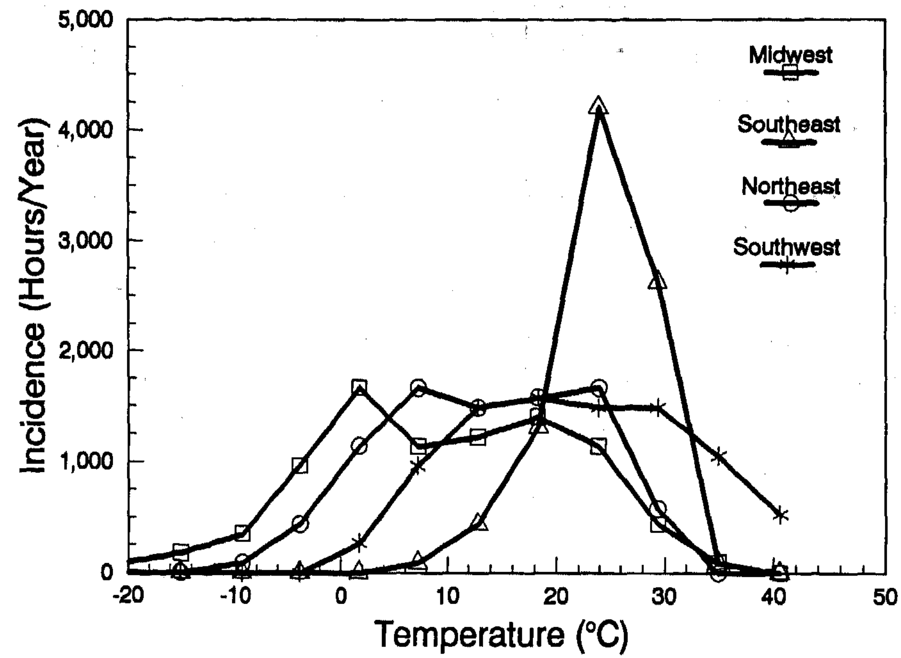


Fig. 1. Ambient temperature distribution for the U.S.

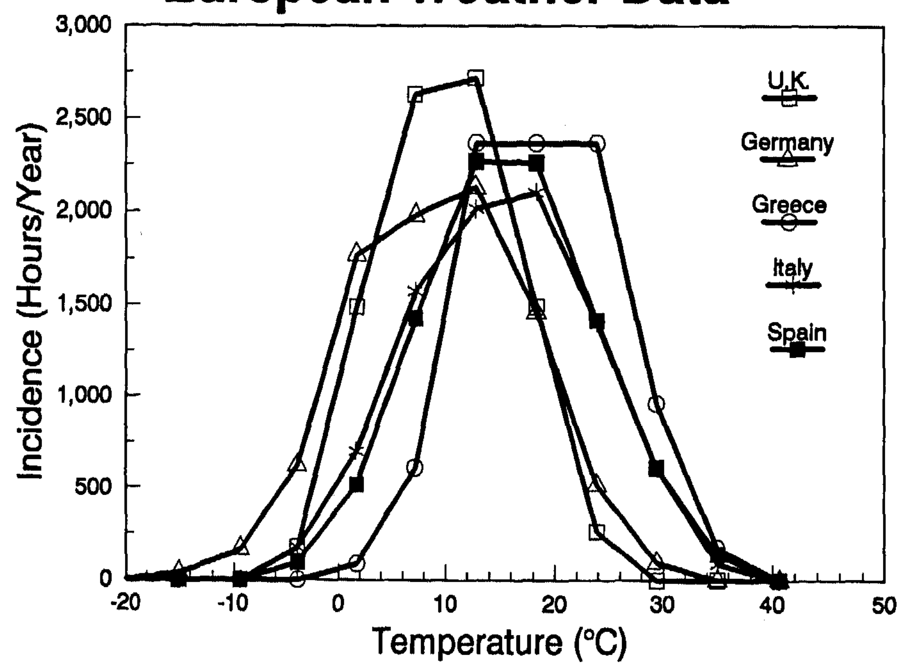


Fig. 2. Ambient temperature distribution for Europe.

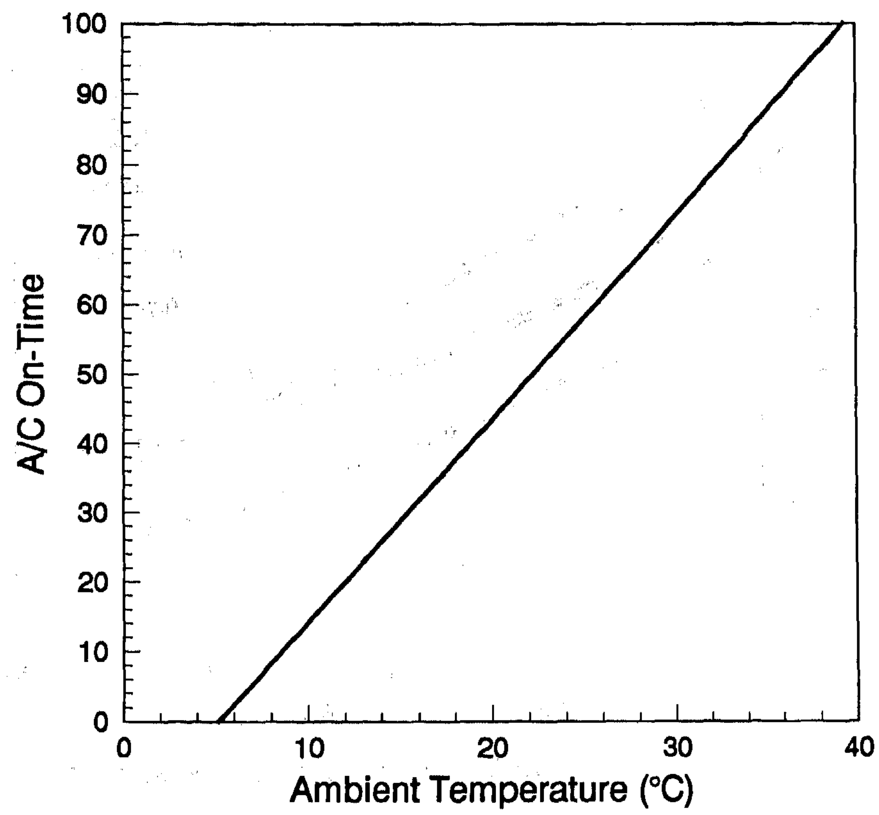


Fig. 3. Air conditioner run time and daily average temperature.

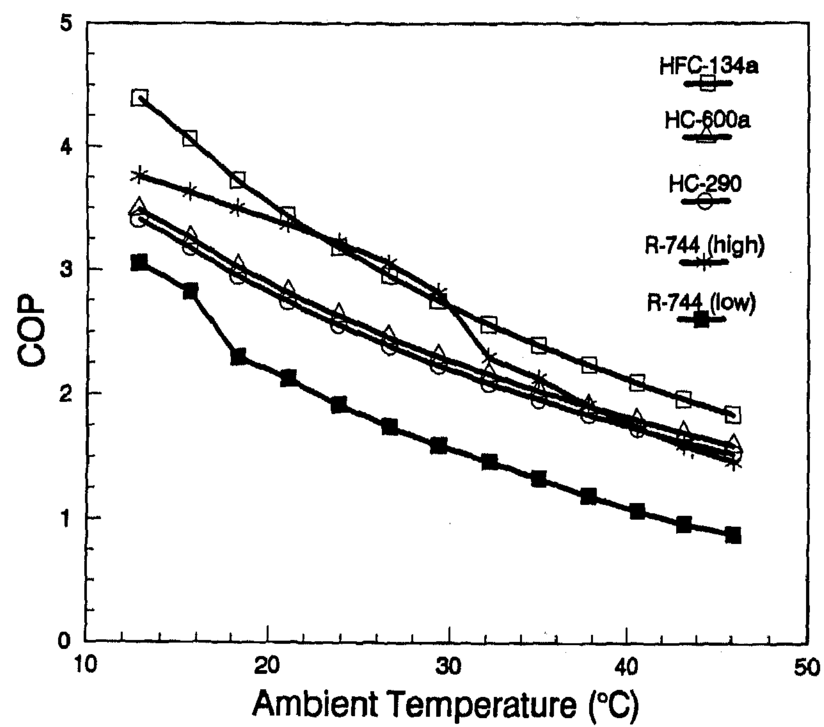


Fig. 4. System COP at highway driving speeds.

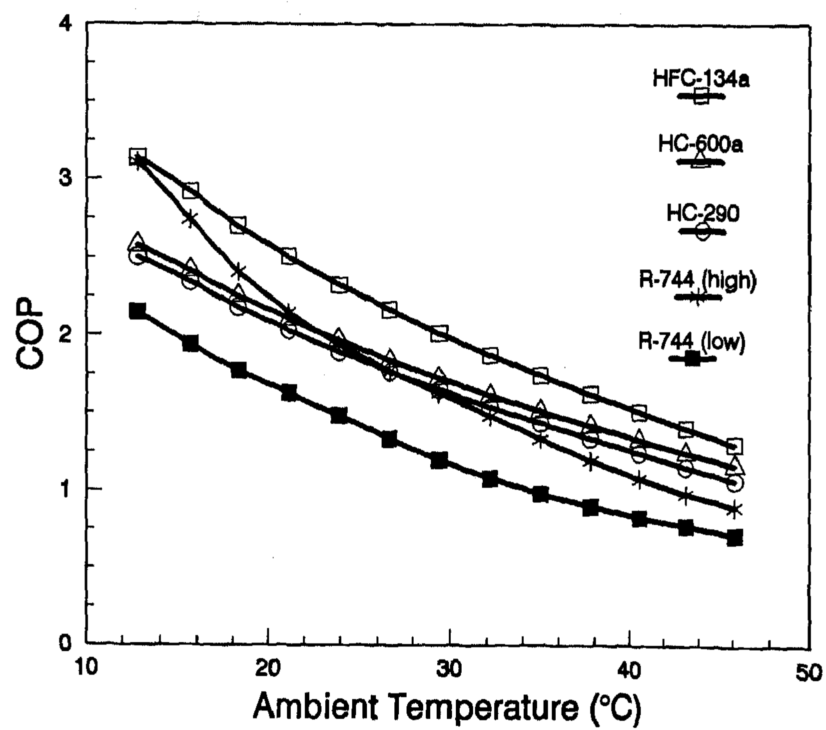


Fig. 5. System COP at idle conditions.

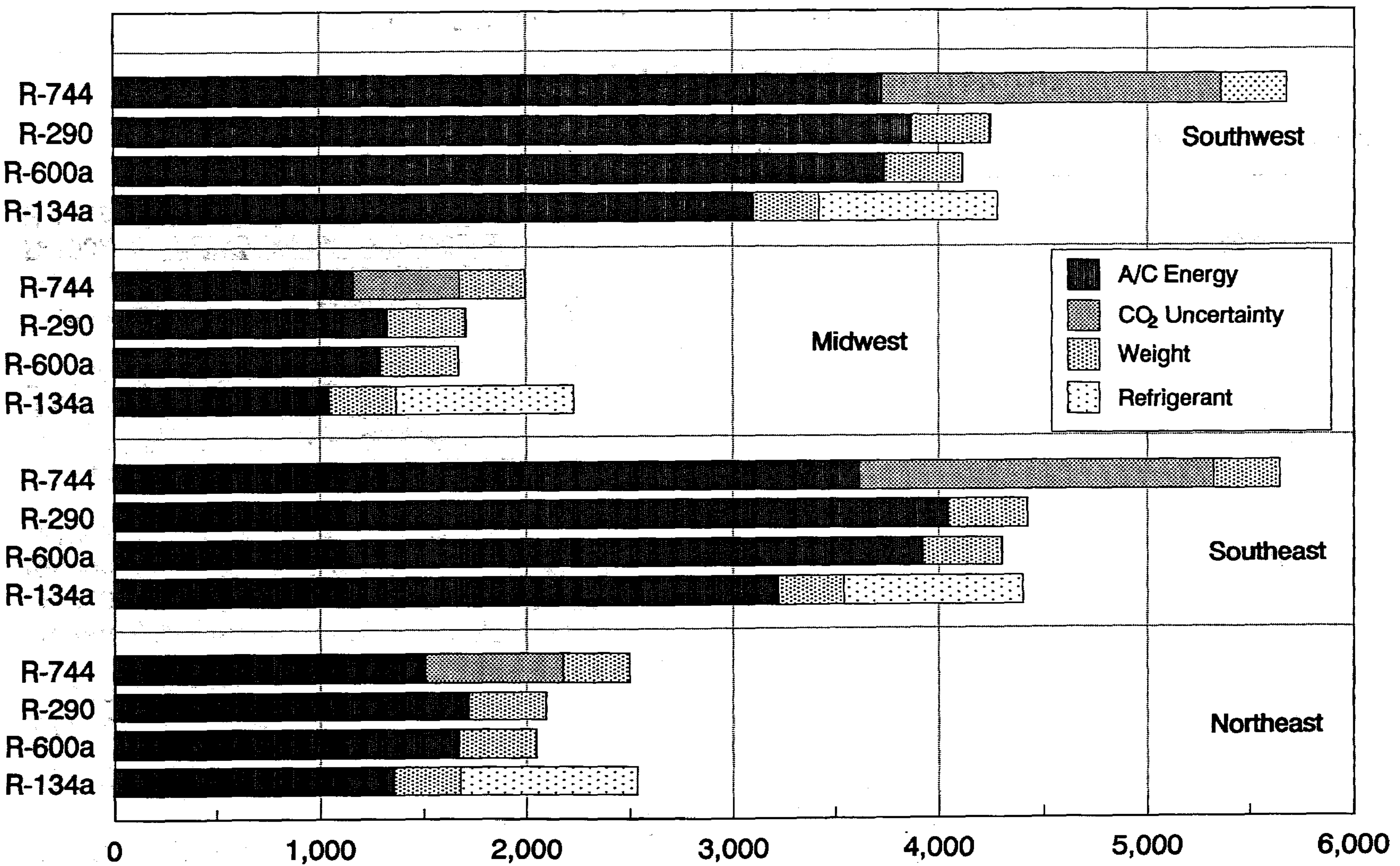


Fig. 6. TEWI of automobile air-conditioning systems in the U.S.

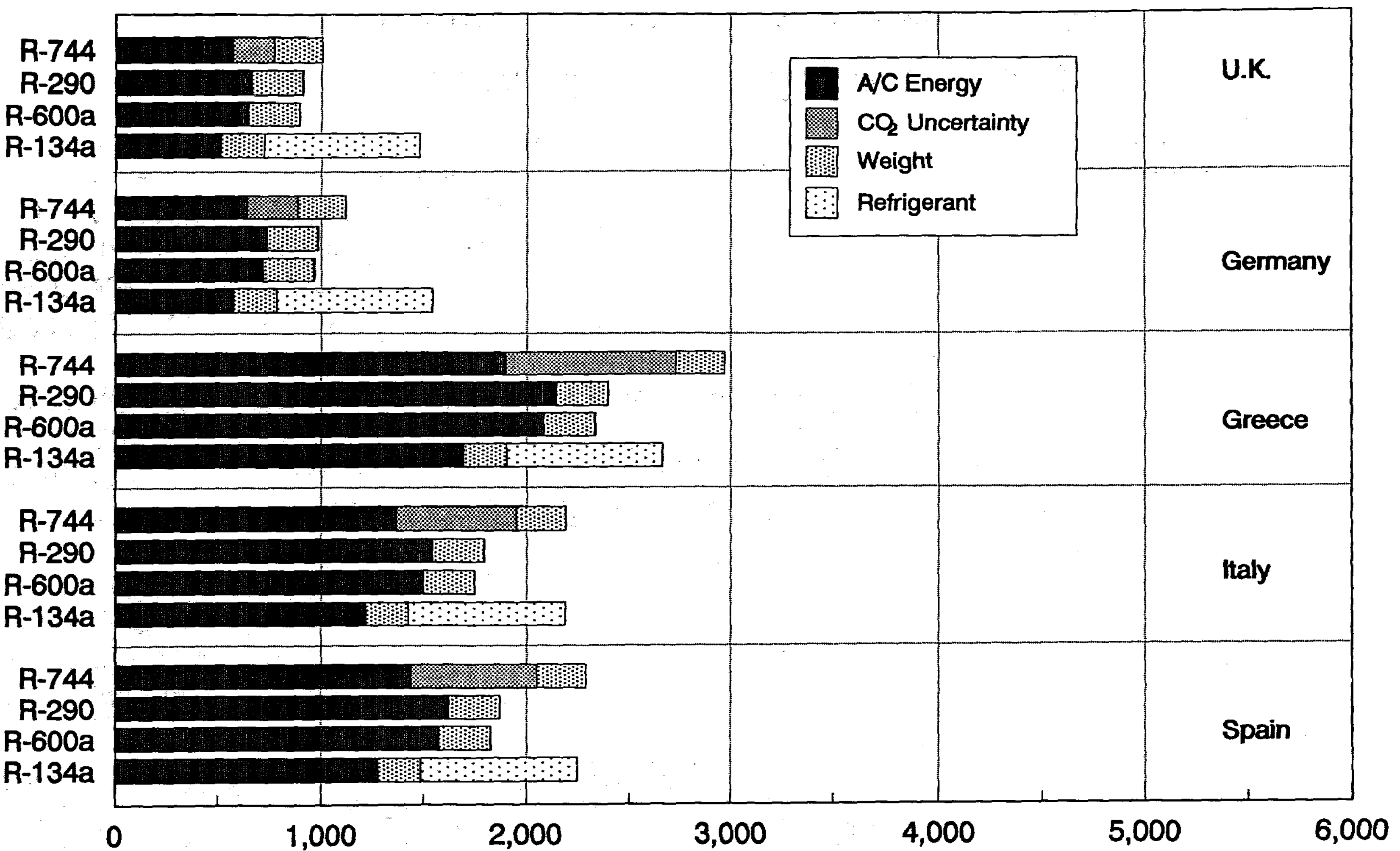


Fig. 7. TEWI of automobile air-conditioning systems in Europe.

970527

A Critical Look at R-744 and R-134a Mobile Air Conditioning Systems

M. S. Bhatti
General Motors Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

There is intense research activity to appraise the merits of the carbon dioxide (R-744) mobile air conditioning system due to its perceived amelioratory effect on the total global warming impact which comprises two components: direct global warming due to refrigerant leakage into the atmosphere and indirect global warming due to power consumption by the system. While the direct global warming impact of R-744 is negligible compared to that of R-134a, the indirect global warming impact of the R-744 system is intrinsically higher than that of the R-134a system. In order to quantify the indirect global warming impact of the R-744 system, an accurate assessment of its coefficient of performance (COP) *vis-a-vis* COP of the present baseline R-134a system is necessary. Presented in this paper is the critical assessment of the R-744 system, including its COP under realistic operating conditions, heat exchanger and compressor characteristics, total global warming impact, and historical evidence that suggests that the proposal to replace R-134a by R-744 is untenable in view of lower energy efficiency and 5 to 12 times higher operating pressure for the R-744 system.

INTRODUCTION

The specter of ozone depletion and global warming has revivified carbon dioxide which at one time (1890 to 1930) was the preferred refrigerant for vapor compression machines despite its unfavorable thermodynamic characteristics (high critical pressure and *ipso facto* high operating pressure) [1-30]. The main appeal of carbon dioxide at the time was that its toxicity and flammability characteristics were superior to those of other available refrigerants like sulfur dioxide, ammonia, methyl chloride, and butane. However, it could not compete with the chlorofluorocarbon refrigerants which emerged in 1930 [31]. The rapid demise of the carbon dioxide systems thereafter is attributable to their lower energy efficiency and higher operating pressure compared to those of the chlorofluorocarbon systems. In this paper, we will review the historic evidence that suggests that the *locus standi* of the R-744 system against the present R-134a system is rather tenuous for the same reasons which led to its phase out starting back in the 1930s.

Up until the 1970s, environmental concerns like ozone depletion and global warming were not even considered in the context of anthropogenic refrigerants. However, the picture changed in 1974 when two scientists named Molina and Rowland revealed the phenomenon of destruction of the stratospheric ozone layer that shields the Earth from the harmful effects of the sun's ultraviolet rays. They theorized that the destruction of the ozone layer is catalyzed by molecules containing chlorine that are propelled to the stratosphere by extremely stable chlorofluorocarbon compounds with inordinately long atmospheric lives of the order of 100 years [32]. In 1995, Molina and Rowland along with a third scientist named Crutzen received the Nobel Prize for their work on the stratospheric ozone depletion [33-35].

As a direct result of the work by Molina *et al.*, production and use of the chlorofluorocarbon refrigerants, including R-12, was banned as mandated by the Montreal Protocol in 1981

[36]. After considerable research between 1986 and 1992, R-134a emerged as the preferred replacement for R-12 in mobile air conditioning systems and beginning in 1992 automakers around the world started replacing R-12 with R-134a. An indirect consequence of the work by Molina *et al.* has been that it heightened our awareness of other environmental concerns such as acid rain and global warming particularly the latter since the strong ozone depleting gases are also generally strong greenhouse gases responsible for global warming.

Recent interest in carbon dioxide as a replacement for R-134a stems from the fact that although the ozone depletion potential of R-134a is zero, its global warming potential is not zero. Carbon dioxide, on the other hand, has zero ozone depletion potential coupled with practically zero global warming potential. Based on the favorable global warming potential index alone, it cannot be asserted *a priori* that the global warming impact of the R-744 system will be less than that of the R-134a system. This is predicated on the realization that there are two components of the global warming impact equation, *viz.*, direct global warming due to refrigerant leakage into the atmosphere and indirect global warming due to power consumption by the system incorporating the refrigerant. The issue of the total global warming impact of the R-744 and R-134a systems will be explored at length in this paper.

As a prelude to assessment of the two systems, the comparative characteristics of the two refrigerants (*vide* Table 1) are reviewed as they shed considerable light on the relative performance of the systems. Following these introductory remarks, the subject matter is covered under four distinct captions: (1) Historical Background, (2) System Performance, (3) Heat Exchanger and Compressor Characteristics, and (4) Global Warming Impact. Substantiated conclusions of the analysis are then clearly stated at end of the paper.

HISTORICAL BACKGROUND

There is a long history of the use of carbon dioxide in the refrigeration and air conditioning industry. It was first proposed as a refrigerant for the vapor compression machines by an American named Alexander Catlin Twining (1801-1884) of New Haven, Connecticut in 1850 [37]. The vapor compression system itself was proposed by another American named Oliver Evans (1755-1819) of Philadelphia, Pennsylvania in the postscript to a treatise bearing the curious title: *The Abortion of the Young Steam Engineer's Guide* published in 1805 [38]. Nearly three decades after its original proposal, Jacob Perkins (1766-1849), a friend of Oliver Evans, patented the vapor compression system obtaining the British Patent No. 6662 on August 14, 1834 [39].

Although there are earlier patents proposing the use of carbon dioxide in vapor compression machines [40], the first use of carbon dioxide in a commercially successful machine is attributed to an American named Thaddeus Sobeski Carlincourt Lowe (1832-1913) of Jefferson, New Hampshire whose fame rests on the construction of the military balloons with carbon dioxide as inflation gas. In 1867, he obtained the British Patent No. 952 for an ice machine with carbon dioxide as refrigerant [41]. He made commercial ice with his machines in Dallas, Texas and in Jackson, Mississippi in 1866. He also constructed a machine on board a ship for transport of frozen meat across the Gulf of Mexico [37].

Karl von Linde (1842-1934) of Germany also experimented with carbon dioxide when he designed a machine for F. Krupp at Essen, Germany in 1882 [42]. In 1884, Wilhelm Raydt of Hanover, Germany received the British Patent No. 15,475 for a carbon dioxide ice--making machine [43]. Also, in 1884, James Harrison (1816-1893) of Australia took out the British Patent No. 1890 on the improvements in the production and use of carbon dioxide for refrigeration and air conditioning [44].

The use of carbon dioxide did not advance significantly until Franz Windhausen (1829-1904) of Brunswick, Germany designed a carbon dioxide compressor receiving the British Patent No. 2864 in 1886 [45]. The following year, this patent was purchased by J&E Hall of Great Britain who improved it further and commenced manufacture of the compressor in 1890 [46]. The British ship *Highland Chief* was the first to be outfitted with Hall's carbon dioxide vapor compression machine in 1890 and thereafter these machines found widespread use on the refrigerated marine cargo ships. Of 356 ships worldwide in 1900, 37% used air cycle machines, 37% ammonia absorption machines, and 25% carbon dioxide vapor compression machines. By 1930, 80% of the ships were outfitted with the carbon dioxide machines and the remaining 20% with the ammonia machines [47]. The last of the ships with carbon dioxide machine went out of service in 1950 due to

Table 1. Comparative Characteristics of R-744 and R-134a Refrigerants

Characteristic	R-744	R-134a	Remarks
Chemical formula	CO_2	$C_2H_2F_4$	Unlike R-134a, R-744 occurs in nature.
Molecular weight, lb_m/lb mole	44	102	Being colligative in nature, the refrigerant charge for R-744 system is 44/102 = 0.43 times that for R-134a system.
Triple point temperature, ^oF	- 69.8	- 153.9	It is the lowest temperature at which the liquid phase exists.
Triple point pressure, psia	75.12	0.06	It is the lowest pressure at which the liquid phase exists.
Molar latent heat of evaporation at triple point, Btu/lb mole	6,553	11,549	The higher the molar latent heat of evaporation the higher is the cooling capacity.
Normal sublimation or boiling point, ^oF	- 109.3	- 15.7	R-744 sublimes at atmospheric pressure. Hence it has no normal boiling point.
Vapor pressure at 32 ^oF	505	43	This suggests that operating pressure of R-744 system is 505/43 = 11.7 times that of R-134a system.
Molar latent heat of evaporation at 32 ^oF, Btu/lb mole	4,384	8,648	This suggests that COP of R-744 system is half (4,384/8,648) COP of R-134a system for fixed compressor power.
Critical temperature, ^oF	87.9	214.0	The higher the critical temperature the higher the cooling capacity generally.
Critical pressure, psia	1,070	590	The higher the critical pressure the higher the operating pressure generally.
Global warming potential (GWP)	1	1,300	It is a measure of the potency of a gas to absorb infrared rays responsible for global warming.
Ozone depletion potential (ODP)	0	0	It is a measure of the potency of a gas to destroy the ozone layer that shields the Earth from harmful effects of the sun's ultraviolet rays.

emergence of the chlorofluorocarbon refrigerants (*vide infra*).

Between 1890 and 1897, several unsuccessful attempts were made to introduce the carbon dioxide refrigeration machines in the USA. Eventually, the first successful machine was installed in 1897 under the supervision of Fred Whittenmier who was engaged by Grommes & Ulrich of Chicago to install a carbon dioxide machine for cooling the wine cellar located in Marquette building in Chicago. The machine was imported from Germany and was built by Scharrer & Gross of Nurnburg. Its refrigeration capacity was 2 tons which was rather inadequate for the intended purpose [48].

On July 30, 1897, the Kroeschell Bros. Ice

Machine Co. was organized at Chicago with Fred Whittenmier as Manager. The company started manufacturing small carbon dioxide compressors using the patents of Julius Sedlacek of Linz, Austria [49]. In 1918, Fred Whittenmier left the employ of Kroeschell Bros. to form his own company called Whittenmier Machinery Company, Chicago. This company successfully installed a number of carbon dioxide refrigeration and air conditioning systems.

Use of carbon dioxide for comfort cooling lagged by three decades its use for refrigeration. It was not until 1919 that carbon dioxide vapor compression machines found widespread use for comfort cooling applications such as air conditioning of theaters (1919), department stores (1919), churches (1920), offices (1927), residences (1930), and later practically all types of commercial establishments and public buildings [50-51]. While the high pressure carbon dioxide systems were acceptable for such stationary applications, they never received consideration for mobile air conditioning ostensibly due to the containment problems which are greatly accentuated in a mobile system.

Development of the first mobile air conditioning system was prompted by the invention of chlorofluorocarbon refrigerants in 1928 by Midgley *et al.* of the Frigidaire Division of General Motors Corporation [52]. On September 23, 1932 a proposal was prepared by the General Motors Research Staff to develop a self-contained automobile air conditioning system with R-12 refrigerant. However, it was not until the summer of 1933 that the work got underway with Charles F. Kettering offering the use of his 1933 Cadillac V-12 Town Sedan to estimate the cooling capacity of the proposed system. Eventually, a prototype self-contained system was built and installed in the trunk of a 1939 Cadillac. While General Motors Corporation was evaluating its trunk-mounted unit, Packard Motor Car Company made history by exhibiting its trunk-mounted automobile air conditioning system on November 4, 1939 at the Chicago Auto Show [53].

The modern day automobile air conditioning system for underhood installation with R-12 refrigerant and a single front-mounted condenser was developed in 1953 by Harrison Radiator Division (now known as Delphi Harrison Thermal Systems) in collaboration with Pontiac Division of General Motors Corporation [54]. The same year, Airtemp unit of Chrysler Corporation introduced an automobile air conditioning system with R-12 refrigerant and two front-mounted condensers rather than one as in present practice [55].

A literature review from 1930 to 1990, reveals that there are practically no patents or technical papers on the carbon dioxide refrigeration or air conditioning systems as they were rapidly supplanted by the R-12 systems. Carbon dioxide systems simply could not compete with the R-12 systems from an energy efficiency viewpoint and high operating pressure which undermined the system integrity and cooling capacity at high ambient temperatures. Although it ceased to be used as a working fluid of vapor compression machines, carbon dioxide gained wide acceptance, between 1930 and the present time, in solid form called dry ice in a variety of cryogenic applications including production and storage of ice cream and other perishable food products [56]. Dry ice was independently discovered in 1835 by a French physicist named Charles Saint-Ange Thilorier (1797-1852) and a Scottish chemist named Kenneth Kemp (1807-1843) [37].

Recent interest in the carbon dioxide air conditioning system [57] is in large measure due to the efforts launched in 1989 by Gustav Lorentzen and his associates at Norwegian Institute of Technology, Trondheim, Norway with funding from Hydro Aluminium a. s., Oslo, Norway. Motivated by the consideration that carbon dioxide is a natural refrigerant and its direct global warming impact is negligible compared to that of R-134a, they initiated a theoretical and experimental investigation which culminated in the establishment of a BMW-led European consortium to explore merits of the carbon dioxide system in 1994. In 1993, Lorentzen *et al.* also succeeded in securing the U. S. Patent No. 5,245,836 relating to pressure regulation in the carbon dioxide system [58]. The particular pressure regulation technique covered by this patent has been known since 1928 [26].

A few remarks about the history of the relatively new refrigerant R-134a are in order, Its invention falls within the overarching scope of a series of patents secured between 1930 and 1940 by Midgley, Henne, and McNary of General Motors Corporation [52]. In 1936, Henne and Renoll [59] synthesized R-134a for the first time at Ohio State University, Columbus, Ohio apparently to prove the point made in the aforesaid patents that the compounds like R-134a

are not as stable as R-12. In view of this consideration, R-134a emerged as the natural substitute for R-12 after the alarm was sounded in 1974 that chlorofluorocarbon refrigerants were powerful ozone depleters *ibid.* [32]. After considerable research between 1978 and 1992, car makers around the world started replacing R-12 with R-134a in 1992. DuPont and ICI were the first chemical companies to start up the world's commercial plants to manufacture R-134a in the last quarter of 1990 at Corpus Christi, Texas, USA and Runcorn, England, respectively.

In view of the foregoing appraisement of the historical facts, there is a compelling reason to believe *a fortiori* that R-744 is not likely to replace R-134a in mobile air conditioning system since R-134a is similar to R-12 as regarded operating pressure and energy efficiency.

SYSTEM PERFORMANCE

The underlying principle of the R-744 as well as R-134a system is the vapor compression cycle first proposed by Evans in 1805 and sometimes alluded to as the Evans-Perkins vapor compression cycle [38-39]. However, there is a significant difference between the two systems from an operational viewpoint. The operating pressure of the R-744 system is significantly higher than that of the R-134a system. The R-744 system is a supercritical system with its compressor discharge pressure greater than the critical pressure of carbon dioxide (1,070 psia). The R-134a system, on the other hand, is a subcritical system with its compressor discharge pressure lower than the critical pressure of R-134a (590 psia). The relative magnitudes of the compressor discharge pressures in the two systems requires that in the supercritical system the high pressure side heat exchanger (gas cooler) must carry single-phase (vapor) refrigerant while in a subcritical system the high pressure side heat exchanger (condenser) must carry a two-phase (liquid + vapor) refrigerant.

Presented in this section is the complete theoretical performance of the R-744 and R-134a mobile air conditioning systems comprehending air as well as refrigerant side of the systems. The calculations were performed both for realistic as well as idealized air conditioning systems. In order to provide performance results over the entire operating range of the vehicle, the results were generated both under idle (0 MPH vehicle speed) and down-the-road (50 MPH vehicle speed) steady state conditions. Since the air conditioning system load is higher in ventilation rather than in recirculation mode, the calculations were executed for the former mode. The computations were performed with the aid of a special Delphi Thermal computer code that has been extensively validated by comparing its predictions with the wind tunnel measurements performed on the R-12 and R-134a mobile air conditioning systems.

REALISTIC DESIGN CONDITIONS

There appears to be considerable confusion in some quarters as to what constitutes a realistic set of design conditions for a mobile air conditioning system. This has led to misrecokening of the coefficient of performance and compounding of misapprehension as to the power consumption by the R-744 system. With this in mind, stated below are the realistic design conditions.

1. The nominal cooling capacity of the mobile air conditioning system is 2 tons which is equivalent to 400 Btu/min.

2. Under down-the-road conditions, the standard dry bulb temperature of the outside air is 100 °F with 40% relative humidity which corresponds to an absolute humidity of 0.0167 lb_m H_2O / lb_m dry air. Under idle conditions, on the other hand, there is 20 °F rise in the dry bulb temperature with absolute humidity remaining the same. This translates to a relative humidity of 22.4%.

3. The dry bulb temperature of the evaporator-conditioned air is 70 °F under idle and 50 °F under down-the-road conditions.

4. The evaporator air volume flow rate is 250 ft^3/min both under idle and down-the-road conditions. This translates to the air mass flow rate of 16.6 lb_m/min under idle and 17.3 lb_m/min under down-the-road conditions.

5. The condenser or gas cooler air volume flow rate is 600 ft^3/min under idle and 2,000 ft^3/min under down-the-road conditions. This translates to the air mass flow rate of 40 lb_m/min under idle and 138 lb_m/min under down-the-road conditions.

SYSTEM IRREVERSIBILITIES

In a real life system, we have to contend with various system irreversibilities that lower

the overall system efficiency. Based on our experience, the following irreversibilities are deemed appropriate for the systems under consideration:

1. The power consumption by the air conditioning system blower is 250 watts which, with an alternator efficiency of 55%, translates to 26 Btu/min.

2. The power transmission loss of the air conditioning system mechanical drive is 3% of the compressor power.

3. The pressure drop in the R-134a evaporator is 10% of the mean evaporating pressure. Likewise the pressure drop in the R-134a condenser is 5% of the mean condensing refrigerant pressure. For the R-744 system, the pressure drop in the two heat exchangers is negligible as elaborated on pp. 12 *et seq.*

4. The pressure drop in the accumulator used to store excess refrigerant charge is 10% of the evaporator out pressure.

5. The isentropic efficiency of the R-134a system compressor is 65% under idle and 60% under down-the road conditions. The isentropic efficiency of the R-744 system is 70% under idle and 65% under down-the road conditions.

6. With orifice tube as the expansion device, there is no evaporator out superheat either under idle or under down-the-road conditions. This means that in both cases the evaporator out pressure equals the saturation vapor pressure corresponding to the evaporator out temperature.

7. In the case of the R-134a system, there is no condenser out subcooling under idle conditions but there is 25 °F subcooling under down-the-road conditions.

8. In the case of R-744 system, the gas cooler refrigerant side temperature effectiveness, *i.e.*, ratio of the refrigerant temperature drop to the inlet temperature difference (ITD) between the refrigerant and air stream has been assigned a value of 0.90 both under idle and down-the-road conditions.

9. The evaporator air side temperature effectiveness, *i.e.*, ratio of the air side temperature drop to the inlet temperature difference (ITD) between the refrigerant and air stream has a value of 0.90 under idle conditions and 0.85 under down-the-road conditions for both the systems.

IDEALIZED SYSTEM

As a prelude to the performance calculations for the realistic air conditioning systems with the aforementioned irreversibilities, it is instructive to perform the calculations for an idealized system devoid of the irreversibilities.

An idealized supercritical system embodies the following idealizations: (1) There is no pressure drop in the evaporator, (2) the evaporator temperature effectiveness is unity so that the conditioned air temperature is equal to the refrigerant evaporating temperature, (3) there is no evaporator out superheat so that the compressor suction temperature equals the saturation temperature corresponding to the evaporating pressure, (4) the isentropic efficiency of the compressor is 100% so that there is neither friction nor heat loss during the compression process, (5) there is no pressure drop in the gas cooler, and (6) the gas cooler temperature effectiveness is unity so that the gas cooler air out temperature is equal to the refrigerant out temperature. Such a system is depicted on the pressure-enthalpy diagram in Fig. 1 (a).

In 1928, Inokuty [60-61] presented a singularly lucid graphical method to find the optimum compressor discharge pressure for the idealized supercritical system described above. According to this method, we draw a tangent to the isotherm corresponding to the gas cooler cooling medium temperature (t_m) starting from the point of intersection of the 90% quality line and the suction pressure (P_{suc}) line on the pressure-enthalpy diagram [*vide* Fig. 1 (a)]. Then the pressure given by the point of contact of the tangent with the t_m isotherm directly gives the optimal compressor discharge pressure (P_{dis}) that will yield the maximum coefficient of performance for prescribed evaporating temperature and pressure.

An idealized subcritical system is similar to the supercritical system described above. In this case, the idealizations (5) and (6) are replaced by: (5a) There is no pressure drop in the condenser and (6a) there is no condenser out subcooling and the condenser out refrigerant temperature is equal to the dry bulb temperature of the incoming condenser cooling air. Such a system is depicted on the pressure-enthalpy diagram in Fig. 1 (b). Placed *vis-a-vis* the

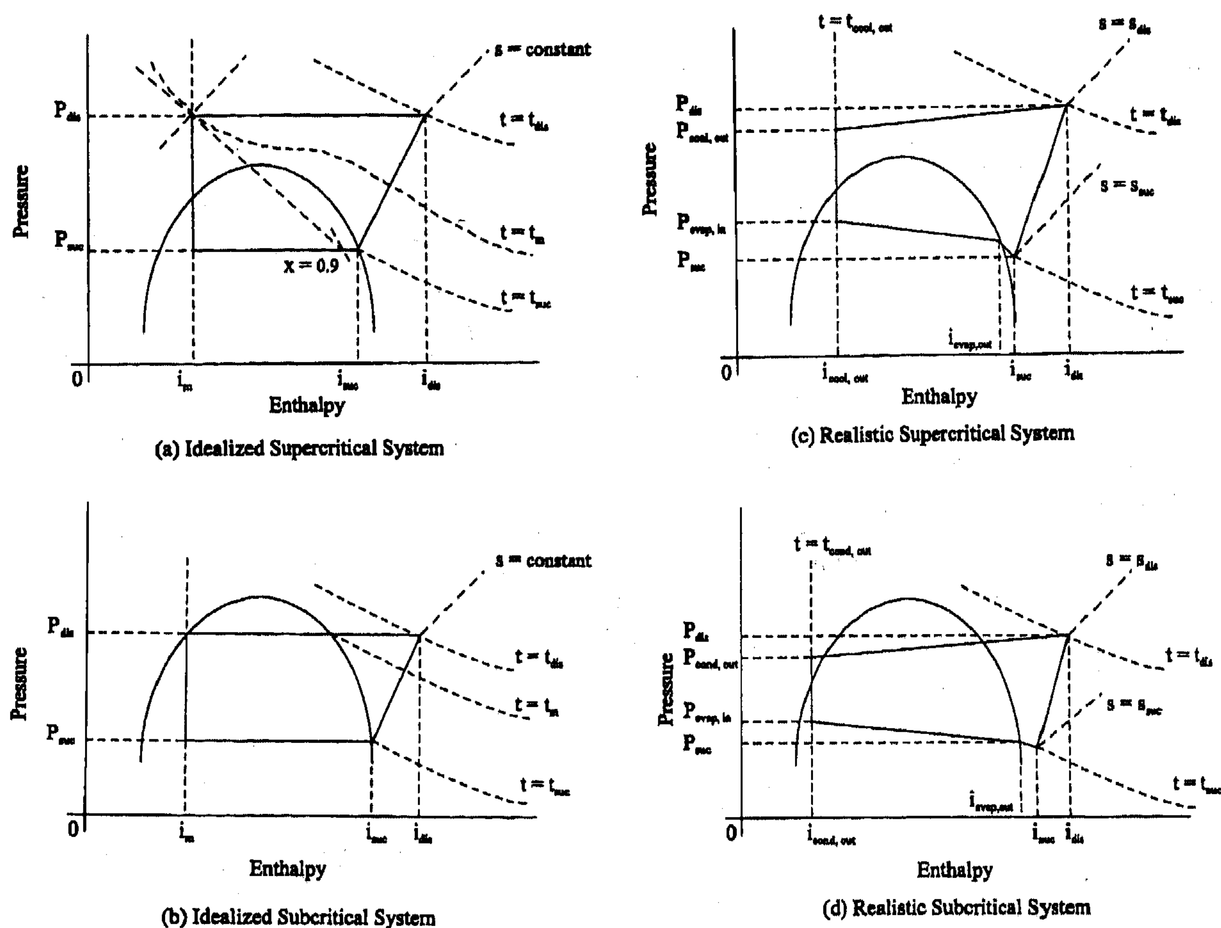


Figure 1. Schematic Representation of (a) Idealized Supercritical, (b) Idealized Subcritical, (c) Realistic Supercritical, and (d) Realistic Subcritical Vapor Compression Systems on the Pressure-Enthalpy Diagram.

corresponding idealized systems in Figs 1(a) and 1(b) are the realistic supercritical and subcritical systems in Figs. 1(c) and 1(d). An examination of these juxtaposed figures clearly elicits the system irreversibilities stated above.

Inokuty's graphical procedure cannot be applied to the idealized subcritical system since it is not possible to draw a tangent from the 90% quality line to the cooling air isotherm which now intersects the two-phase dome as shown in Fig. 1(b). However, in this case simple physical reasoning shows that thermodynamically possible maximum heat transfer from the condenser will occur if the condensing temperature equals the incoming air temperature. Accordingly, the compressor discharge pressure for the optimum coefficient of performance must equal the saturation vapor pressure corresponding to the condenser cooling medium temperature.

A little reflection shows that the idealizations embodied in the foregoing systems are the same as those in the absolute datum called the Carnot cycle with the important difference that the expansion process in these systems is isenthalpic rather than isentropic as in the Carnot Cycle. The implication of the isenthalpic expansion is that the cycle COP of the idealized system will be lower than that of the Carnot cycle. This latter COP is independent of the working fluid and is given simply as $T_l/(T_h - T_l)$ where T_l is the absolute temperature of the low pressure side and T_h that of the high pressure side.

IDEALIZED SYSTEMS PERFORMANCE

Presented in Table 2 is the predicted performance of the idealized R-744 and R-134a systems. The cycle COP's and other quantities in Table 2 must be viewed as the upper bounds that no real life system can exceed. A perusal of the

Table 2. Air Conditioning Performance Comparison of Idealized R-744 and R-134a Systems

Road Load Condition	Idle		Down-the-Road	
Refrigerant	R-744	R-134a	R-744	R-134a
Outside air temperature, °F	120	120	100	100
Outside air relative humidity, %	22.4	22.4	40	40
Outside air absolute humidity, lb_m H_2O/lb_m dry air	0.0167	0.0167	0.0167	0.0167
Evaporator air volume flow rate, ft^3/min	250	250	250	250
Evaporator air mass flow rate, lb_m/min	16.6	16.6	17.3	17.3
Conditioned air temperature, °F	70	70	50	50
Condenser/gas cooler air volume flow rate, ft^3/min	600	600	2,000	2,000
Condenser/gas cooler air mass flow rate, lb_m/min	40	40	138	138
Condenser/gas cooler air out temperature, °F	149	147	134	113
Refrigerant charge, lb_m	0.917	2.125	0.917	2.125
Refrigerant mass flow rate, lb_m/min	4.26	3.66	7.08	5.92
Compressor suction pressure, psia	852	86	653	60
Compressor suction temperature, °F	70	70	50	50
Compressor discharge pressure, psia	2,000	186	1,500	139
Compressor discharge temperature, °F	182	124	165	105
Compressor isentropic efficiency	1.0	1.0	1.0	1.0
Compressor power, HP	1.20	0.59	2.24	1.04
Condenser/gas cooler out pressure, psia	2,000	186	1,500	139
Condenser/gas cooler out temperature, °F	120	120	100	100
Condenser/gas cooler temperature effectiveness	1.0	1.0	1.0	1.0
Condenser/gas cooler cooling capacity, Btu/min	270	243	477	426
Evaporator in pressure, psia	852	86	653	60
Evaporator in temperature, °F	70	70	50	50
Evaporator out pressure, psia	852	86	653	60
Evaporator out temperature, °F	70	70	50	50
Evaporator temperature effectiveness	1.0	1.0	1.0	1.0
Latent load on evaporator, Btu/min	20	20	175	175
Sensible load on evaporator, Btu/min	199	199	207	207
Evaporator cooling capacity, Btu/min	219	219	382	382
Compressor power, Btu/min	51	25	95	44
Compressor power transmission loss, Btu/min	0	0	0	0
A/C blower power, Btu/min	14	14	14	14
System power, Btu/min	65	39	109	58
Cycle COP	4.26	8.80	4.00	8.71
System COP	3.37	5.62	3.50	6.59

results in Table 2 shows that, for equal cooling capacities, the compressor power for the R-744 system is 2.04 times the compressor power for the R-134a system under idle conditions and 2.16 times under down-the-road steady state conditions. Since, for fixed cooling capacity, the COP varies inversely with compressor power, the COP of the R-744 system is 0.48 times the COP of the R-134a system under idle and 0.46 times under down-the-road conditions. These COP ratios are consistent with the values anticipated from an examination of the molar latent heat of evaporation at 32 °F (*cf.* p. 3).

The numerical values in Table 2 suggest that the cycle COP decreases with increasing cooling capacity q_{cool}. To explore this trend further, a set of calculations was performed varying the dry bulb temperature of the outside air from 60 to 110 °F while keeping the conditioned air temperature fixed at 50 °F. For convenience, only sensible evaporator load was considered for this purpose. The results in Table 3 confirm the before-mentioned trend over a wider temperature and hence cooling capacity range. They also show that R-744 COP is always lower than R-134a COP. Explored subsequently in the paper will be the variation of the global warming impact with the cooling capacity.

Table 3. Variation of COP With Cooling Capacity

Ambient air, °F	q_{cool} Btu/min	R-744 COP	R-134a COP	COP ratio
60	41	16.9	48.4	0.35
70	82	9.4	24.0	0.39
80	124	6.3	15.5	0.40
90	165	4.6	11.3	0.41
100	207	4.0	8.7	0.46
110	248	2.9	7.0	0.42

REALISTIC SYSTEMS PERFORMANCE

Presented in Table 4 is the predicted performance of the realistic R-744 and R-134a systems. These performance comparisons are for the identical cooling capacities both under idle and down-the-road conditions. The compressor power indicated is the power derived from the driveshaft while the system power includes the power loss in the compressor mechanical drive and the evaporator blower power in addition to the compressor power. Two types of COP's are presented, *viz.*, cycle COP and system COP. The cycle COP is calculated taking into account only the compressor power while the system COP is calculated taking into account the system power. In both cases, the power used to force air over the condenser or gas cooler is not charged to the air conditioning system as the engine cooling fan supplying the cooling airflow is a part of the engine cooling system used to cool the radiator located directly behind the condenser or gas cooler.

A perusal of the tabular results shows that for equal cooling capacities, the compressor power consumption by the R-744 system is 1.10 times the power consumption by the R-134a system under idle conditions and 1.96 times the power consumption by the R-134a system under down-the-road conditions. Consistently, the cycle COP's of the R-744 system are 0.94 and 0.53 times those of the R-134a system under idle and down-the-road conditions. Likewise, the system COP's of the R-744 system are 0.92 and 0.54 times those of the R-134a system under idle and down-the-road conditions.

A comparison of the cycle COP's for the idealized systems (Table 2) with the cycle COP's for the realistic system (Table 4) shows that the idealized COP's are significantly higher than the corresponding realistic COP's due to absence of the system irreversibilities. The disparity between idealized and realistic cycle COP's of the R-744 system is seen to be less than that for the R-134a system. This is due to the fact that the compressor discharge pressure of the realistic R-744 system has been optimized as described by Inokuty *op. cit.* If a similar optimization is carried out for the realistic R-134a system, then the disparity between R-134a idealized and realistic COP's will narrow and the realistic R-134a COP's will be even higher than the realistic R-744 COP's.

A measure of the degree of optimization of the system is the ratio of the actual compressor discharge pressure to the idealized compressor discharge pressure. Obviously, for a perfectly optimized system the ratio is unity. It is seen from Tables 2 and 4 that the discharge pressure ratios for the R-744 system are respectively 1.06 and 1.17 under idle and down-the-road conditions. The corresponding discharge pressure ratios for the R-134a system are 2.16 and 2.04 which attest to the extent of unrealized optimization with the R-134a system.

Assuming that the idling time is 15% of the on-road time, the weighted average cycle COP of the realistic R-134a system can be taken as 0.15

Table 4. Air Conditioning Performance Comparison of Realistic R-744 and R-134a Systems

Road Load Condition	Idle		Down-the-Road	
Refrigerant	R-744	R-134a	R-744	R-134a
Outside air temperature, °F	120	120	100	100
Outside air relative humidity, %	22.4	22.4	40	40
Outside air absolute humidity, lb_m H_2O/lb_m dry air	0.0167	0.0167	0.0167	0.0167
Evaporator air volume flow rate, ft^3/min	250	250	250	250
Evaporator air mass flow rate, lb_m/min	16.6	16.6	17.3	17.3
Conditioned air temperature, °F	70	70	50	50
Condenser/gas cooler air volume flow rate, ft^3/min	600	600	2,000	2,000
Condenser/gas cooler air mass flow rate, lb_m/min	40	40	138	138
Condenser/gas cooler air out temperature, °F	162	159	123	118
Refrigerant charge, lb_m	0.917	2.125	0.917	2.125
Refrigerant mass flow rate, lb_m/min	5.65	6.06	9.04	6.97
Compressor suction pressure, psia	713	67	518	44
Compressor suction temperature, °F	64.4	61.6	41.0	38.5
Compressor discharge pressure, psia	2,125	401	1,750	283
Compressor discharge temperature, °F	251	219	256	201
Compressor isentropic efficiency	0.70	0.65	0.65	0.60
Compressor power, HP	3.83	3.49	9.06	4.60
Condenser/gas cooler out pressure, psia	2,125	381	1,750	269
Condenser/gas cooler out temperature, °F	133	176	116	122
Condenser/gas cooler temperature effectiveness	0.90	0.50	0.90	0.50
Condenser/gas cooler cooling capacity, Btu/min	380	371	759	580
Evaporator in pressure, psia	792	82	578	53
Evaporator in temperature, °F	64.4	67.1	41.0	43.5
Evaporator out pressure, psia	792	74	578	48
Evaporator out temperature, °F	64.4	61.6	41.0	38.5
Evaporator temperature effectiveness	0.90	0.90	0.85	0.85
Latent load on evaporator, Btu/min	20	20	175	175
Sensible load on evaporator, Btu/min	199	199	207	207
Evaporator cooling capacity, Btu/min	219	219	382	382
Compressor power, Btu/min	162	148	382	195
Compressor power transmission loss, Btu/min	5	4	12	6
A/C blower power, Btu/min	26	26	26	26
System power, Btu/min	193	178	420	227
Cycle COP	1.36	1.48	1.03	1.96
System COP	1.13	1.23	0.91	1.68

x 1.48 + 0.85 x 1.96 = 1.89 while that of the R-744 system as 0.15 x 1.36 + 0.85 x 1.00 = 1.08 corresponding to an average cooling capacity of 0.15 x 219 + 0.85 x 382 = 358 Btu/min. This gives the cycle COP ratio as 1.08/1.89 = 0.57. A similar calculation for the system COP's shows the weighted average COP ratio to be 0.94/1.61 = 0.58. This latter COP ratio will be used in the global warming impact analysis presented below.

HEAT EXCHANGER AND COMPRESSOR CHARACTERISTICS

Since the operating pressure and transport properties of R-744 are significantly different from those of R-134a, it is expected that the types of heat exchangers and compressor suitable for the R-744 system will be different from those used in the present R-134a system. In order to bring out the differences between the two sets of components, it is necessary to present first some basic heat transfer and fluid flow relations.

HEAT TRANSFER RATE

The heat transfer rate q (Btu/hr) of a heat exchanger of the type used in the systems under consideration is given by the following relation analogous to Ohm's law [62]:

$$q = \frac{t_h - t_c}{\frac{1}{h_a A_a} + \frac{\delta_w}{k_w A_w} + \frac{1}{h_r A_r}} \qquad (1)$$

where

A_a is the effective air side heat transfer area including that of the fins, ft^2
A_r is the effective refrigerant side heat transfer area including that of the fins, ft^2
h_a is the air side heat transfer coefficient, Btu/hr ft^2 °F
h_r is the refrigerant side heat transfer coefficient, Btu/hr ft^2 °F
k_w is thermal conductivity of the flow passage wall, Btu/hr ft °F
t_c is the cold fluid mean temperature, °F
t_h is the hot fluid mean temperature, °F
δ_w is wall thickness of the flow passage, ft.

The denominator on the right hand side of Eq. (1) represents the total thermal resistance R of the heat exchanger comprising three distinct resistances: the air side resistance $R_a = (1/h_a A_a)$, flow passage wall resistance $R_w = (\delta_w/k_w A_w)$, and refrigerant side resistance $R_r = (1/h_r A_r)$. For the metal heat exchangers of the type used in present mobile air conditioning systems, the approximate values of R_w, R_r, and R_a are $R_w/R = 0.005$, $R_r/R = 0.10$, and $R_a/R = 0.895$.

Introduction of R-744 in the mobile air conditioning system is expected to increase h_r due to superior transport properties and increased refrigerant flow rate. The upshot of this would be to lower $R_r = 1/h_r A_r$. Also, introduction of R-744 is expected to leave R_a unchanged while increasing $R_w = \delta_w/k_w A_w$ due to an increase in δ_w to accommodate higher pressure. This is easy to infer from the following radial stress relation for a thin-walled circular cylinder which shows that for a permissible radial stress, the wall thickness is directly proportional to the internal pressure:

$$\sigma_{rad} = \frac{P_i r_i}{\delta_w} \qquad (2)$$

where

P_i is the internal pressure, lb_f/ft^2
r_i is the inside radius of the cylinder, ft
δ_w is the wall thickness of the cylinder, ft
σ_{rad} is the radial stress, lb_f/ft^2.

The effect of refrigerant flow rate, flow passage geometry, and refrigerant transport properties on h_r can be inferred from an appropriate turbulent flow correlation for the Nusselt number. An exhaustive compendium of all available turbulent flow correlations has been prepared by Bhatti and Shah [63]. For the purpose of the present analysis, the following correlation due to Dittus and Boelter has been selected in view of its simplicity:

$$Nu = 0.024 Re^{4/5} Pr^{2/5} \qquad (3)$$

where

Nu is the dimensionless Nusselt number = hd_h/k,
Pr is the dimensionless Prandtl number = $\mu c_p/k$,
Re is the dimensionless Reynolds number = $\rho u_m d_h/\mu$,
c_p is isobaric specific heat of the fluid, Btu/lb_m°R
d_h is the flow passage hydraulic diameter, ft
h is the heat transfer coefficient, Btu/hr ft^2 °R
k is thermal conductivity of the fluid, Btu/hr ft °R

u_m is the fluid mean velocity, ft/hr
μ is dynamic viscosity of the fluid, lb_m/hr ft
ρ is the fluid density, lb_m/ft^3.

Introducing the defining relations for Nu, Re, and Pr into Eq. (3), we obtain the following expression for the heat transfer coefficient:

$$h = 0.024 u_m^{4/5} d_h^{-1/5} c_p^{2/5} k^{3/5} \rho^{4/5} \mu^{-2/5} \quad (4)$$

The fluid mean velocity u_m is expressible in terms of fluid mass flow rate $\dot{m}$ (lb_m/hr) as

$$u_m = \frac{4\dot{m}}{\pi \rho d_h^2} \quad (5)$$

Introducing Eq. (5) into Eq. (4) and denoting the resulting combination of transport properties by

$$\Phi_h = c_p^{2/5} k^{3/5} \mu^{-2/5} \quad (6)$$

we obtain,

$$h = 0.06 \dot{m}^{4/5} d_h^{-9/5} \Phi_h \quad (7)$$

which brings out with illuminating distinctness dependence of h on the fluid mass flow rate ($\dot{m}$), flow passage geometry (d_h), and fluid transport properties (Φ_h).

Applying Eq. (7) to the two systems with identical geometry (d_h), we obtain

$$\frac{h_{744}}{h_{134}} = \left(\frac{\dot{m}_{744}}{\dot{m}_{134}}\right)^{4/5} \left(\frac{\Phi_{h,744}}{\Phi_{h,134}}\right) \quad (8)$$

It is also useful to obtain an expression for the relative total thermal resistance of the two systems

$$\frac{R_{744}}{R_{134}} = \frac{R_{a,744} + R_{w,744} + R_{r,744}}{R_{a,134} + R_{w,134} + R_{r,134}} \quad (9)$$

where in view of the foregoing considerations

$$R_{a,744} = R_{a,134} \quad (10)$$

$$R_{w,744} = \left(\frac{P_{744}}{P_{134}}\right) R_{w,134} \quad (11)$$

$$R_{r,744} = \left(\frac{h_{134}}{h_{744}}\right) R_{r,134} \quad (12)$$

Knowing relative thermal resistance from Eqs. (9) through (12) in terms of the resistances of the baseline system, the relative heat transfer rate follows from Eq. (1) as

$$\frac{\dot{q}_{744}}{\dot{q}_{134}} = \left(\frac{\Delta t_{744}}{\Delta t_{134}}\right) \left(\frac{R_{134}}{R_{744}}\right) \quad (13)$$

where Δt is the difference between the hot and cold fluid mean temperatures.

PRESSURE DROP

The pressure drop Δp (lb_f/ft^2) in the flow passage of a heat exchanger of the type used in the systems under consideration is given by [62]

$$\Delta p = \frac{2 f l \rho u_m^2}{g_c d_h} \quad (14)$$

where ρ, u_m, and d_h are previously defined and the remaining symbols have the following meanings:

f is the dimensionless Fanning friction factor,
g_c is the proportionality constant in Newton's second law of motion = 32.174 x $3{,}600^2$ lb_m ft/lb_f hr^2
l is the flow passage length, ft.

There are a number correlations for the turbulent flow Fanning friction factor which have been brought together by Bhatti and Shah *op. cit.* For the purpose of the present analysis, the following correlation due to Blausius has been selected owing to its simplicity and accuracy:

$$f = \frac{0.0791}{Re^{1/4}} \quad (15)$$

where the dimensionless Reynolds number Re has been previously defined as $Re = \rho u_m d_h / \mu$.

Combining Eqs. (5), (14), (15) and denoting the resulting combination of transport properties by

$$\Phi_p = \rho^{-1}\mu^{1/4} \quad (16)$$

we obtain,

$$\Delta p = \frac{0.2414}{g_c} \dot{m}^{7/4} \frac{l}{d_h^{19/4}} \Phi_p \quad (17)$$

which shows with illuminating distinctness dependence of Δp on the fluid mass flow rate ($\dot{m}$), flow passage geometry (l, d_h), and fluid transport properties (Φ_p).

Applying Eq. (17) to the two systems with identical geometry (l, d_h), we obtain

$$\frac{\Delta p_{744}}{\Delta p_{134}} = \left(\frac{\dot{m}_{744}}{\dot{m}_{134}}\right)^{7/4} \left(\frac{\Phi_{p,744}}{\Phi_{p,134}}\right) \quad (18)$$

Table 5. Transport Properties of R-744 and R-134a [64-66]

	40 °F		120 °F	
Property	R-744	R-134a	R-744	R-134a
$c_{p,f}$, Btu/lb$_m$°F	0.6560	0.3276	-	0.3660
$c_{p,g}$, Btu/lb$_m$°F	0.4940	0.1967	1.2045	0.2260
k_f, Btu/hr ft°F	0.0610	0.0542	-	0.0434
k_g, Btu/hr ft°F	0.0114	0.0069	0.0438	0.0093
ρ_f, lb$_m$/ft^3	56.099	79.577	-	68.139
ρ_g, lb$_m$/ft^3	7.0240	1.0560	35.226	4.0323
μ_f, lb$_m$/ft hr	0.2270	0.5978	-	0.4829
μ_g, lb$_m$/ft hr	0.0392	0.0315	0.1037	0.0378
$\Phi_{h,f}$	0.2855	0.1367	-	0.1363
$\Phi_{h,g}$	0.1880	0.1051	0.4082	0.1235
Mean Φ_h	0.2368	0.1209	0.4082	0.1299
$\Phi_{p,f}$	0.0123	0.0110	-	0.0122
$\Phi_{p,g}$	0.0634	0.3989	0.0161	0.1094
Mean Φ_p	0.0379	0.2050	0.0161	0.0608

It is apparent from Eqs (8) and (18) that in order to determine the relative values of h and Δp for the two systems, we need the values of the transport property dimensional parameters Φ_h (Btu/lb$_m^{4/5}$ hr$^{1/5}$ °R) and Φ_p (ft$^{11/4}$/lb$_m^{3/4}$ hr$^{1/4}$) defined in Eqs. (6) and (16). These values are presented in Table 5 together with the precursor transport properties. The property values are given at two temperatures of 40 and 120 °F which are close to the evaporating and gas cooling or condensing temperatures under down-the-road conditions of Table 4. The subscripts f and g respectively denote liquid and vapor. Also, it may be noted that the tabular results for R-744 at 120 °F are those of superheated vapor corresponding to 1,750 psia whereas the remaining results are for saturated liquid and vapor for both the refrigerants.

EVAPORATOR HEAT TRANSFER RATE AND PRESSURE DROP

From Table 4 we note that $\dot{m}_{744}/\dot{m}_{134}$ = 9.04/6.97 under down-the-road conditions. Also from Table 5 we note that $\Phi_{h,744}/\Phi_{h,134}$ = 0.2368/0.1209 for the evaporator at 40 °F. With this value, Eq. (8) yields h_{744}/h_{134} = 2.50 for the evaporator. This dramatic increase in h_r for R-744 does not result in a dramatic decrease in $R = R_a + R_w + R_r$ since $R_r = 1/h_r A_r$ is not the dominant resistance. Moreover, for R-744 there is an increase in $R_w = \delta_w/k_w A_w$ due to increased δ_w which wipes out part of the gain due to increased h_r.

To provide a quantitative idea about the effect of increase in h_r, we note that for the evaporator $R_{a,134}$ = 0.0797 hr °F/Btu, $R_{w,134}$ = 0.0004 hr °F/Btu, and $R_{r,134}$ = 0.0089 hr °F/Btu. Introducing these values in Eqs. (9) through (12) together with P_{744}/P_{134} = 578/53 (*vide* Table 4) and h_{744}/h_{134} = 2.50, we obtain R_{744}/R_{134} = 0.0877/0.0890 = 0.99 which shows that there is only marginal reduction in overall thermal resistance of the R-744 system evaporator. According to Eq. (13), with R_{744}/R_{134} = 0.99 and $\Delta t_{744}/\Delta t_{134}$ = 34/34 (*vide* Table 4), we get $\dot{q}_{744}/\dot{q}_{134}$ = 1.01 indicating practically no change in the heat transfer rate of the R-744 system evaporator compared to that of the R-134a system. Since the required cooling capacity (heat transfer rate) of the R-744 as well as R-134a evaporator is the same, *viz.*, 382 Btu/min (*vide* Table 4), it follows from $\dot{q}_{744}/\dot{q}_{134}$ = 1.01 that despite favorable value of h_{744}/h_{134} = 2.50 there is no scope to reduce the thermal size of the R-744 evaporator.

We note from Table 5 that $\Phi_{p,744}/\Phi_{p,134}$ = 0.0379/0.2050 for the evaporator at 40 °F. With this value together with the previously stated value of $\dot{m}_{744}/\dot{m}_{134}$ = 9.04/6.97, Eq. (18) yields $\Delta p_{744}/\Delta p_{134}$ = 0.29 for the evaporator. This result shows that the pressure drop of the R-744 system evaporator is significantly lower than that of the R-134a system evaporator. This fact was taken into account in executing the calculations reported in Table 4.

CONDENSER OR GAS COOLER HEAT TRANSFER RATE AND PRESSURE DROP

From Table 4 we note that $\dot{m}_{744}/\dot{m}_{134}$ = 9.04/6.97 under down-the-road conditions. Also from Table 5 we note that $\Phi_{h,744}/\Phi_{h,134}$ = 0.4082/0.1299 for the condenser or gas cooler at 120 °F. With this value, Eq. (8) yields h_{744}/h_{134} = 3.87 for the condenser or gas cooler.

We note that for the condenser $R_{a,134}$ = 0.0814 hr °F/Btu, $R_{w,134}$ = 0.0005 hr °F/Btu, and $R_{r,134}$ = 0.0081 hr °F/Btu. Introducing these values in Eq. (9) through (12) together with P_{744}/P_{134} = 1,750/283 (*vide* Table 4) and h_{744}/h_{134} = 3.87, we obtain R_{744}/R_{134} = 0.0866/0.0900 = 0.96 which shows that there is only marginal reduction in overall thermal resistance of the R-744 system gas cooler. According to Eq. (13), with R_{744}/R_{134} = 0.96 and $\Delta t_{744}/\Delta t_{134}$ = 74.5/52.5 (*vide* Table 4), we get $\dot{q}_{744}/\dot{q}_{134}$ = 1.48 indicating that there is significant increase in the heat transfer rate of the R-744 system gas cooler compared to the R-134a system condenser. However, the required cooling capacity of the R-744 gas cooler is 382 + 382 = 764 Btu/min compared to the 382 + 195 = 577 Btu/min for the R-134a condenser (*vide* Table 4). This gives required value of $\dot{q}_{744}/\dot{q}_{134}$ = 1.32 compared to the achievable value of 1.48 with the R-744 gas cooler of the same thermal size as the condenser of the R-134a system. Thus in this case there is some scope for designing a more compact gas cooler for the R-744 system.

We note from Table 5 that $\Phi_{p,744}/\Phi_{p,134}$ = 0.0161/0.0608 for the condenser or gas cooler at 120 °F. With this value together with the previously stated value of $\dot{m}_{744}/\dot{m}_{134}$ = 9.04/6.97, Eq. (18) yields $\Delta p_{744}/\Delta p_{134}$ = 0.42 for the condenser or gas cooler. This result shows that the pressure drop of the R-744 system gas cooler is significantly lower than that of the R-134a condenser. This fact was taken into account in executing the calculations reported in Table 4.

COMPRESSOR EFFICIENCY

By definition, the isentropic efficiency η of the compressor is given by

$$\eta = \frac{\Delta i_o}{\Delta i} \qquad (19)$$

where

Δi_o is the enthalpy change across compressor for the adiabatic isentropic compression, Btu/lb_m
Δi is the enthalpy change across compressor for the actual compression process, Btu/lb_m.

For the compression process involving real gases, the expressions for Δi_o and Δi in terms of absolute temperature T (°R) and pressure P (lb_f/ft^2) are complex. However, for the ideal gas the expressions are simple and provide a good idea of the dependence of η on P_{dis}/P_{suc} and T_{dis}/T_{suc} where the subscripts $_{dis}$ and $_{suc}$ refer to the compressor discharge and suction sides respectively.

For the ideal gas with equation of state

$$Pv = RT \qquad (20)$$

undergoing adiabatic isentropic compression under the constraint imposed by the polytropic equation

$$Pv^{\gamma} = \text{constant} \qquad (21)$$

it can be shown that [67]

$$\Delta i_o = \left(\frac{\gamma}{\gamma - 1}\right)\left(\frac{RT_{suc}}{J}\right)\left[\left(\frac{P_{dis}}{P_{suc}}\right)^{(\gamma - 1)/\gamma} - 1\right] \qquad (22)$$

where

J is the mechanical-to-thermal energy conversion factor = 778.163 ft lb_f/Btu,
R = R_u/M is the gas constant, lb_f ft/lb_m °R,
R_u is the universal gas constant = 1545.32 lb_f ft/lb mole °R,
M is the molecular weight, lb_m/lb mole
v is the specific volume, ft^3/lb_m
γ is the specific heat ratio = c_p/c_v,

c_v is the isochoric specific heat, Btu/lb_m°R.

It may be noted that the polytropic equation [Eq. (21)] is merely a process equation connecting the process endpoints such that

$$P_{suc}v_{suc}^{\gamma} = P_{dis}v_{dis}^{\gamma} \qquad (23)$$

As such, it is distinct from the equation of state [Eq. (20)] which is applicable everywhere. Also it may be noted that Eq. (21) is a good representation of the adiabatic isentropic compression even for real gases.

For the ideal gas, the enthalpy change Δi entering Eq. (19) is exclusively a function of temperature being given by

$$\Delta i = A_o ln\left[\frac{T_{dis}}{T_{suc}}\right] + B_o(T_{dis}-T_{suc}) + \frac{C_o}{2}(T_{dis}^2-T_{suc}^2)$$

$$+ \frac{D_o}{3}(T_{dis}^3-T_{suc}^3) + \frac{E_o}{4}(T_{dis}^4-T_{suc}^4) \qquad (24)$$

where A_o, B_o, C_o, D_o, and E_o are the experimentally determined constants of the zero-pressure isobaric specific heat represented by

$$c_p = \frac{A_o}{T} + B_o + C_oT + D_oT^2 + E_oT^3 \qquad (25)$$

The experimental values of A_o through E_o, in English units, for the two refrigerants are as follows. $A_o = 0$, $B_o = 0.1076$, $C_o = 2.1867 \times 10^{-4}$, $D_o = -8.9232 \times 10^{-8}$, $E_o = 1.3871 \times 10^{-11}$ for R-744 [67] and $A_o = 6.8021$, $B_o = -1.2557 \times 10^{-3}$, $C_o = 4.3743 \times 10^{-4}$, $D_o = -1.4811 \times 10^{-7}$, $E_o = 0$ for R-134a [68].

For R-744, $\gamma = 1.10$, R = 35.1209 lb_f ft/lb_m °R, and typical value of T_{suc} = 500 °R and P_{dis}/P_{suc} = 3.5. For R-134a, on the other hand, $\gamma = 1.11$, R = 15.1457 lb_f ft/lb_m °R, and typical value of T_{suc} = 500 °R and P_{dis}/P_{suc} = 6.5. Equation (19) in conjunction with Eq. (22) shows that for high η, R as well as P_{dis}/P_{suc} should be as high as possible. An examination of the just-mentioned values then reveals that while higher R is a definite advantage low P_{dis}/P_{suc} is a disadvantage to the R-744 system compared to the R-134a system. Since $R = R_u/M$, we can state that lower molecule weight M of R-744 is a decided advantage as far as isentropic efficiency is concerned.

In order to obtain the numerical values of η for the two systems, we have the following data from Table 4 under down-the-road conditions: T_{suc} = 501 °R, T_{dis} = 716 °R, P_{dis}/P_{suc} = 1,750/518 = 3.38 for the R-744 system while T_{suc} = 498.5 °R, T_{dis} = 661 °R, P_{dis}/P_{suc} = 283/44 = 6.43 for the R-134a system. Using these values in Eqs. (19), (22), and (24) together with the values of the constants A_o through E_o, we obtain η = 28.69/45.26 = 0.63 for R-744 and η = 19.86/34.75 = 0.57 for R-134a. Carrying out a similar calculation under idle conditions of Table 4, we obtain η = 26.78/39.51 = 0.68 for R-744 and η = 19.90/34.44 = 0.58 for R-134a. The higher isentropic efficiency of the R-744 compressor was taken into account in performing the calculations reported in Table 4.

GLOBAL WARMING IMPACT

The global warming impact of a mobile air conditioning system is attributable to two sources: refrigerant leakage from the air conditioning system (direct impact) and fraction of the tailpipe emissions chargeable to the air conditioning system (indirect impact). As a prelude to the global warming impact analysis, certain preliminary information is presented below. It pertains to the refrigerant emission rate, emission of the tailpipe gases, fuel consumption, and global warming potential.

RATE OF REFRIGERANT EMISSION FROM THE AIR CONDITIONING SYSTEMS

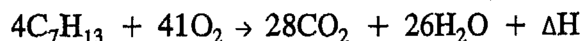
According to the estimates developed by the car makers around the world, the emission rate of R-134a from the R-134a system into the atmosphere is 55 gms per year [69]. Since the average refrigerant charge for the R-134a system is 1,000 gms, this annual emission rate amounts to 5.5% of the refrigerant charge.

As per Table 1, the refrigerant charge for the R-744 system should be 430 gms which is 43% the refrigerant charge for the R-134a. Assuming the R-744 emission rate to be the same fraction of the refrigerant charge as for R-134a while ignoring higher pressure of R-744, the emission rate of R-744 can be taken as 24 gms per year. For ready reference, these values together with the corresponding values for R-134a system are presented in Table 6.

EMISSION RATES OF THE TAILPIPE GREENHOUSE GASES

The principal greenhouse gas issuing from

the gasoline-powered motor vehicle is R-744, *i.e.*, CO_2. Its average emission rate is taken as 8,868 gms per gallon of gasoline. This value is arrived at as follows. Gasoline is a blend of hundreds of chemical compounds called hydrocarbons. For the purpose of stoichiometric calculations, it is assigned the formula C_7H_{13} and accordingly the chemical equation for the complete oxidation of gasoline can be written as



$$4C_7H_{13} + 41O_2 \rightarrow 28CO_2 + 26H_2O + \Delta H$$

where ΔH is the heat of combustion of gasoline. It has a value of 18,515 Btu per lb_m of gasoline which translates to 114,000 Btu per gallon of gasoline since one gallon of gasoline has a mass of 6.2 lb_m (*vide infra*).

Table 6. Refrigerant Charge and Annual Emission Rates for the R-134a and R-744 A/C Systems

A/C system	Refrigerant charge, gm	Emission rate, gm/yr
R-134a	1,000	55
R-744	430	24

Noting that the atomic weights of H, C, and O are 1, 12, and 16, respectively, we see from the preceding equation that 388 lb_m of gasoline (C_7H_{13}) combine with 1,312 lb_m of O_2 to produce 1,232 lb_m of CO_2 and 468 lb_m of H_2O. Thus upon complete combustion, one lb_m of gasoline produces 2,716/388 = 3.2 lb_m of CO_2. Since 1 gallon = 0.13368 ft^3 and density of gasoline is 46.1 lb_m/ft^3, we see that mass of one gallon of gasoline is 6.2 lb_m. Thus complete combustion of one gallon of gasoline produces 3.2 x 6.2 = 19.6 lb_m or equivalently 8,868 gms of CO_2.

In addition to CO_2, two more greenhouse gases, *viz.*, methane (CH_4) and nitrous oxide (N_2O) issue from the tailpipe and their amounts, determined from the chemical composition of the tailpipe gases, are 0.29 and 0.15 gm per gallon, respectively. For ready reference, the foregoing emission rates are collected in Table 7.

FUEL CONSUMPTION BY THE AIR CONDITIONING SYSTEMS

The annual fuel consumption by the R-134a mobile air conditioning system is taken as 23.5 gallons. This value is arrived at as follows. The average power consumption over the course of the year by the R-134a mobile air conditioning system is 2.5 HP which is 50% of the rated power to provide 2 tons of comfort cooling. Since 1 HP = 42.41 Btu/min, it is expressible as 2.5 x 42.41 = 106 Btu/min. This power is derived from the combustion of gasoline in the internal combustion engine with an average energy conversion efficiency of 27.5% so that the actual power consumption by the air conditioning system is 106/0.275 = 386 Btu/min.

Table 7. Emission Rates of the Principal Greenhouse Gases Issuing from the Tailpipe of a Gasoline-Powered Motor Vehicle

Greenhouse gas	Emission rate, gm/gallon
R-744 (CO_2)	8,868
CH_4	0.29
N_2O	0.15

Now on an average a vehicle in the USA is driven 10,000 miles per year. The cooling period being 1/3rd the driving season, the vehicle is driven 10,000/3 = 3,333 miles during the cooling period. The amount of time the air conditioning system is used for defrosting purposes is negligible compared to the time for comfort cooling. In order to determine the length of time the air conditioning system is on, the weighted average vehicle speed can be taken as 33 MPH which is the mean of idling and maximum legal speed limit of 65 MPH in the USA. As seen from Table 8, this value is in consonance with the value 0.45 x 48.24 + 0.55 x 20.23 = 32.9 MPH which follows from the U. S. Environment Protection Agency (EPA) test procedure to measure fuel consumption and emission for the corporate average fuel economy (CAFE) calculation [70].

Table 8. Particulars of the EPA Test Procedure

Test particular	Highway schedule	Urban schedule
Relative weighting	0.45	0.55
Total distance, miles	10.25	7.46
Total time, sec	765	1,327
Speed, MPH	48.24	20.23

With average vehicle speed of 33 MPH, the air conditioning system is on for 3,333/33 = 101 hrs, *i. e.*, 6,060 mins per year. Based upon actual average customer usage profile across the nation, the number of hours the mobile air conditioning system compressor is on is

approximately 100 hrs in substantial agreement with the value determined in accordance with the just-described duty cycle.

Taking cognizance of the foregoing duty cycle, we see that the annual energy consumption rate by the air conditioning system is 386 x 6,060 = 2,339,160 Btu/yr. With the calorific value of gasoline as 114,000 Btu/gallon (*vide supra*), the annual fuel consumption to operate the air conditioning system is 2,339,160/114,000 = 20.5 gallons.

According to the fuel economy tests conducted with the air conditioning system "on and off" on a 1994 GM vehicle under high load conditions (95 °F ambient dry bulb temperature, 40% relative humidity, and 850 W/m^2 solar load), the fuel usage rate to operate the air conditioning system was found to be 0.55 fluid oz/min. The same test under low load conditions (75 °F ambient dry bulb temperature, 50% relative humidity, and zero solar load) gave a value of 0.44 fluid oz/min. From these measurements, a useful rule of thumb could be deduced, *viz.*, the average fuel consumption rate to operate the mobile air conditioning system is 0.50 fluid oz/min. In order to compare these values with the estimated fuel usage rate of 20.5 gallons per year (6,060 min of operation), we note that 1 gallon is equal to 128 fluid oz whence the estimated fuel consumption to operate the air conditioning system is expressible as 128 x 20.5/6,060 = 0.43 fluid oz/min in good accord with the experimental values.

There is an additional source of fuel consumption to carry the air conditioning system year round. According to the estimates made by General Motors Corporation and others, 10 gallons of fuel is required for each additional 100 lb_m (45.36 kg) of vehicle mass per 10,000 miles [69]. Since the mass of a typical air conditioning system is 30 lb_m, the annual fuel consumption to carry the air conditioning system year round is 10 x 30/100 = 3 gallons. Adding this to the foregoing contribution, the annual fuel consumption by the R-134a air conditioning system is 23.5 gallons. This value is within the range of 20 to 25 estimated by automakers around the world [69].

Since the fuel consumption to operate the air conditioning system is inversely proportional to its COP, the annual fuel consumption to operate the R-744 system can be taken as 20.5/0.58 = 35.3 gallons since the COP of the realistic R-744 system is 0.58 times the COP of the present R-134a system as established in the preceding section.

Since the high pressure R-744 system is expected to be more robust than the low pressure R-134a system, we take its mass to be 45 lb_m which is 50% more than that of the R-134a system. Using the foregoing rule of thumb, we determine fuel consumption to carry the R-744 system year round as 10 x 45/100 = 4.5 gallons. Adding this to the foregoing contribution, the annual fuel consumption by the R-744 system is 39.8 gallons/yr. For ready reference, various computed results are presented in Table 9.

Table 9. Annual Fuel Consumption by the Air Conditioning Systems

	Fuel consumption, gallons		
A/C system	To operate A/C system	To carry A/C system	Total
R-134a	20.5	3.0	23.5
R-744	35.3	4.5	39.8

GLOBAL WARMING POTENTIAL OF SOME GREENHOUSE GASES

Table 10. Atmospheric Lives and Global Warming Potentials of Some Greenhouse Gases [71]

Greenhouse gas	Atmospheric life, yrs	GWP (mass basis)
R-744	50 - 200	1.0
CH_4	14.5	24.5
N_2O	120.0	320
R-134a	14.0	1,300
R-12	102.0	8,500

The global warming potential (GWP) of a greenhouse gas is a measure of its potency to absorb long wavelength infrared rays compared to the potency of the reference gas R-744. This unit of measurement was developed by the Intergovernmental Panel on Climate Change (IPCC). The GWP index varies with the timeframe called integration time horizon over which the gas is compared with R-744. A shorter timeframe emphasizes the climate forcing potential of shorter lived gases while the longer timeframe is more representative of the cumulative effect of a gas on climate change over its lifetime. Table 10 contains the values of atmospheric lives and GWP indices of the greenhouse gases based on an integration time horizon of 100

years as recommended by IPCC [71]. It may be noted that no single atmospheric life can be assigned to R-744 in view of different uptake rates of this gas by multiple sink processes.

GLOBAL WARMING IMPACT OF THE R-134A SYSTEM

As shown in Table 9, the annual fuel consumption by the R-134a air conditioning system is 23.5 gallons. Using the emission rates from Table 7, we get the absolute amounts of R-744, CH_4, and N_2O attributable to R-134a system as 208.4051 x 10^3, 6.8150, and 3.5250 gm/yr, respectively. Factoring in the GWP values of CH_4 and N_2O from Table 10, we obtain R-744 equivalent amounts of CH_4 and N_2O as 166.97 and 1,128.00 gm/yr, respectively. Adding these values, we obtain the indirect equivalent warming impact of the R-134a system as (208.4051 + 0.1670 + 1.1280) x 10^3 = 209.7 kg/yr. This shows that the equivalent warming impact of one gallon of gasoline is 209.7/23.5 = 8.9 kg/yr.

In order to estimate the direct global warming impact of the R-134a system, we note that the emission rate of R-134a system is 55 gm/yr (*vide* Table 6). Factoring in the GWP value of the R-134a from Table 10, this translates to direct equivalent warming impact of the R-134a system as 71.5 kg/yr. Adding the direct and indirect equivalent warming contributions, we see that total equivalent warming impact of the R-134a system is 281.2 kg/yr. For ready reference, various results are collected in Table 11.

Table 11. Global Warming Impact of the R-134a Mobile Air Conditioning System

Green-house gas	GWP	Absolute amount, kg/yr	R-744 equivalent amount, kg/yr
Direct Impact			
R-134a	1,300	0.0550	71.5000
Indirect Impact			
R-744 CH_4 N_2O	1 24.5 320	208.4051 0.0068 0.0035	208.4051 0.1670 1.1280
Total Impact			
R-134a + R-744 + CH_4 + N_2O	-	208.5	281.2

GLOBAL WARMING IMPACT OF THE R-744 SYSTEM

As presented in Table 9, the annual fuel consumption by the R-744 air conditioning system is 39.8 gallons per year. Using the emission rates from Table 7, we get the absolute amounts of R-744, CH_4, and N_2O attributable to the R-744 system as 353.3559 x 10^3, 11.5550, and 5.9767 gm/yr, respectively. Factoring in the GWP values of CH_4 and N_2O from Table 10, we obtain R-744 equivalent amounts of CH_4 and N_2O as 283.10 and 1,912.55 gm/yr, respectively. Adding these values, we obtain the indirect equivalent warming impact of the R-744 system as (353.3559 + 0.2831 + 1.9126) x 10^3 = 355.6 kg/yr.

Adding the direct equivalent warming impact of 24 gm/yr from Table 6, we obtain the total equivalent warming impact of the R-744 system as 355.6 kg/yr. For ready reference, various results are collected in Table 12.

Table 12. Global Warming Impact of the R-744 Mobile Air Conditioning System

Green-house gas	GWP	Absolute amount, kg/yr	R-744 equivalent amount, kg/yr
Direct Impact			
R-744	1	0.0240	0.0240
Indirect Impact			
R-744 CH_4 N_2O	1 24.5 320	353.3559 0.0116 0.0060	353.3559 0.2831 1.9126
Total Impact			
R-744 + CH_4 + N_2O	-	355.4	355.6

RELATIVE WARMING IMPACT OF THE R-134A AND R-744 SYSTEMS

A perusal of Tables 11 and 12 provides an idea about the total equivalent warming impact of the R-134a and R-744 systems. For the sake of clarity, the global warming results relating to direct, indirect, and total equivalent warming impact are brought together in Table 13.

In order to place the total equivalent warming impact of the air conditioning systems in the context of the total equivalent warming impact of the entire gasoline-powered vehicle,

it is instructive to develop an estimate of the latter impact. To that end, it is noted that according to the U. S. Department of Energy, the total fuel consumption by the nationwide fleet comprising 170 million motor vehicles is 118.4 billion gallons per year so that on an average a motor vehicle in the USA consumes 696.5 gallons of fuel per year [72]. Since the total equivalent warming impact of one gallon of gasoline is 8.9 kg/yr, we find that the total equivalent warming impact of the entire vehicle is 8.9 x 696.5/23.5 = 6,215.2 kg/yr.

Table 13. Total Equivalent Warming Impact in kg/yr of the R-134a and R-744 Systems

Equivalent warming	R-134a system	R-744 system	Ratio
Direct	71.5	0.0	0.0
Indirect	209.7	355.6	1.7
Total	281.2	355.6	1.3

Using 6,215.2 kg/yr as base, we see in conjunction with the results in Table 13 that the total equivalent warming impact of the R-134a system is 100 x 281.2/6,215.2 = 4.5% of the total equivalent warming impact of the fuel consumed by the entire vehicle. Similarly, total equivalent warming impact of the R-744 system is 5.7% of the total equivalent warming impact of the fuel consumed by the entire vehicle.

It is also instructive to compare *en passant* the total equivalent warming impact of the recently supplanted R-12 system with that of the R-134a system. To that end, the indirect impact of the R-12 system can be taken as 209.7 kg/yr which is the same as that of the R-134a system (see Table 13). Furthermore, we note that without conservation and containment measures in place, the emission rate of R-12 was 345 gm/yr. Factoring in the GWP value of R-12 from Table 10, this translates to direct equivalent warming impact of 2,932.5 kg/yr. Adding the direct and indirect contributions, we see that total equivalent warming impact of the R-12 system was 3,142.2 kg/yr which amounted to 100 x 2,932.5/6,215.2 = 50.6% of the total equivalent warming impact of the fuel consumed by the entire vehicle. Thus introduction of R-134a has greatly ameliorated the global warming impact of the mobile air conditioning system.

FACTORS GOVERNING GLOBAL WARMING

As a prelude to the sensitivity analysis, it is instructive to derive an analytical expression comprehending various factors governing the global warming impact. To that end, we note that the total equivalent warming impact (TEWI) is the sum of the direct equivalent warming impact (DEWI) due to refrigerant emission, indirect equivalent warming impact due to air conditioning system mass ($IEWI_m$), and indirect equivalent warming impact due to air conditioning system operation ($IEWI_o$).

DEWI is the product of two factors, *viz.*, refrigerant global warming potential (GWP) and refrigerant emission rate (m_{ref}). Likewise, $IEWI_m$ is the product of three factors, *viz.*, volumetric fuel consumption rate to carry unit mass year round (f), equivalent warming impact of unit volume of fuel (e), and air conditioning system mass (m_{sys}). As inferred from the rule of thumb, alluded to above, that 10 gallons of fuel is required per year for each additional 100 lb_m (45.36 kg) of vehicle mass, f = 10/45.36 = 0.22 gallon/kg (*loc. cit.* p. 17). The equivalent warming impact of one gallon of gasoline has already been established as e = 8.9 kg/yr and so is the mass of the air conditioning systems as m_{sys} = 13.6 kg for the R-134a and 20.4 kg for the R-744 system (*vide* pp. 17 *et seq.*).

As regards $IEWI_o$, it is expressible as $IEWI_o = \kappa \dot{q}_{sys}$ where κ [(kg/yr)/(Btu/min)] is the constant of proportionality and $\dot{q}_{sys}$ (Btu/min) is the power consumption by the air conditioning system. $\dot{q}_{sys}$ is expressible in terms of comfort cooling capacity $\dot{q}_{cool}$ (Btu/min) as $\dot{q}_{cool}/COP_{sys}$ where COP_{sys} is the coefficient of performance of the system. Thus $IEWI_o = \kappa \dot{q}_{cool}/COP_{sys}$. In order to find κ, we use the previously reported results for the R-744 system that $IEWI_o$ = 35.3 x 8.9 = 314.2 kg/yr corresponding to a weighted average $\dot{q}_{cool}$ = 358 Btu/min and COP_{sys} = 0.94. These values lead to κ = 0.83 (kg/yr)/(Btu/min). The same value of κ is obtained by using the previously reported results for the R-134a system, *viz.*, $IEWI_o$ = 20.5 x 8.9 = 182.9 kg/yr, $\dot{q}_{cool}$ = 358 Btu/min, and COP_{sys} = 1.61. Physically, κ represents the indirect equivalent warming impact of the unit cooling capacity system.

Taking into account the foregoing contributions, TEWI is expressible as follows in terms of refrigerant global warming potential

(GWP), refrigerant emission rate (m_{ref}), volumetric fuel consumption rate to carry unit mass year round (f), equivalent warming impact of unit volume of fuel (e), system mass (m_{sys}), indirect equivalent warming impact of the unit cooling capacity system (κ), system cooling capacity ($\dot{q}_{cool}$), and system coefficient of performance (COP_{sys}).

$$TEWI = GWPm_{ref} + fem_{sys} + \frac{\kappa \dot{q}_{cool}}{COP_{sys}} \quad (26)$$

This relation is used in performing the sensitivity analysis through appropriate variation of m_{ref}, COP_{sys}, m_{sys}, and $\dot{q}_{cool}$.

EFFECT OF R-134a EMISSION ON TEWI

It is apparent from the foregoing calculations that TEWI of the R-134a system is a strong function of the R-134a emission rate while TEWI of the R-744 system is quite insensitive to the R-744 emission rate. To explore the sensitivity of TEWI to m_{ref}, we assign a series of values to m_{ref} for the R-134a system while keeping all other factors in Eq. (26) fixed. The calculated results are reported in Table 14.

Table 14. Effect of R-134a Emission Rate on Total Equivalent Warming Impact

	TEWI, kg/yr		
m_{134} gm/yr	R-134a system	R-744 system	TEWI ratio
0	211	356	0.59
55	283	356	0.79
110	354	356	0.99
165	426	356	1.20

These results show that if the R-134a emission rate could be reduced to zero, TEWI of the R-134a system would be only 59% of the TEWI of the R-744 system. At the R-134a emission rate of 110 gm/yr, TEWI of the R-134a system just about matches TEWI of the R-744 system while at the R-134a emission rate of 165 gm/yr, TEWI of the R-134a system is 20% more than TEWI of the R-744 system.

EFFECT OF R-134a SYSTEM COP ON TEWI

As noted above, the baseline R-134a system selected for the purpose of the foregoing analysis is unoptimized whereas the R-744 system is optimized. COP_{sys} of the optimized R-134a system will be higher than the value of 1.61 used in the foregoing analysis. To explore the sensitivity of TEWI to COP_{sys}, we assign a series of values to COP_{sys} for the R-134a while keeping all other factors in Eq. (26) fixed. The calculated results are reported in Table 15.

Table 15. Effect of the R-134a System COP on Total Equivalent Warming Impact

	TEWI, kg/yr		
R-134a COP_{sys}	R-134a system	R-744 system	TEWI ratio
1.0	395	356	1.11
1.5	296	356	0.83
2.0	247	356	0.69
2.5	217	356	0.61
3.0	197	356	0.55

These results show that there is considerable scope for reducing TEWI of the R-134a system *vis-a-vis* TEWI of the R-744 system even with modest improvement in the R-134a COP_{sys} through realizable optimization (*loc. cit.* p. 9).

EFFECT OF R-744 SYSTEM MASS ON TEWI

In the foregoing analysis, the mass of the R-744 system was estimated to be 20.4 kg compared to 13.6 kg for the R-134a system. Since the actual mass of the R-744 system has not been definitely established, it is desirable to perform a sensitivity analysis by varying m_{sys} for the R-744 system. This is done in Table 16 with the aid of Eq. (26). The tabular results show that the effect of change in the R-744 system mass is minor.

Table 16. Effect of the R-744 System Mass on Total Equivalent Warming Impact

	TEWI, kg/yr		
R-744 m_{sys}, kg	R-134a system	R-744 system	TEWI ratio
13.6	283	343	0.83
15.0	283	346	0.82
17.5	283	350	0.81
20.0	283	355	0.80
22.5	283	360	0.79

EFFECT OF COOLING CAPACITY ON TEWI

The preceding global warming impact calculations were performed at the standard design point. It is desirable to obtain an idea as to how TEWI varies with the comfort cooling capacity since in practice the air conditioning system operates over a range of the evaporator loads.

In order to calculate TEWI from Eq. (26) for a variable q_{cool}, a knowledge of the corresponding COP_{sys} is essential. The results in Tables 2 and 4 show that there is no definite trend for the variation of COP_{sys} with q_{cool}. Therefore, for the purpose of the present calculation, the average values of COP_{sys} from Table 4, *viz.*, COP_{sys} = 1.46 for the R-134a system and 1.02 for the R-744 system will be used as a *modus vivendi.*

Varying q_{cool} with the above-mentioned values of COP_{sys} and using the same values of GWP, m_{ref}, f, e, and m_{sys} as before in Eq. (26), we obtain the results in Table 17. The tabular results show that at low q_{cool}, TEWI of the R-744 system is lower than that of the R-134a system. However, over the normal operating range of the mobile air conditioning systems (200 to 400 Btu/min), TEWI of the R-134a system is just about equal to or lower than that of the R-744 system.

Table 17. Effect of Comfort Cooling Capacity on Total Equivalent Warming Impact

	TEWI, kg/yr		
q_{cool}, Btu/min	R-134a system	R-744 system	TEWI ratio
100	155	121	1.28
200	212	203	1.04
300	269	284	0.95
400	326	365	0.89
500	382	447	0.85

CONCLUSIONS

Several substantiated conclusions, drawn from the foregoing review of the historical facts, system performance, heat exchanger characteristics, compressor characteristics, and global warming impact analysis, are stated below.

1. There is undeniable historical evidence suggesting that the *locus standi* of the high pressure R-744 system against the low pressure R-134a system is tenuous. At one time (1890 to 1930), R-744 was the preferred refrigerant for vapor compression machines for refrigeration and air conditioning. However, with the proliferation (starting in 1930) of the chlorofluorocarbon refrigerants, R-744 rapidly disappeared from service. On account of its lower COP and higher operating pressure, it could not compete with the chlorofluorocarbon refrigerants notably R-12. Since the COP and operating pressures of the R-134a system are similar to those of the R-12 system, it is reasonable to expect that the R-744 system is unlikely to compete with the R-134a system on these counts.

2. The R-744 system is receiving a great deal of attention due to the fact that it is a natural refrigerant and its direct global warming impact (due to direct release into the atmosphere) is negligible compared to that of the R-134a system. However, its indirect global warming impact (due to fuel consumption by the system) is 70% more than that of the R-134a system. This counterbalancing effect renders its total equivalent warming impact (direct + indirect) 30% more than that of the R-134a system.

3. The total equivalent warming impact (TEWI) of the present day R-134a mobile air conditioning system is no more than 4.5% of TEWI of the fuel consumed by the entire vehicle. This represents a dramatic improvement over TEWI of the recently supplanted R-12 system which was 50.4% of TEWI of the fuel consumed by the entire vehicle.

4. The fuel consumption by the R-744 system is 39.8 gallons/yr compared to 23.5 gallons/yr for the R-134a system. Thus introduction of R-744 would exacerbate the fuel usage at a time when the car makers around the world are striving hard to improve the fuel economy of their vehicles.

5. In the present analysis, an unoptimized R-134a system was used as a base and its performance was compared with an optimized R-744 system. The COP of the optimized R-744 system turns out to be 0.92 times the COP of the baseline R-134a system under idling and 0.54 times the COP of the baseline R-134a system under down-the-road conditions. The higher indirect and *ipso facto* total equivalent warming impact of the R-744 system is directly attributable to its lower COP. The optimization

of the R-134a system would render the outlook for the R-744 system even gloomier. A measure of the degree of optimization of the system is the ratio of the actual compressor discharge pressure to the idealized compressor discharge pressure. For a fully optimized system, this ratio is unity. For the R-744 system the ratio has the values of 1.06 and 1.17 under idling and down-the-road conditions, respectively. The corresponding values for the baseline R-134a system are 2.16 and 2.04 which attest to the extent of as yet unrealized optimization possible with the R-134a system.

6. The operating pressure of the R-744 system is significantly higher than that of the baseline R-134a system. Under idling conditions, the suction and discharge pressures of the R-744 are respectively 10.6 and 5.3 times the corresponding pressures of the baseline R-134a system. Likewise under down-the-road conditions, the suction and discharge pressures of the R-744 system are respectively 11.8 and 6.2 times the corresponding pressures of the baseline R-134a system. Aside from safety concerns, the high operating pressures pose refrigerant containment problems which are particularly accentuated in a mobile system. For this reason, the high pressure R-744 system is apt to be costlier than the low pressure R-134a system.

7. Combination of the transport properties including density, isobaric specific heat, dynamic viscosity, and thermal conductivity favors R-744 over R-134a as far as overall heat transfer rate, pressure drop, and compressor efficiency are concerned. However, these benefits are outweighed by the unfavorable thermodynamic characteristics of R-744 including temperature, pressure, and enthalpy change.

8. TEWI of the R-134a system is a strong function of the emission rate of R-134a while TEWI of the R-744 system is quite insensitive to the emission rate of R-744. If the R-134a emission rate could be reduced to zero, TEWI of the R-134a system would be only 59% of the TEWI of the R-744 system. At the R-134a emission rate of 110 gm/yr, TEWI of the R-134a system just about matches TEWI of the R-744 system while at the R-134a emission rate of 165 gm/yr, TEWI of the R-134a system is 20% more than TEWI of the R-744 system.

9. There is considerable scope for reducing TEWI of the R-134a system *vis-a-vis* TEWI of the R-744 system even with modest improvement in the R-134a COP through realizable optimization. For example, with a COP of 3.00 (attainable through optimization) instead of 1.61, TEWI of the enhanced R-134a system is 55% compared to 79% of TEWI of the optimized R-744 system.

10. The mass of the R-744 system has not been definitely established although there is no gainsaying the fact that it would be more than that of the R-134a system due to higher operating pressure. The sensitivity analysis shows that TEWI of the R-744 system is a weak function of its mass.

11. The effect of a decrease in the comfort cooling capacity is to lower TEWI of the R-744 system compared to TEWI of the R-134a system. However, over the typical operating range of the mobile air conditioning system (200 to 400 Btu/min), TEWI of the R-744 system is just about equal to higher than that of the R-134a system.

Taking due cognizance of the foregoing findings, it is concluded that it would be imprudent to enact the proposal to replace the R-134a system by the R-744 system since upon closer scrutiny the R-744 system is found to consume more fuel and is environmentally less benign. Moreover, being a higher pressure system, it poses significant safety and containment problems *inter alia*. Instead of investing in the R-744 system, our time and resources would be more judiciously spent in continuing to minimize emissions of R-134a from the present system or in exploring some other system that unreservedly could offer a net environmental benefit over the present system.

ACKNOWLEDGEMENT

Throughout the course of preparation of this paper the author has had numerous stimulating discussions with James A. Baker of Delphi Harrison Thermal Systems, General Motors Corporation. The author is appreciative of his useful input and constructive criticism which contributed in no small measure to the value of the paper.

REFERENCES

1. Anon., Historical Review of the Rise of Mechanical Refrigeration, Ice and Refrigeration, Vol. 21, pp. 45-59, August 1901.

2. J. C. Goosmann, Carbonic Acid Industry, Ice and Refrigeration, Vol. 29, pp. 292-294, Dec. 1905; Vol. 30, pp. 18-20, January 1906; Vol. 30, p. 57, February 1906; Vol. 30, pp. 93-94, March

1906; Vol. 30, pp. 223-224, 1906.

3. G. T. Voorhess, Operation of Multiple Effect of Compressors, Ice and Refrigeration, Vol. 31, pp. 52-59, January 1907.

4 J. A. Ewing, Mechanical Production of Cold, Cambridge University Press, London, 1908.

5. L. Horst, Verbesserte CO_2-Expansions-Kaltemaschine, Zeitschrift fur Sauerstoff-und Stickstoff-Industrie, Heft XII, III Jahrgang, pp. 241-246, Dezember 1911.

6. R. Plank, Die Verbesserung der Kohlensaure-Kaltemaschine durch Einfuhrung eines Expansionszylinders, Zeitschrift fur die gesamte Kalte-Industrie, Heft 3, XIX Jahrgang, pp. 41-44, Marz 1912.

7. R. Plank, Beitrag zur Erhohung der Kalteleistung und der Leistungsziffer von Kohlensaurekaltemaschine insbesondere bei hohen Kondensatordrucken, Zeitschrift fur die gesamte Kalte-Industrie, Heft 10, XX Jahrgang, pp.189-196, Oktober 1913.

8. A. H. Goelz, Carbonic Anhydride Refrigeration, Ice and Refrigeration, Vol. 59, pp. 17-18, 1920.

9. R. Plank, Versuche an einer Kohlensaure-kaltemaschine mit Zusatzkompression bei hohen Kondensatordrucken, Zeitschrift fur die gesamte Kalte-Industrie, 11 Heft, 28 Jahrgang, pp. 157-162, November 1921.

10. H. J. MacIntire, Carbon-Dioxide Refrigerating Machine, Power, Vol. 55, No. 1, pp. 24-26, January 3, 1922.

11. A. R. Stevenson, High Speed Carbon Dioxide Compression, Vol.10, No. 2, Refrigerating Engineering, August 1923.

12. Anon., Improvements in CO_2 Machines, Refrigerating Engineering, Vol. 10, No. 4, pp. 141-142, October 1923.

13. H. Brier, Pre-Cooling by Primary Evaporation and Multiple Effect Compression as applied to CO_2 Refrigerating Machines, Proc. Fourth Int. Cong. Refrig., pp. 473-500, 1924.

14. D. W. Mclenegan, Some Problems in the Application of Direct Connected Synchronous Motors to Carbon Dioxide Compressors, Refrigerating Engineering, Vol. 13, No. 7, pp. 220-229, January 1927.

15. J. C. Goosmann, Factors Governing the Liquefaction of Carbon Dioxide, Refrigerating Engineering, Vol. 14, No. 1, pp. 13-22, 29, July 1927.

16. E. H. Lamb, Tests on CO_2 Cycle, Refrigerating Engineering, Vol. 14, No. 3, pp. 108-109, September 1927.

17. Anon., Liquefaction of Carbon Dioxide, Refrigerating Engineering, Vol. 14, No. 5, pp. 150-153, November 1927.

18. J. C. Goosmann, The Progressive Development of Carbon Dioxide Refrigerating Methods, Refrigerating Engineering, Vol. 14, No. 6, pp. 188-189, December 1927.

19. J. C. Goosmann and F. R. Zumbro, Recent Improvements in CO_2 Equipment, Refrigerating Engineering, Vol. 16, No. 1, pp. 1-10, July 1928.

20. J. C. Goosmann, Carbon Dioxide Equipment, Refrigerating Engineering, Vol. 16, No. 3, pp. 83, 89, September 1928.

21. N. H. Hiller, The CO_2 Situation, Refrigerating Engineering, Vol. 16, No. 4, pp. 120-122, October 1928.

22. H. Brier, Pre-Cooling by Primary Evaporation Multiple Effect Compression & Automatic Regulation as Applied to CO_2 Refrigerating Machines, Volume Troisieme, Actes V[eme] Congres Int. du Froid, Rome, pp. 193-208, 9-15 Avril, 1928.

23. E. H. Lamb, Liquid Pre-Cooling in the Carbonic Acid Machine, Volume Troisieme, Actes V[eme] Congres Int. du Froid, Rome, pp. 209-241, 9-15 Avril, 1928.

24. F. A. Willcox and F. J. G. Foot, Liquid Pre-Cooling in the Carbonic Acid Machine, Volume Troisieme, Actes V[eme] Congres Int. du Froid, Rome, pp. 242-259, 9-15 Avril, 1928.

25. E. Holton, "Sabroe" Compressors for Motor-Vessels, Volume Troisieme, Actes V[eme] Congres Int. du Froid, Rome, pp. 265-276, 9-15 Avril, 1928.

26. H. Inokuty, Theory and Experiments on Vapour Compression Refrigerating Machine with Some Modified Cycles, Volume Troisieme, Actes V[eme]

Congres Int. du Froid, Rome, pp. 136-184, 9-15 Avril, 1928.

27. N. H. Hiller, Jr., A Mollier Diagram for Carbon Dioxide and its Use with Regard to Simple and Multiple Effect Compression, Volume Troisieme, Actes V[eme] Congres Int. du Froid, Rome, pp. 122-135, 9-15 Avril, 1928.

28. J. C. Goosmann, Carbon Dioxide Thermodynamics, Refrigerating Engineering, pp. 190-193, October 1933; pp. 245-248, November 1933; pp. 304-306, 318, 325, December 1933.

29. H. Inokuty and Z. Nagaoka, A Method of Controlling Automatically the Pressure of Multiple-Effect Refrigerating Machines, Proc. Int. Cong. Refrig., pp. 314-330, 1936.

30. J. P. Brouquet, Regard sur les Anhydrides, Rev. Prat. Froid, 71, Vol. 12, Nr. 32, pp. 25-27, 1956.

31. T. Midgley, Jr. and A. L. Henne, Organic Fluorides as Refrigerants, Ind. Eng. Chem., Vol. 22, No. 5, pp. 542-545, May 1930.

32. M. J. Molina and F. S. Rowland, Stratospheric Sink for Chlorofluoromethanes: Chlorine Atom-Catalyzed Destruction of Ozone, Nature, Vol. 249, pp. 810-812, 1974.

33. P. J. Crutzen, My Life with O_3, NO_x, and Other YZO_x Compounds, 1995 Nobel Lecture, Angewandte Chemie, International English Edition, Vol. 35, No. 16, pp. 1758-1777, September 6, 1996.

34. M. J. Molina, Polar Ozone Depletion, 1995 Nobel Lecture, Angewandte Chemie, Vol. 35, No. 16, International English Edition, pp. 1778-1785, September 6, 1996.

35. F. S. Rowland, Stratospheric Ozone Depletion by Chlorofluorocarbons, 1995 Nobel Lecture, Angewandte Chemie, International English Edition, Vol. 35, No. 16, pp. 1786-1798, September 6, 1996.

36. Anon., United Nations Environmental Programme (UNEP), Governing Council Decision 9/13B, May 1981.

37. W. R. Woolrich, The Men Who Created Cold, Exposition Press, New York, 1967.

38. O. Evans, The Abortion of the Young Steam Engineer's Guide, published by Fry and Krammer, Philadelphia, 1805.

39. J. Perkins, Apparatus for Producing Ice and Cooling Fluids, British Patent No. 6662, August 14, 1834.

40. G. B. Piatti, Improvements in the Production of Ice, British Patent No. 2397, 1856.

41. T. S. C. Lowe, An Improved Process for Manufacturing Ice and other Refrigerating Purposes, British Patent No. 952, March 30, 1867.

42. C. Linde, The Refrigerating Machine of Today, Ice and Refrigeration, Vol. 5, p. 173, 1893.

43. W. Raydt, Improvements in Machines or Apparatus for the Manufacture of Ice, British Patent No. 15, 475, November 24, 1884.

44. J. Harrison, Improvements in the Production and Use of Carbonic Acid for Refrigerating and Aerating, British Patent No. 1890, January 22, 1884.

45. F. Windhausen, Improvements in Apparatus for Refrigerating Purposes, British Patent No. 2864, February 27, 1886.

46. H. Miller, Halls of Dartford: 1785-1985, Hutchinson Benham, London, pp. 71-84, 1985.

47. R. Thevenot, A History of Refrigeration Throughout the World, translated from French by J. C. Fidler, International Institute of Refrigeration, Paris, France, 1979.

48. F. Whittenmier, Development of Carbon Dioxide Refrigerating Machines, Ice and Refrigeration, Vol. 51, p. 165, November 1916.

49. J. Sedlacek, Condensing Water and Air Cooling Apparatus for Refrigerating Machines, U. S. Patent No. 601,716, April 5, 1898; Refrigerating Machine, U. S. Patent No. 602, 199, April 12, 1898; Cooling and Sealing, U. S. Patent No. 604,162, May 17, 1898.

50. B. A. Nagengast, The 1920s: The First Realization of Public Air Conditioning, ASHRAE Journal, Vol. 35, pp. S49-S50, S52, S54, S56, January 1993.

51. B. Donaldson and B. Nagengast, Heat & Cold: Mastering the Great Indoors, ASHRAE, Atlanta, Georgia, 1994.

52. M. S. Bhatti, A Historical Look at Chlorofluorocarbon Refrigerants, Harrison Division, General Motors Corporation, April 29, 1993.

53. Anon., Air Conditioning for Packards ··· $274, Motor, p. 114, December 1939.

54. J. R. Holmes, Development of an Automobile Air Conditioning System for Underhood Installation, General Motors Engineering Journal, pp. 2-9, May-June, 1955.

55. P. J. Kent, Automobile Passenger Compartment Cooling From the Vehicle Manufacturer's Viewpoint, SAE National Passenger Car, Body, and Materials Meeting, Detroit, Michigan, March 3-5, 1953.

56. J. C. Goosmann, Gaseous, Liquid and Solid Carbon Dioxide, Proc. Int. Cong. Refrig., pp. 827-852, 1936.

57. G. Lorentzen and J. Pettersen, A New, Efficient and Environmentally Benign System for Car Air-Conditioning, International Journal of Refrigeration, Vol. 16, No. 1, pp. 4-12, 1993.

58. G. Lorentzen, J. Pettersen, and R. Bang, Method and Device for High Side Pressure Regulation in Transcritical Vapor Compression Cycle, U. S. Patent No. 5,245,836, September 21, 1993.

59. A. L. Henne and M. W. Renoll, Fluoro Derivatives of Ethane and Ethylene IV and Fluoroethanes and Fluoroethylene V, J. Am. Chem. Soc., Vol. 58, pp. 887-890, June 1936.

60. H. Inokuty, Graphical Method of Finding Compression Pressure of CO_2 Refrigerating Machines for Maximum Coefficient of Performance, Volume Troisieme, Actes V^{eme} Congres Int. du Froid, pp 185-192, Rome, 9-15 Avril, 1928.

61. H. Inokuty, Optimum Operating Conditions for Carbon Dioxide, Refrigerating Engineering, pp. 122-123, Vol. 16, No. 4, October 1928.

62. W. M. Kays and A. L. London, Compact Heat Exchangers, McGraw Hill Book Company, Third Edition, New York, 1984.

63. M. S. Bhatti and R. K. Shah, Turbulent and Transition Flow Convective Heat Transfer in Ducts, Chap. 4 in Handbook of Single-Phase Convective Heat Transfer, pp. 4-1 through 4-166,, edited by S. Kakac, R. K. Shah, and W. Aung, John Wiley & Sons, New York, 1987.

64. P. E. Liley and P. D. Desai, Thermophysical Properties of Refrigerants, ASHRAE, Atlanta, 1993.

65. R. B. Stewart, R. T. Jacobsen, and S. G. Penoncello, Thermodynamic Properties of Refrigerants, ASHRAE, Atlanta, 1986.

66. Anon., 1993 ASHRAE Handbook: Fundamentals, Atlanta, 1993.

67. K. Wark, Thermodynamics, McGraw-Hill Book Co., New York, 1983.

68. D. P. Wilson and R. S. Basu, Thermodynamic Properties of a New Stratospherically Safe Working Fluid - Refrigerant R-134a, ASHRAE Trans. 94, Part 2, pp. 2095-2118, 1988.

69. S. K. Fischer, P. J. Hughes, P. D. Fairchild, C. L. Kusik, J. T. Dieckmann, E. M. McMahon, and N. Hobday, Energy and Global Warming Impacts of CFC Alternative Technologies, Report for AFEAS and U. S. Department of Energy, December 1991.

70. J. D. Murrell and R. M. Heavenrich, Downward Trend in Passenger Car Fuel Economy - A Review of Recent Data, Technical Report, U. S. Environmental Protection Agency, January 1990.

71. D. L. Albritton, R. G. Derwent, I. S. A. Isaksen, M. Lal, and D. J. Wuebbles in Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emissions Scenarios, pp. 205-231, edited by J. T. Houghton, *et al.*, Cambridge University Press, London, 1994.

72. Anon., Annual Energy Review 1989, Report No. DOE/EIA-0384 (89), Energy Information Administration, U. S. Department of Energy, Washington, DC 20585, May 24, 1990.

970529

Numerical Simulation of the Flow in a Passenger Compartment and Evaluation of the Thermal Comfort of the Occupants

Joachim Currle
Mercedes-Benz AG

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The present study shows how the application of computational fluid dynamics can help to understand and optimize the flow field in a passenger compartment in order to achieve an optimum of thermal comfort for the occupants.

The flow field and temperature distribution in a passenger compartment have been calculated using the commercial CFD program STAR-CD. In combination with a thermophysiological model for the passengers, the computational results are used to evaluate the thermal comfort of the occupants and compare different geometrical modifications. The computational mesh consisting of around 3 millions hexahedra cells resolves all geometrical details of the car cabin including the air ducts, air nozzles and louvers. Natural convection, heat conduction and radiation are taken into account. One standard climatisation mode, the winter heat-up mode has been simulated. A special emphasis of the numerical investigations is the optimization of the ventilation of the front and rear legroom.

INTRODUCTION

Comfort aspects of passenger cars are a significant sales argument and have become a topic of rising importance during the development process of a new car. One important comfort aspect is the thermal comfort of the driver and other occupants, which is not only a question of comfortable and stress-free driving, but also a question of safety, because a freezing or sweating driver has a reduced concentration for traffic.

The microclimate is established by the heating, ventilation and air-conditioning system of the car (HVAC-system). Cool or warm air is distributed in the air distribution chamber of the HVAC-system to several air ducts, through which the air is directed to the vents located at different regions in the passenger compartment. Apart from the thermal comfort, the HVAC-system has to fulfill other tasks like e.g. optimal defrosting and defogging of the windows. Moreover the reduction of pressure drop and noise in the different components of the HVAC-system is a critical issue in order to achieve maximum overall performance of the whole system.

All these aspects have to be considered during the development of a new car. In the last decades, most of the investigations have been done experimentally by road testing or in wind tunnel facilities and other test rigs. Only recently, a new development tool has gained increased interest, namely computational fluid dynamics (CFD). Several authors have applied numerical flow simulation successfully for the optimization of different components of the HVAC-system [1-5]. Also the flow in the passenger compartment has been calculated by different authors [6-8]. On the other hand, computer programs have been developed to estimate the thermal state of the passengers and to predict the thermal comfort [9-11].

In the present paper, the numerical simulation of the local flow and temperature field in the passenger compartment using the general purpose CFD-code STAR-CD is coupled with an in-house program called TIM [9] to evaluate the thermal comfort of occupants. The paper is based upon extensive investigations with respect to the accuracy of the numerical results [12,13]. The calculated flow and temperature field in an idealized passenger compartment showed good agreement with measurements, not only for the overall flow topology, but also for the local values.

DESCRIPTION OF THE NUMERICAL MODEL

The flow and temperature field in the passenger compartment and in the air ducts is calculated with STAR-CD. The basic conservation equations, namely the Navier-Stokes and the energy equations, are solved numerically using the finite volume technique as implemented in STAR-CD [14]. The turbulent nature of the flow is modeled by the standard k,ε-model with logarithmic law of the wall.

The computational mesh consisting of around 3 million cells (hexahedra) is shown in figure 1. The mesh includes the passenger compartment of a new

Mercedes with four occupants and the air ducts. At every duct, a homogeneous velocity distribution normal to the inlet area is prescribed. At the rear outlets, where the air is extracted from the compartment, an artificial duct extension was added (Fig. 1) to assure numerical convergence.

At the walls of the passenger compartment (windows, floor, roof, doors, dashboard etc.) the ambient temperature together with a heat resistance is used as a thermal boundary condition [14] to model the complete heat transfer to the environment. The specified heat resistance is the sum of the resistance for the heat conduction through the walls and the resistance for the convective and radiative heat transfer from the outer surface to the ambient air. The heat transfer at the inner surface of the wall comes up as a result of the calculation and has not to be specified.

The interchange of thermal radiation between solid surfaces is calculated using the so-called 'discrete beam' approach [14]. In this approach, the solid surfaces, which are assumed to be of the gray Lambert type, are subdivided into contiguous non-overlapping areas or 'patches'. The number and extend of the patches are problem dependent and has to be chosen so as to resolve adequately the surface geometry as well as the radiation flux variations. The incoming and leaving radiation is assumed to be constant in one patch. For the present application, the windscreen, the side windows, the pillars, the doors, the roof and the floor for example are one patch respectively. The 4 occupants are divided into 14 patches, which correspond to the segments used in TIM (see figure 2). The view factor between two patches, which is the fraction of the total radiation leaving one patch that is intercepted by the other, is computed with the help of discrete beams. Over the enclosing unit hemisphere, a user-specified number of equally distributed beams are emitted from the nominal center of each patch. Each beam is traced through the fluid until it intercepts an opposing patch, thus defining a pair that may undergo radiant interchange.

EVALUATION OF THERMAL COMFORT

For the evaluation of the thermal comfort of the occupants an in-house program called TIM (thermo-physiological model of passenger) is used in conjunction with STAR-CD.

TIM is a multi-layer model which allows the computation of the thermal state of any occupant under a given heat load. In the model, the human body is divided into 14 segments as shown in figure 2. All segments are connected with their neighbors through the central blood-vessel system. Each segment consists of 4 layers (core, muscle, fat and skin). Clothes are also taken into account. The computational model comprises the heat transport mechanisms of the body (passive system) and the thermophysiological regulation (active system). The heat loss through the skin could be calculated for each segment individually. A more detailed description of TIM is given in [9]. Results of the program are the skin temperature of the segments and the determination of the thermal comfort of the passenger by using suitable thermal comfort criteria. Apart from global comfort criteria like the predicted mean vote (PMV) [15] and the thermal sensation vote (TSV) [16], also local comfort criteria, like the local equivalent temperature of the segments [17] are implemented. The equivalent temperature is defined as the temperature of an ideal reference room with equal air and enclosure temperature, in which the passenger would emit the same heat as in the real situation under consideration. Because of the strongly inhomogeneous conditions in the passenger compartment, local comfort criteria instead of global ones are of great importance.

Both programs, TIM and STAR-CD are coupled together in an iteration loop. The convective and radiative heat load at the surface of the passengers, as calculated by STAR-CD, is the input for TIM, whereas the calculated temperature at the skin or at the clothes of the occupants, is used as an updated thermal boundary condition for STAR-CD. The iteration loop is shown in figure 3. At the beginning of the loop, a neutral thermal state for the occupants is specified. Normally, 3-5 iterations are sufficient to get a converged result.

RESULTS

The numerical calculations presented in this study refer to a typical winter heat-up situation. Only the steady state is considered so far. Several geometrical modifications for the air ducts have been investigated with regard to the thermal comfort of the passengers. Only two of them will be presented here to show the successful application of the above described numerical method. These two modifications are related to the evaluation of the ventilation of the front and rear legrooms.

Figure 4 shows the two air ducts underneath the dashboard (providing the ventilation of the frontal legrooms) as well as the center console air duct (providing the ventilation of the rear legrooms). In the numerical simulation, only these air ducts are supplied with warm air (60°C). The air mass flux is 100 kg/h for each air duct. The ambient air temperature for the simulation was -20°C and the car is assumed to drive at constant speed of 100 km/h. An air-guiding element is contained in each of the two air duct underneath the dashboard (Fig. 4). The air duct at the driver's side shows three outlets, whereas the air duct at the passenger side has only two (Fig. 4). This corresponds to two different design stages, i.e., by comparing the computational results for the left and right side of the compartment, the two different design stages can be evaluated and compared with each other.

The numerical computation provides a large amount of results. At every cell center of the computational mesh (around 3 millions) values for the three velocity components, pressure, turbulence quantities, density

and temperature are available. Temperature distribution, convective and radiant heat fluxes at solid surfaces (i.e. walls of the compartment, seats and occupants) are also part of the computational results. On the one hand, these results can be used to get a very detailed insight into the local flow and temperature field in the compartment, especially in the proximity of the passengers. Cold or warm air streams at different parts of the body could be traced back to their origin. On the other hand, averaged values like averaged heat fluxes or temperatures for different components (e.g. roof, floor) or certain fluid volumes (e.g. driver's legroom, head region at driver's side) can be derived from the results and used as easy criterias for the comparison of different geometrical modifications during the development process. In the following figures and tables, a brief overview will be given about how the numerical results could be exploited in that sense and how they can be used for practical purposes.

In figure 5, two selected section plots for the velocity field are shown. The section x=390mm clearly shows the air jets coming out of the air outlets in the ducts underneath the dashboard. The third air outlet at the driver's side strongly reduces the amount of air flowing through the outlet in the middle. The section z=200mm gives insight into the ventilation of the rear legroom. The air is guided by 7 louvers at each side and is directed towards the inner foot of the rear passengers.

Figure 6 shows the temperature distribution at the two sections x=390mm and y=-335mm (seat center plane of driver's side). The typical temperature distribution versus the height of the compartment, i.e. higher temperature in the bottom regions and lower temperature in the top regions, is represented in figure 6b.

Very revealing with regard to thermal comfort aspects (see explanations for figure 9) are plots showing the temperature distribution in the vicinity of the occupants (Fig. 7).

The distribution of temperature and convective heat flux is shown in figure 8 for the inner surface of the floor and could be investigated in the same manner for every interesting solid surface. Note, that the negative values for the convective heat flux in figure 8b imply, that the fluid looses heat at the floor. The high negative heat fluxes at the floor of the frontal legrooms lead to a strong cooling of the warm air on its way to the rear.

Table 1 shows averaged values for temperature, convective heat flux and thermal radiation for a selection of solid surfaces.

Concerning the evaluation of the thermal comfort, it has proved reasonable to represent the results of TIM, in diagrams as shown in figure 9 [18]. For every occupant, the local equivalent temperatures for the 14 segments of TIM are plotted together with two borderlines. On top of each diagram, the same is done for the global equivalence temperature. The two borderlines determine the limits, in between 80% of the people will feel comfortable according to the PMV-index of Fanger. Those parts of the body, having local equivalence temperatures outside of the borderlines, reveal, that the ventilation doesn't perform well in this area and should be improved.

Comparing the diagrams for the driver and the passenger clearly shows, that the three outlets in the air duct underneath the dashboard at the driver's side results in better comfort values for the lower parts of the bodies (upper legs, lower legs, feet). The warm left foot of the passenger in figure 9 is also indicated in figure 7. Looking at figure 5a, it can be concluded, that too much warm air is directed towards this foot. The comparison of the top parts of the bodies doesn't appear reasonable, because the opening of the panel vents would change the situation at these parts.

The study of the diagrams for the rear occupants shows, that the ventilation of the rear legroom is far from being optimal, because the feet lying close to the center of the compartment, are experiencing the warm air jets coming out of the center console air duct, whereas the outer feet are to cold (see also figure 5b). Of course this observation is obvious without doing any simulation. Another fact is worth mentioning: The comfort values for the feet are not the same for the driver's and the passenger's side. This is mainly due to the different air flow coming from the frontal legrooms.

	T [^{0}C]	q [W/m^2]	q_{rad} (leaving) [W/m^2]	q_{rad} (incoming) [W/m^2]
floor	**14.0**	**-188.0**	**391.5**	**438.2**
roof	**13.9**	**-47.0**	**387.9**	**417.9**
driv. door	**24.2**	**-64.6**	**440.9**	**412.8**
pass. door	**17.6**	**-34.4**	**405.3**	**401.9**
driv. seat	**26.9**	**adiab.**	**456.0**	**427.9**
pass. seat	**26.4**	**adiab.**	**452.7**	**422.9**
driv.B-pillar	**19.3**	**-71.6**	**417.0**	**439.6**
pass.B-pill.	**17.6**	**-74.1**	**407.4**	**423.6**
rear door	**22.4**	**-51.3**	**430.7**	**414.5**

Table 1: Averaged temperature, convective heat flux, leaving and incoming thermal radiation for different components

CONCLUSIONS

The present study has shown, how computational fluid dynamic can be applied for the evaluation of the thermal comfort of the passengers in car cabins. The computational results are able to give a detailed insight into the complex flow field in the passenger compartment and by this means increase the understanding of the flow physics. The numerical method has proved to be a powerful tool for a systematic comparison of different ventilation concepts and moreover, it allows a sophisticated optimization of different components of the HVAC-system with regard to the microclimate in the compartment and the thermal comfort of the occupants.

An optimal use of the method is believed to be in an early stage of the development process, where no prototypes are available. Thus, the number of

experimental tests as well as time and costs during the development process of new cars could be greatly reduced.

In the future, the method will be applied to a summer cool-down mode, in the course of which also solar radiation will be considered. First numerical tests have shown encouraging results. Moreover, the numerical calculation of the real transient behavior during a winter heat-up or a summer cool-down situation is planned for the near future, including the complete conjugate heat transfer problem.

REFERENCES

1 Currle, J., Application of Computational Fluid Dynamics for the Optimization of Air Ducts, 29th ISATA, Florence, 1996.

2 Currle, J., Simulation of the Two-Phase Flow in the Waterbox of a HVAC System, SAE Paper 960962, 1996.

3 Anderson, G., Hall, T., CFD analysis of climate control systems, Automotive Engineer, 1994.

4 Werner, F., Frik, S., Optimization of an automotive HVAC module by means of computational fluid dynamics, VDI Berichte Nr. 1153, 1994.

5 Fischer, D., Airflow Simulation Through Automotive Blowers Using Computational Fluid Dynamics, SAE Paper 950438, 1995.

6 Chao-Hsin Lin et al., An Experimental and Computational Study of Cooling in a Simplified GM-10 Passenger Compartment, SAE Paper 910216, 1991.

7 Komoriya, T., Kobayashi, T., Taniguchi, N., Numerical Simulation of the Flow in a Vehicle Passenger Compartment using General Co-Ordinate System with Finite Volume Method, JSAE Review, Vol 12, No.3, 1991.

8 Ishihara, Y. et al., Determination of Flow Velocity Distribution in a Vehicle Interior Using a Visualization and Computational Techniques, SAE Paper 910310, 1991.

9 Maue, J., Wahl, D., Breitling, T., Rößner, W., Theoretische Untersuchungen zum thermischen Komfort von Fahrzeuginsassen, Haus der Technik e.V., Tagung: Thermische Behaglichkeit in Fahrzeugen, 1995.

10 Althabegoity, F., Cordier, A., Thellier, F., MATHER - a model to study thermal comfort in a vehicle, Proc. 2nd Inter. Cong. on Physiological Anthropology, Kiel, Germany, 1994.

11 Ingersoll, J.G., Kalman, T.G., Maxwell, L.M., Automobile Passenger Compartment Thermal Comfort Model - Part II: Human Thermal Comfort Calculation, SAE Paper 920266, 1992.

12 Herwig, H., Klemp, K., Schmücker A., Currle, J., Ventilation of passenger car cabins: comparison of experimental data and numerical calculation, Engineering Research, Vol. 62, No. 6, 1996.

13 Currle, J., Numerical Analysis of the Microclimate in a Passenger Compartment, VDI Berichte Nr. 1153, 1994.

14 Computational Dynamics Ltd., STAR-CD Manual Version 2.3, 1996.

15 Fanger, P.O., Thermal Comfort, Kopenhagen, Danish Technical Press, 1970.

16 Taniguchi, Y. et al., Study on Car Air Conditioning System Controlled by Car Occupant's Skin Temperatures. Part 1: Research on a Method of Quantitative Evaluation of Car Occupants' Thermal Sensations by Skin Temperatures, SAE Paper 920169, 1992.

17 Holmer, I., Nilsson, H., Bohm, M., Browen, A., Noren, O., Evaluation of Thermal Environment in Automotive Vehicles, 2nd Int. Conf. Vehicle Comfort, Bologna, 1992.

18 Wahl, D., Rechnerische Untersuchungen zum thermischen Komfort in Fahrzeugkabinen, Daimler-Benz Report, 1995.

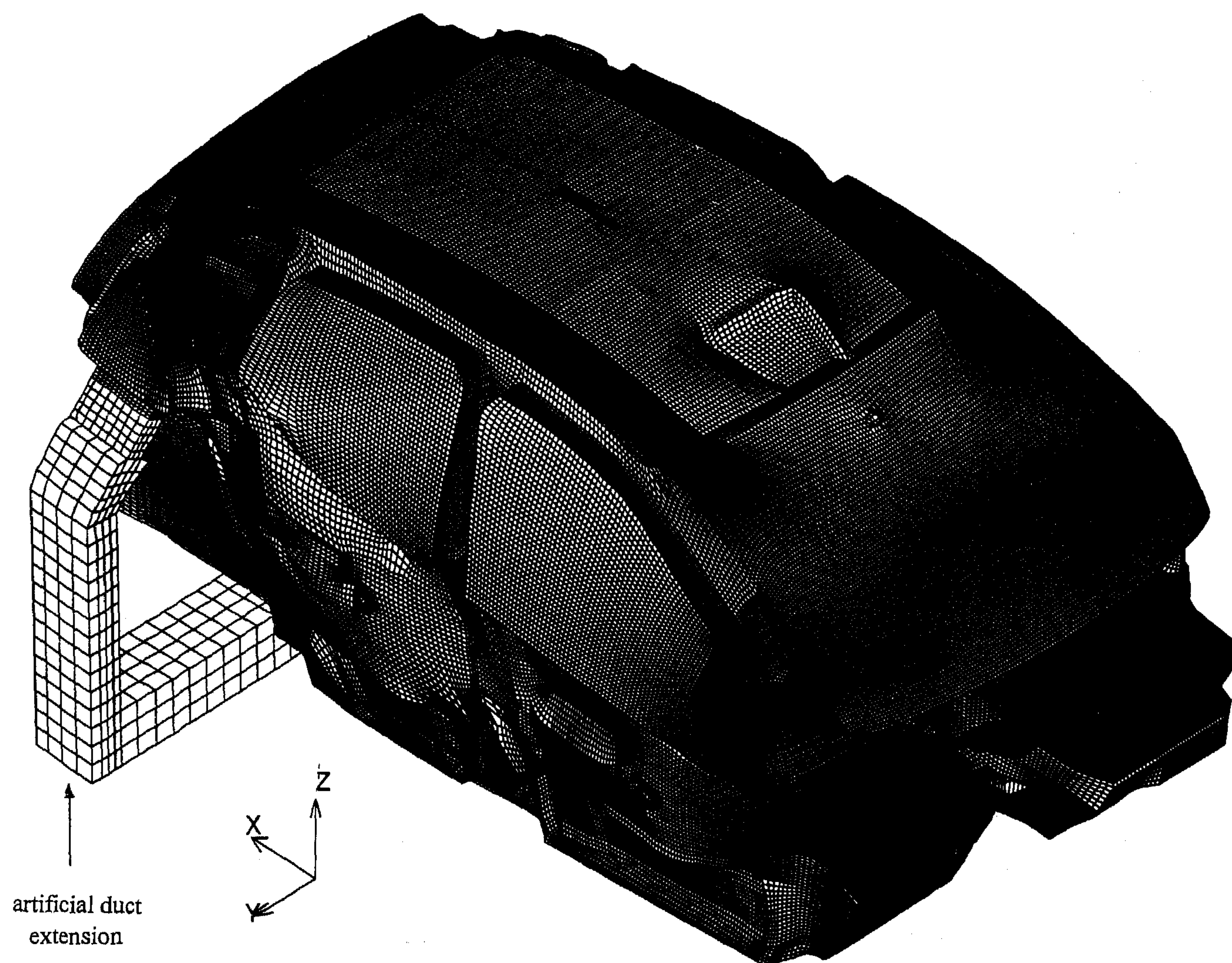


Figure 1: Computational mesh

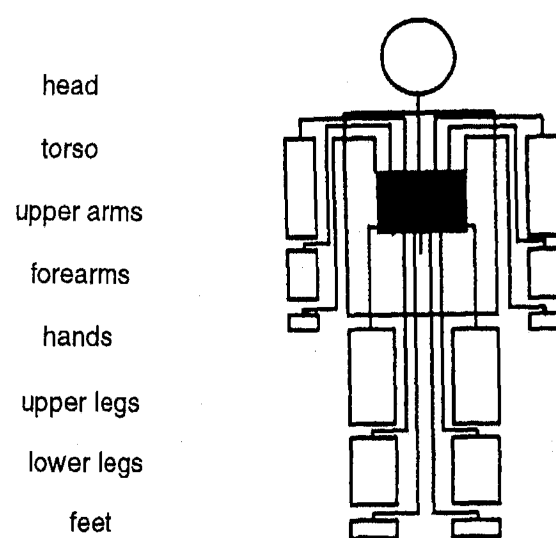


Figure 2: Segments in TIM (thermophysiological model of passenger)

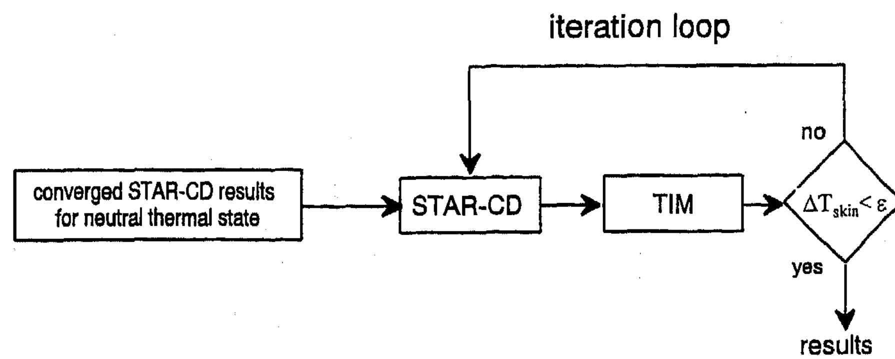


Figure 3: Coupling of STAR-CD and TIM

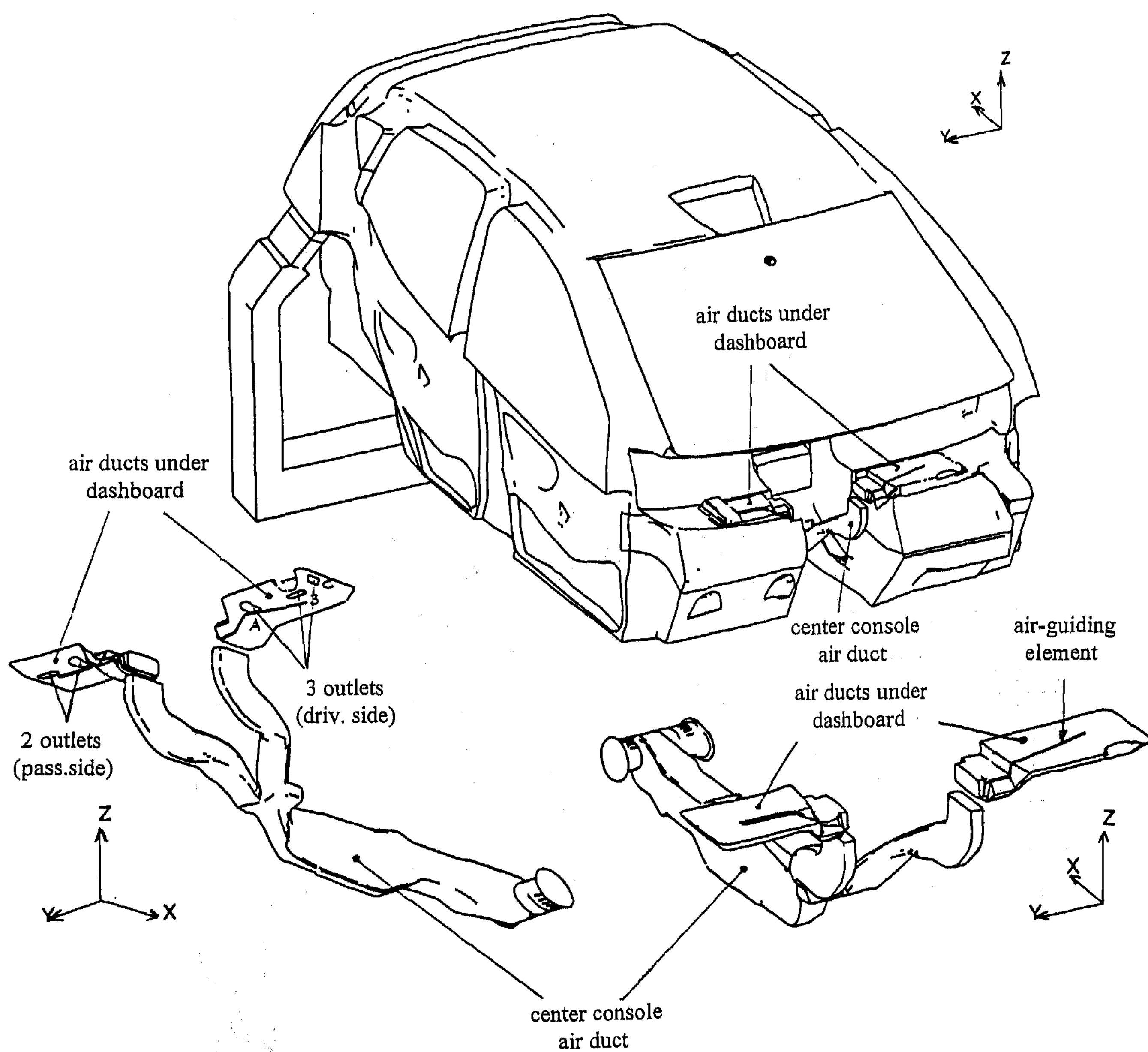
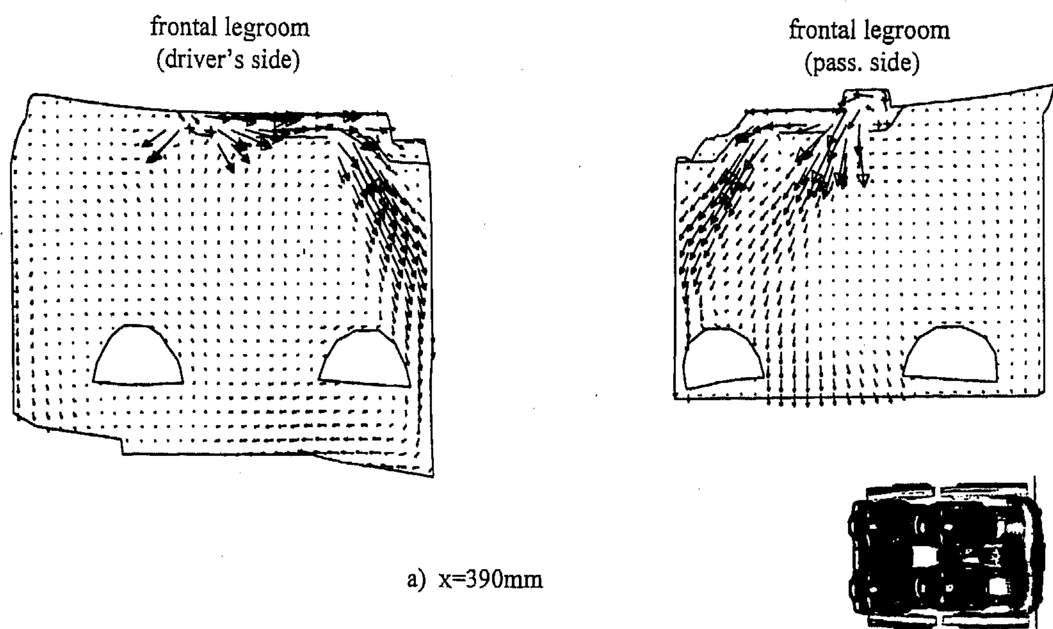
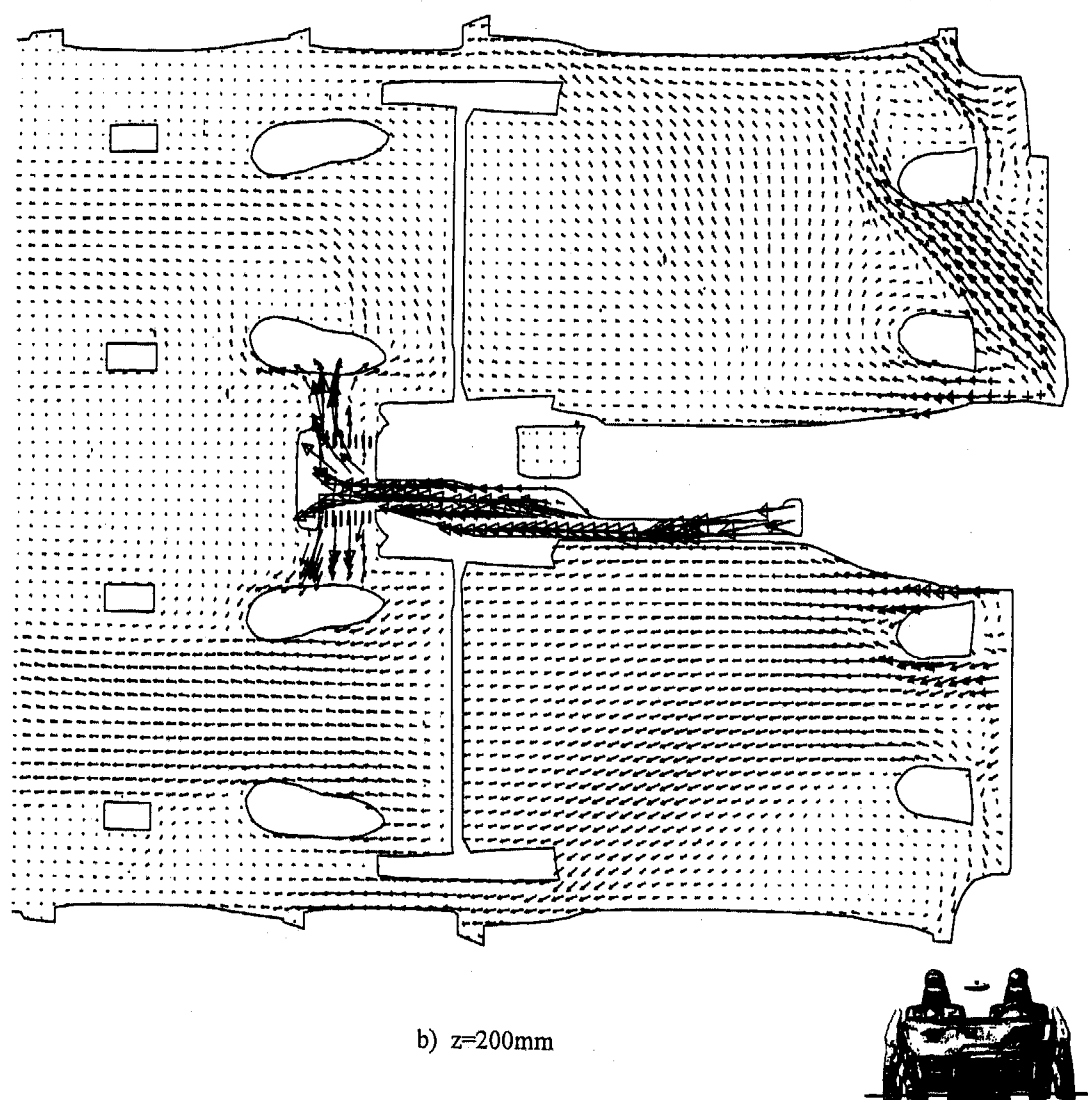


Figure 4: Air ducts of the computational model (2 views)

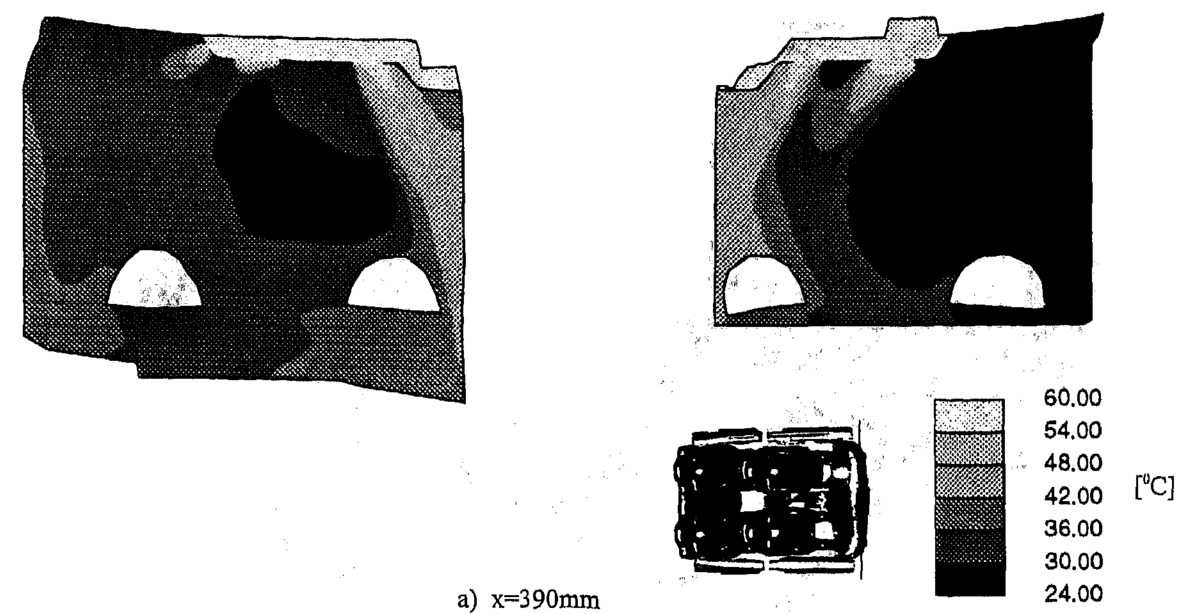


a) x=390mm

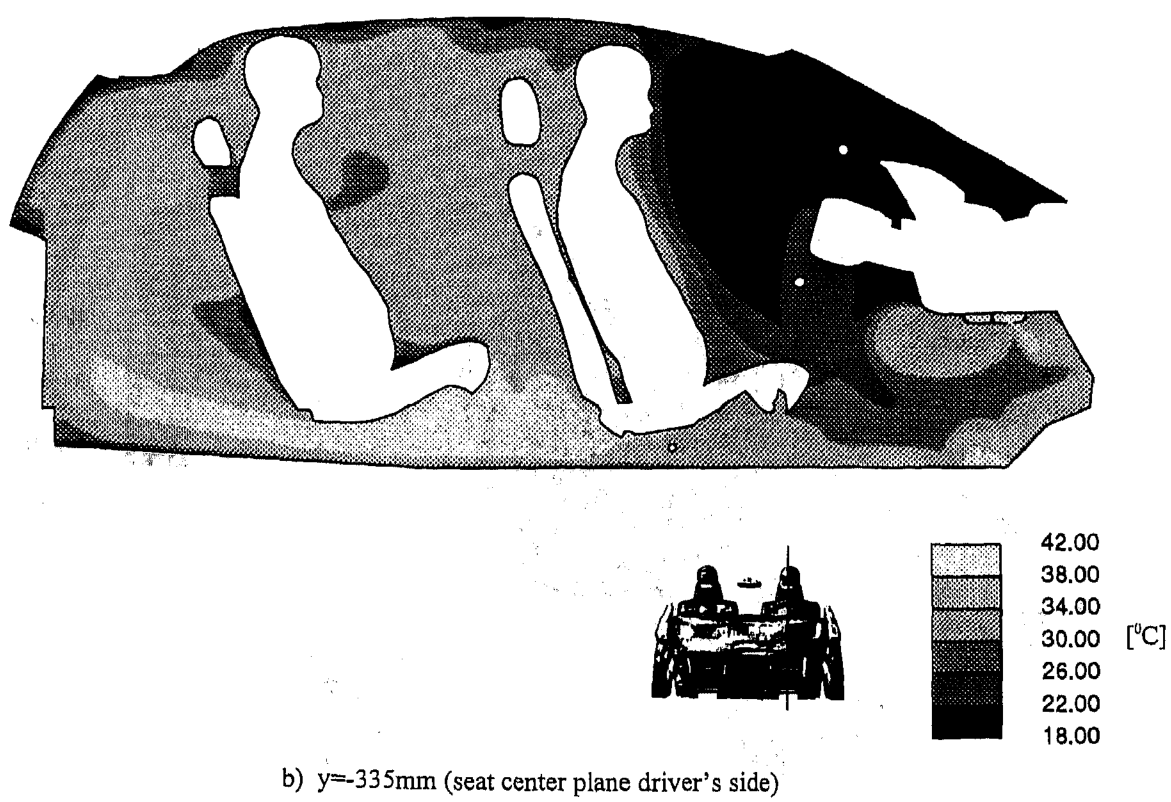


b) z=200mm

Figure 5: Velocity vectors at two sections



a) x=390mm



b) y=-335mm (seat center plane driver's side)

Figure 6: Temperature distribution at two sections

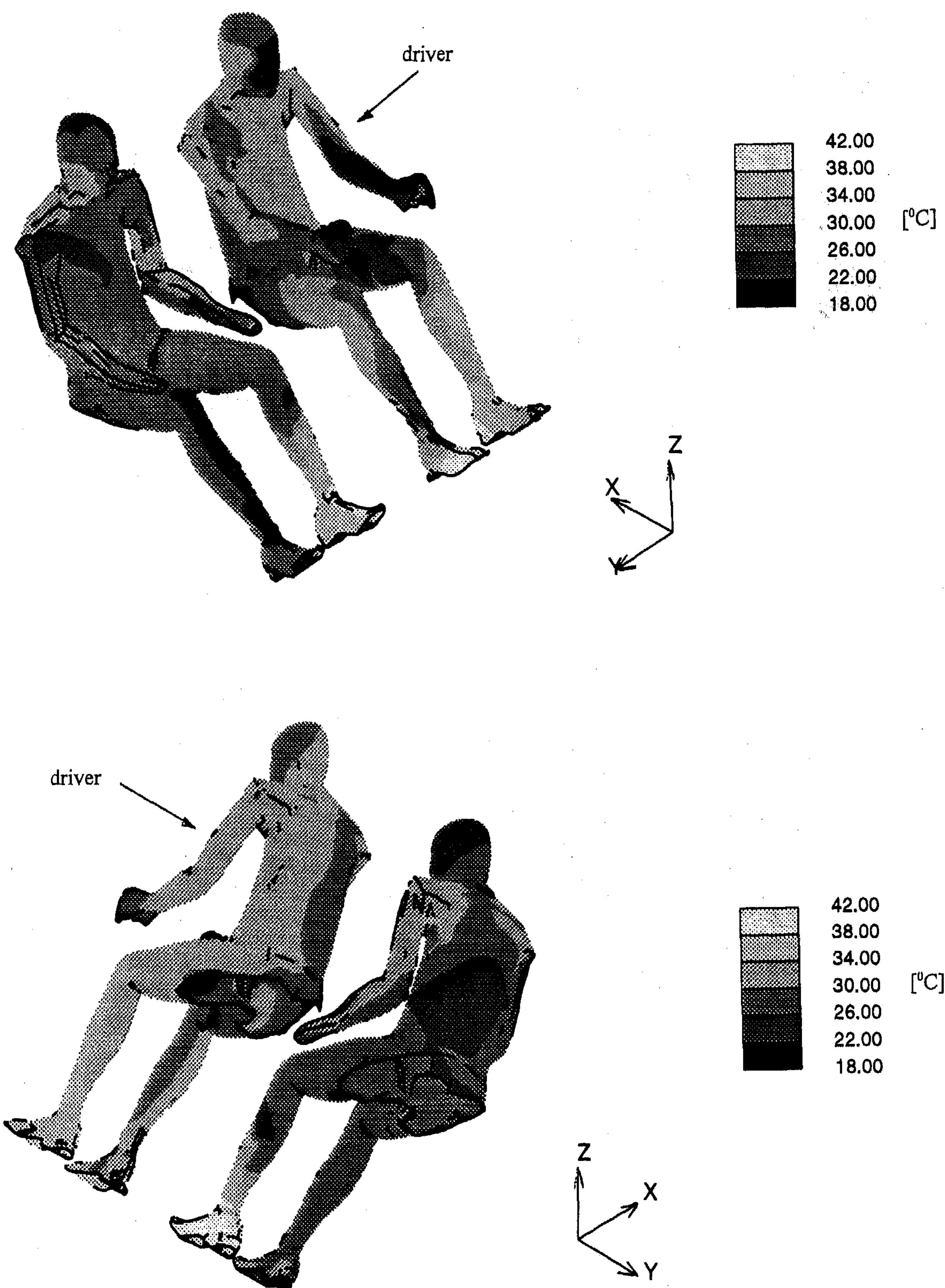
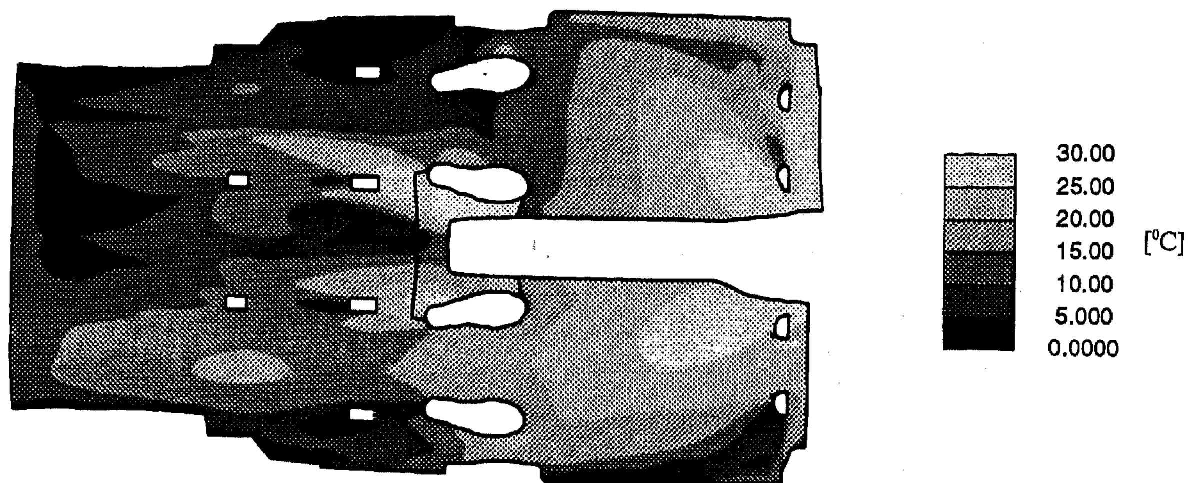
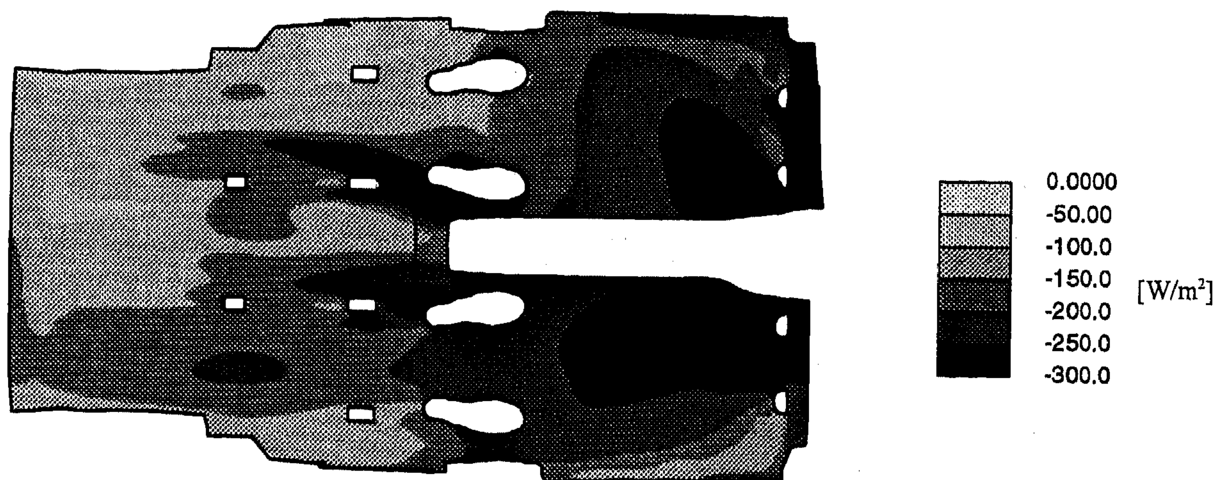


Figure 7: Air temperature in the vicinity of the frontal occupants (2 views)

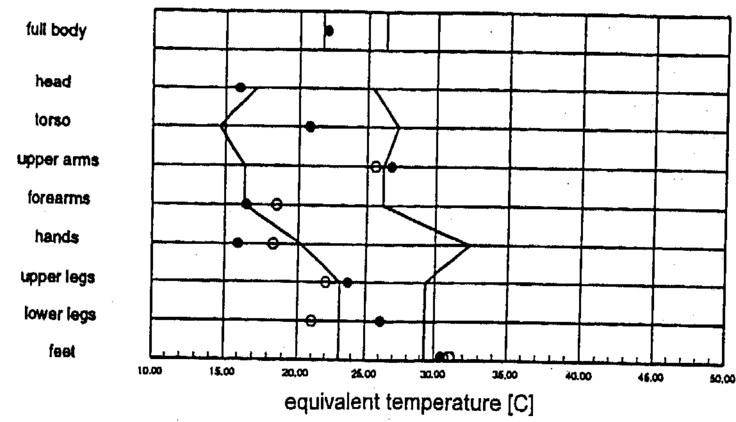


a) temperature

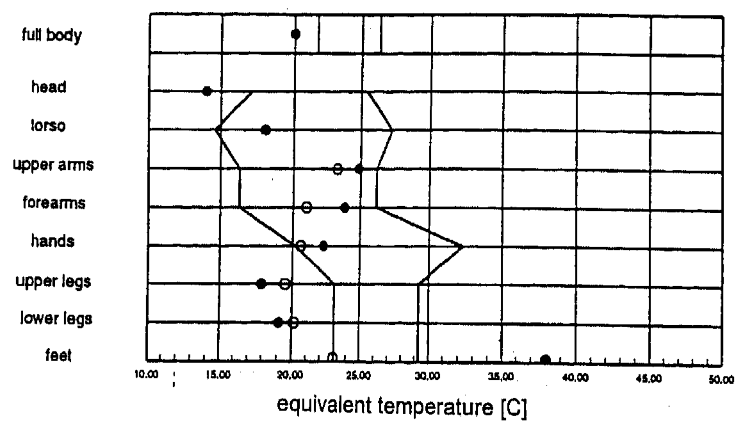


b) heat flux

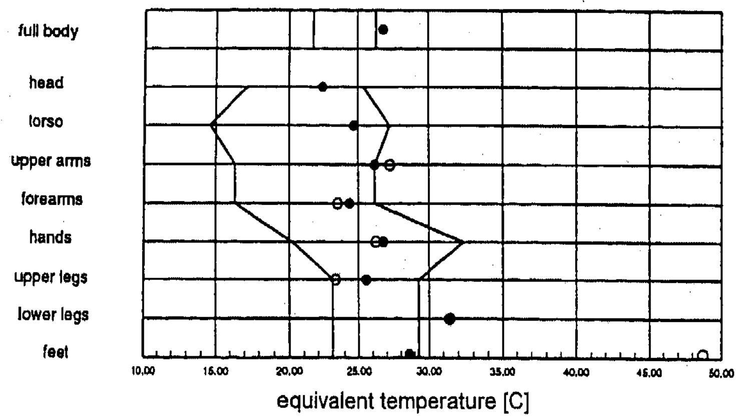
Figure 8: Temperature and heat flux distribution at the floor



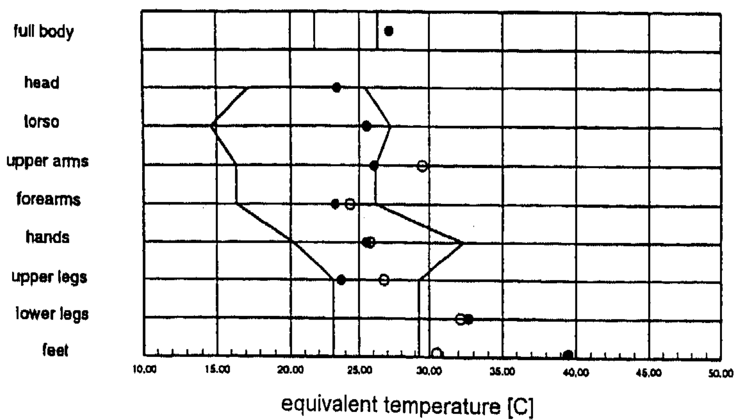
driver



passenger



rear passenger
(driver's side)



rear passenger
(pass. side)

● left
○ right

Figure 9: Thermal comfort diagrams for all 4 occupants

970547

Measurement of Rubber Elasticity and Correlation with Seal Life

Boris Dinzburg
Chicago Rawhide

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Two new methods will be presented to measure rubber and seal elasticity - Recovery from Bending (RFB) and Loss of Seal Interference (LOSI). It will be demonstrated, the kinetics of rubber elasticity change during immersion in oil with aeration and oil change for different rubber compounds as NBR, HNBR, ACM, AEM, FKM and VMQ. Using the kinetics of Recovery From Bending, the estimation of seal life can be done.

INTRODUCTION

To find a correlation between rubber properties of elastomeric radial lip seal life, it is necessary to know the sealing mechanism. In this case, it is possible to select rubber characteristics which are responsible for seal life and investigate how these properties change in the application environment. Currently, it is possible to formulate the basic seal properties which are responsible for seal life. One of the most important seal properties is seal lip **follow-ability**, or an ability of a seal to follow the shaft wobbling due to the eccentricity caused by Dynamic Run-out (DRO), Shaft-to-Bore Misalignment (STBM), or roughness of the shaft surface. The second factor is seal **pumping ability** - an ability of a seal to pump oil from the air side to the oil side. The third factor is **resistance to wear**, or more importantly, resistance to lubricated wear. A fourth factor is **resistance to swelling** in oil; and a fifth factor is the appearance of **defects at the contact point** during seal usage such as carbonization, blisters, cracks, stick-slip, etc. which are not due to the factors mentioned above. This presentation is dedicated to investigating the influence of rubber properties which are responsible for seal lip follow-ability as a main factor of seal leakage.

BACKGROUND

The first publication about the importance of seal follow-ability was done by Akaoka [1] in 1957. The experiment was conducted on a torque instrument to measure the influence of speed, torque, shaft eccentricity and spring force magnitude. He came to a conclusion that to maintain seal performance, no gap should be between the shaft surface and the seal lip. The material has to provide fast elastic recovery speed. According to Akaoka, the spring has no effect on wobbling action, in fact, a stiff spring has a negative effect.

Later, in 1977, Akaoka and Nitani [2] derived a mathematical model which includes seal parameters, rubber elasticity, and the gap between seal and shaft. They divided the seal into three elements; lip head, lip arm and spring. After simple calculations, they connected all three elements into one mathematical model. From their paper, it is not clear if the mathematical model fits with the experimental data, but the principal statement is correct, "...a gap occurs when the displacement velocity of the seal lip cannot follow the displacement velocity of the shaft." In 1982, Akaoka and Nitani [3] modified the instrument described in [2]. They changed the relative displacement magnitude and the displacement velocity measured by a strain gauge attached to the spring and recorded by an oscilloscope. Displacement of the seal lip was read from the film taken by an optical camera.

The final conclusion drawn from this very accurate experiment is that seal follow-ability is greatly affected by the viscoelastic characteristics of rubber material. The recovery is very fast and the material is quite elastic during the initial process of recovery. The recovery becomes quite slow in the viscoelastic area which takes a long time to recover. Elastic material properties are more effective for seal follow-ability than are viscous properties.

Theoretical and experimental descriptions of dynamic behavior of the seal lip were also made by Prati [4,5]. His experimental study was performed on a device with a static seal and a vibrator. The viscoelastic seal characteristics are determined by applying an oscillatory motion. The loss of contact is evidenced by optical fiber light passing through the narrow gap between the seal and the shaft. The passing light was visualized by a photo detector and oscilloscope.

According to Prati, the dynamic eccentricity produces the oscillatory motion of the seal lip. As the frequency is increased, the seal lip is unable to follow the radial shaft displacement and the lip lifts from the shaft and oscillatory clearance occurs between the seal lip and the

shaft. Horve [6], together with Prati, conducted experiments with seals made from compounds based on NBR, FKM and VMQ rubbers which illustrated the ability of Prati's device to measure seal follow-ability. Horve also made some calculations to fit the experimental data to the analytical models [7]. The conclusion was made that this method can be used to estimate seal follow-ability.

In 1989, Mims [8] created an optical device to measure the oscillatory clearance with rotational motion. The equation was derived which correlated the theory with the experiment. According to Mims, the stress relaxation or creep measurement has to be used to calculate the loss of follow-ability. The silicone compound demonstrated the best follow-ability.

The investigations described above (1-8) were conducted in dry conditions without confirming results of calculation and measurement with actual oil seal leakage. Also, these investigations did not take into account an interaction between oil and rubber.

According to Stakenborg [9], the clearance appearing between the shaft and the seal lip is influenced not only by the rubber viscous properties, but also by the viscoelastohydrodynamic effect of nonuniform fluid film which was created between the seal lip and the shaft.

In Russia, calculations were made to predict oil leakage through the oscillatory clearance when rubber elasticity or recovery was altered by changing the temperature [10]. The lower the temperature was, the lower was the elasticity and the recovery coefficient diminished. The oil leakage was calculated using the modified equation for fluid flow through the static clearance. Experimental data and calculations show a satisfactory correlation.

In spite of the complicated correlation, the experimental data fits very well with the theoretical prediction and provides an ability to use the coefficient of recovery for calculation of the temperature limit of seal leakage. It proves the theoretical statement of correlation between the rate of rubber recovery and the clearance between the seal lip and the shaft..

Works cited here represent only a small portion of the investigations performed in this area. Yet, they once more prove the importance of seal follow-ability or the ability of a seal to maintain a certain extent of elasticity during operational conditions and protecting against seal leakage.

This work tries to address the question of what kind of rubber properties are responsible for seal follow-ability, and what method should be used to measure the loss of elasticity.

There are many different methods for measuring seal elasticity. One of the most popular in the seal industry is to measure elongation change of s specimen during immersion in oil which to some extent represents elasticity, however this method did not show satisfactory correlation with seal life, even when aging times were from 1,000 to 2,000 hours.

A modification of this method was developed by Chicago Rawhide in which the kinetics of elongation loss is measured [11]. The empirical equation was derived which provides a satisfactory correlation of elongation change and seal life during rubber immersion in different oils, temperatures, and aging times.

$$\frac{E_0}{E_t} = \exp(b\,t^{1/2}) \qquad (1)$$

where E_0 = Initial ultimate elongation, %
E_t =Ultimate elongation after aging, %
b =Regression index
t =Aging time, hours

Using Equation (1), straight lines are obtained when ln (E_0/E_t) versus $t^{1/2}$ is plotted. The slope of such a straight line "b" characterizes the kinetics of aging. The Heat Resistance Index (HRI) was calculated as:

$$HRI-e = \frac{1}{b}\ x\ 10^{-1} \qquad (2)$$

This method has some shortcomings. First of all, Equation (1) does not provide good correlation for materials with high heat resistance like FKM and, in some cases, HNBR. For high heat resistance materials, the ultimate elongation did not change much during a long period of aging. Secondly, this method cannot be used for silicone, and sometimes for polyacrylic compounds due to an elongation increase during aging in oil. In spite of these shortcomings, this method is being successfully used to develop some new compounds with high heat resistance, and also to compare aggressiveness of different oil types.

The method of evaluating HRI using the elongation change utilizes the elongation at break. The value of ultimate elongation can reach very high levels, but in seal operating conditions, the deformation does not exceed 5 to 10%. Because the difference between "large" and "small" deformation of rubber is well known, a method was created to measure the stiffness change in bending at small deformation [12]. Using the kinetics of stiffness change, an HRI-s equation analogous to the equation for elongation change can be calculated. This method provides better correlation with seal life [13], but also has the same shortcomings as the method when ultimate elongation change was used. FKM, some HNBR compounds, and VMQ did not show visible change in stiffness.

All methods described above were static methods of measuring the change of rubber elasticity.

Bjork, Dickman and Stenberg from the Sweden Royal Institute of Technology [14], have undertaken a long term study to measure dynamic rubber properties and to correlate them with sealing performance. Cylindrical buttons of equal height and diameter were tested in air and oil for up to 1000 hours. The minimum and the maximum forces were recorded at different temperatures (up to 155°C), frequency (up to 32 Hz), amplitude (up to 10%) and pre-strain (up to 30%). Continuous and intermittent stress relaxation were measured to correlate with dynamic stress relaxation. The experimental results showed that

intermittent and continuous stress relaxation gives only an approximation of the dynamic stress relaxation. However, continuous and intermittent stress relaxation are important characteristics of dynamic seal life. The real seal life was not measured, but the method of dynamic stress relaxation could be very valuable for investigation of rubber properties change during immersion in oil. The experimental equipment required use of only 4 buttons for simultaneous testing and took a long time, up to 1000 hours which did not give a lot of advantages against a regular seal bench test.

EXPERIMENTAL

SAMPLE PREPARATION - The compounds were mixed in the laboratory internal K-1 mixer. The slabs were cured in an electrical press at 177±2°C (350 ±5°F) at t_{90}. The ACM, AEM, VMQ and FKM compounds were post cured according to a standard procedure.

Our first attempt was to measure the classic dynamic rubber characteristics and to determine how these characteristics would change during aging in hot oil. The results did not show much change of dynamic characteristics for the FKM compound when the compound was aged in the oil up to 1,500 hours [25].

Analyzing the dynamic properties change during aging in oil and the seal follow-ability phenomena, an assumption was made that the most important rubber characteristic is the process of rubber recovery. How it was demonstrated in the background section, this assumption is not new and has been investigated previously [7-8].

For this research, the effort was directed to measure not the stress, but the strain relaxation. It is necessary to remember that elasticity is an ability of rubber to sustain large deformation and to come back when the load is removed. The importance of strain relaxation for the seal follow-ability was proven by a special research project jointly made with the University of Akron [15].

Because the seal rotation speed is very high and can reach 8,000 rpm or 0.0075 seconds per one rotation, a special apparatus was designed and fabricated to measure recoverable strain after imposing a constant compressive strain on the rubber sample ("Restrain"). The apparatus can measure the relaxation of rubber button beginning from 0.0067 seconds when the external load was removed. The distance between the top rubber button and the planger head was measured by a Hall effect sensor. Recoverable strain (RS) of a rubber sample was calculated according to the following equation:

$$\epsilon = \frac{h_t - h_1}{h_o - h_1} = \frac{\Delta h}{\Delta h_o}. \quad (3)$$

where ϵ = recoverable strain,
h_t = current height of the sample during recovery,
h_1 = height of samples after 20% of compression,
h_o = initial height of sample.

Using the results of this investigation, some important conclusions can be reached. First, the recoverable strain for HNBR and even for FKM compound changed during aging rubber in air or in oil. Recoverable strain measurement is a lot more sensitive than elongation and stiffness change and dynamic stress relaxation. Second, the comparison of the recovery for FKM and HNBR rubbers did not show a big difference between results taken after 0.02 seconds of relaxation time and one hour, even though the relaxation times differ in a factor of 180,000 (Fig. 1).

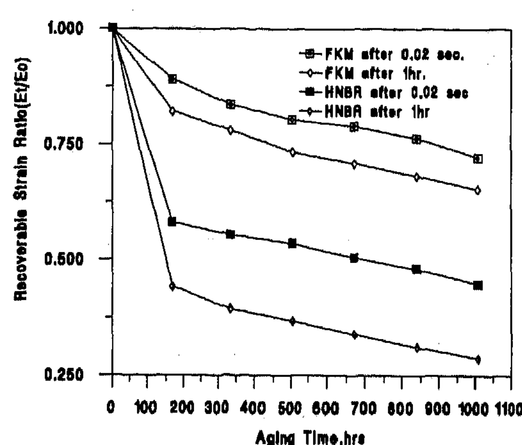


Fig. 1 - Comparison recovery after 0.02 seconds and 1 hour (3,600 seconds).

This experiment leads to the conclusion that for technical applications it is possible to compare different compounds after relaxation time or time of recovery longer than a fraction of a second, which does not require complicated instrumentation.

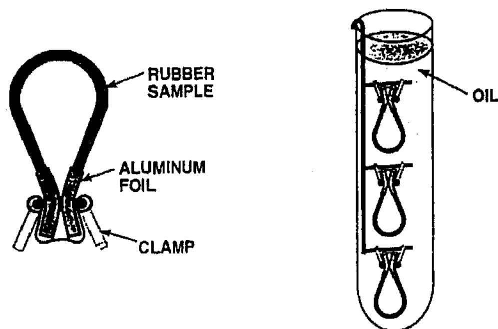
In the present paper, two new methods were introduced which are able to measure strain set deformation after "long" periods of time. One is "Recovery From Bending" (RFB) which can be done on rubber samples, and the second is "Loss of Seal Interference" (LOSI) which can be done on seals.

Bending deformation is the predominant type of deformation which is created in the flex section of the radial lip seal during shaft rotation. Resistance of rubber to bending deformation is one of the most important rubber properties which is responsible for seal elasticity or seal follow-ability. The principle of this method was taken from the SKF Engineering Research Center method of "Permanent Set Measurement".

RECOVERY FROM BENDING

PROCEDURE - A strip 100 mm long, 2 mm thick and 10 mm width was bent and clamped with paper clamps. The clamped sample is immersed in oil as shown in Fig. 2 at the required temperature.

RECOVERY FROM BENDING



MEASUREMENT AND CALCULATION

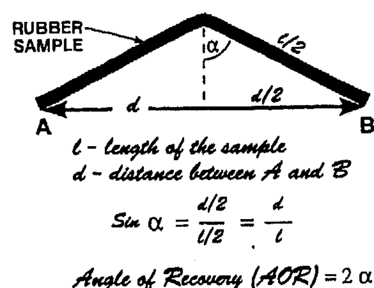


Fig. 2 - Scheme of procedure to measure Recovery From Bending (RFB).

After aging, the sample is cleaned and released from the clamp. After 15 minutes of recovery at ambient temperature, the distance between the end points of the sample are measured. The sample is then placed on the preheated metal plate and put in the preheated oven for 15 minutes at 212°F. After 15 minutes, the distance between the end points is measured while the plate is still hot.

Recovery From Bending (RFB) is calculated as the ratio:

$$R = \frac{l}{d} \tag{4}$$

where d = the distance between the end points of the sample,

l = the length of the sample after immersion in oil and recovery.

The 1/R also represents the sine of recoverable angle (α)

$$\frac{1}{R} = \sin\alpha \tag{5}$$

The Angle of Recovery is 2α. The complete recovery is 180°.

When the strip is bent, the stress is concentrated on the top of the loop. The FEA stress comparison on the RFB loop with the stress on the seal when the seal is installed on the shaft shows the good correlation.

It is very important that the recoverable angle is measured not at ambient temperature, but at the hot temperature which better represents the real temperature of seal usage. Table 1 demonstrates the difference between the recoverable angle measured at ambient temperature and at 212°F. For some seal compounds the difference in percentage of recovery is about two times.

Table 1
Comparison of the angle of recovery after bending for samples swelled at 168 hours at ambient temperature

Compound	After 15 minutes of recovery at ambient temperature		After additional heating at 212°F for 15 minutes	
	Angle of Recovery (°)	% Recovery	Angle of Recovery (°)	% Recovery
NBR	130.6	72.5	171.8	95.4
HNBR	93.5	51.9	171.8	95.4
ACM	124.6	69.2	168.4	93.5
VMQ	148.3	82.3	165.8	92.1
FKM	104.8	58.2	168.4	93.5

The measurement of the Recovery From Bending has an advantage in the accuracy with comparison to the measurement of elongation change after aging in oil.

The RFB demonstrates high reproducibility and less variation. The comparison of these two methods for HNBR after aging in oil for 70 hours at 302°F (ten samples were tested) shows that elongation has a variation coefficient of 12.7%, and the RFB is 1.7%. The absolute variation for elongation was 220% ±40% and for RFB the angle of recovery was 75 ± 2°.

All RFB tests were made after immersion in oil with a small amount of interrupted aeration (2 ml per min. and 3 hours on and 21 hours off) and oil change after 168 hours. The conditions of interrupted aeration was previously investigated and correlated with engine field conditions for passenger cars. The influence of aeration was published in [16, 17, 18].

INFLUENCE OF VOLUME SWELL - Volume swell is a very important rubber property which is responsible for seal life. Usually when the volume swell increases, the elongation diminished and, in a troubleshooting

situation, it is very difficult to determine the cause of seal leakage. A seal can leak due to rubber swell or loss of elasticity.

To verify the influence of volume swell on elasticity change, Recovery From Bending was measured. Three heat resistant compounds: FKM, VMQ and HNBR were tested at a relatively low temperature, 212°F for 70 hours to avoid aging. This condition normally does not change elongation, but is enough to reach maximum swell. Different oils were selected to reach various volume swell.

Test results for Recovery From Bending (Fig. 3) demonstrates that swelling in oil did not diminish rubber elasticity and did not reduce the seal lip follow-ability.

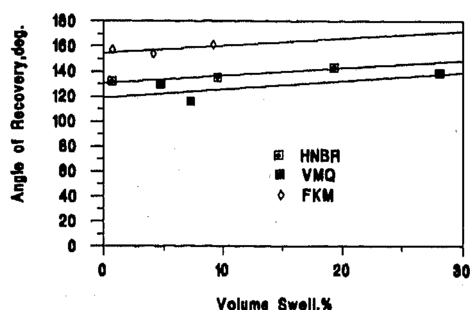


Fig. 3 - Influence of volume swell on Recovery From Bending.

This very important advantage of Recovery From Bending shows the ability to separate the cause of seal leakage due to loss of elasticity or volume swell. When a seal leaks and rubber has high volume swell but Recovery From Bending did not change, the conclusion could be reached that the cause of leakage was not the elasticity loss but different factors. For instance, high swell could deform the lip or change the radial load. This experiment could be used as an explanation for the reason why seals work properly in spite of high volume swell up to 30% because the seals do not lose their elasticity and proper seal design guarantees the necessary radial load and absence of lip deformation.

Fig. 4 shows the angle of recovery for nine tested compounds.

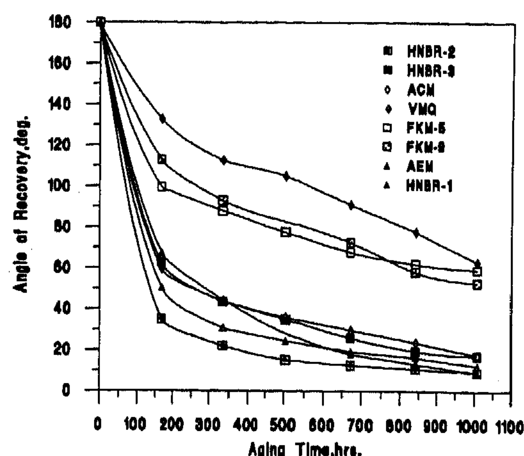


Fig. 4 - Angle of recovery after Recovery From Bending.

RFB is able to measure the change in elasticity for all tested compounds. The silicone compound has the higher elasticity, the FKM compounds are the second highest, the NBR are the lowest, and HNBR, ACM and AEM belong to the middle group. This statement is confirmed in Fig. 5 where the Heat Resistance Index from Recovery From Bending (HRI-r) is calculated using the equations (4) and (5) with replacement of the elongation on the rate of Recovery From Bending.

$$R = \exp(-bt^{\frac{1}{2}}) \qquad (6)$$

and

$$HRI\text{-}r = -\frac{1}{b} \qquad (7)$$

where R = Recovery From Bending,
t = Aging time, hours
b = Coefficient
HRI-r = Heat Resistance Index (calculated from the kinetics of Recovery From Bending).

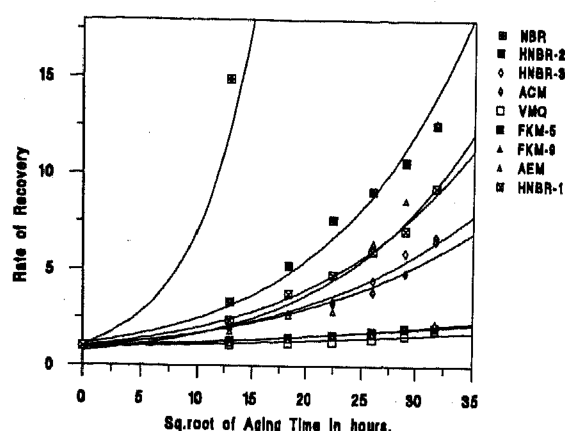


Fig. 5 -Kinetic of recovery change (with aeration and oil change).

The correlation coefficient for RFB is higher than the one for elongation change. The comparison of correlation coefficients for the Heat Resistance Index indexes is shown in Table 2.

Table 2

Comparison of the correlation coefficient for Heat Resistance Index, calculated from kinetic of elongation and Recovery from Bending change

Compound	Elongation	RFB
NBR	0.948	0.991
HNBR-2	0.915	0.987
HNBR-3	0.915	0.990
ACM	0.921	0.989
FKM-5	0.957	0.993
FKM-9	0.889	0.942
AEM	0.987	0.905
HNBR	0.871	0.997
VMQ	—	0.831

In the previous paper [25], the correlation was found between HRI-r, calculated from immersion test with aeration and without oil change, and seals tested in monograde oil SAE-30 at 275°F (135°C) and 2165 rpm. The aim of the present experiment is to find the correlation in more severe conditions using an immersion test with aeration and oil change.

To find a correlation between Recovery From Bending and seal life, 3" CRW seals were molded using the same compounds which were tested above. The seals were tested in the bench test machine with automatic speed and temperature controls. The following parameters were used:

1. Oil: SAE 10w30 multigrade oil (same oil used for immersion test)
2. Temperature: 302°F (150°C)
3. Oil Change: Every 168 hours
4. Speed: 3600 rpm
5. Conditions: First day - 1 hour at 0 rpm and ambient temperature, 9 hours at 1500 rpm, 10 hours at 3600 rpm, 4 hours at 0 rpm. Next days - 20 hours on and 4 hours off.
6. At least 6 seals were tested for each compound.
7. DRO: 0.010" (0.254mm) TIR
8. STBM: 0.010" (0.254mm) TIR

The seal test results and Heat Resistance Index, calculated from RFB are shown in Table 3.

Table 3

Compound	Characteristic Life		HRI-r
	Life, hours	Wiebull Slope	
NBR	90	3.1	5.3
HNBR-2	614	4.4	12.4
ACM	651	2.8	17.7
VMQ	1022	3.2	53.3
FKM-5	2969	3.6	44.0
FKM-9	3349	14.3	38.7
AEM	427	7.4	12.5
HNBR-1	690	11.0	14.4

Compound HNBR-3 was not tested because it was designed for reciprocating applications.

The correlation of HRI-r and seal life is shown in Fig. 6. The correlation coefficient is 0.948.

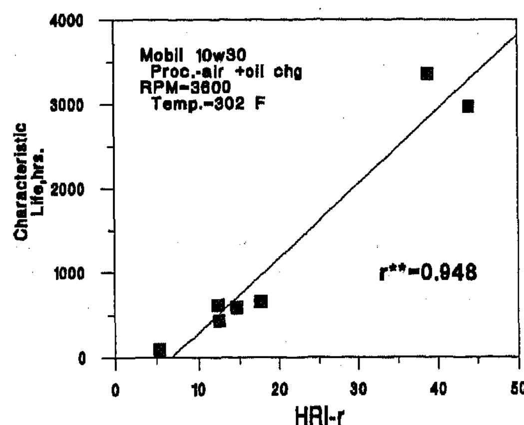


Fig. 6 - Correlation HRI-r with seal life.

This correlation coefficient was reached without excluding the silicone rubber compound VMQ because the failure mode for most of the silicone seals in test conditions was a breakdown in the flex cross-section area. An attempt to use the Seal Life Index (SLI) which was introduced in the previous publication [25] and is HRI-r divided by the Lubricated Wear Index (LWI) did not improve the situation. This experiment demonstrates that the SLI can be used only if the main cause of the leakage is loss of follow-ability and lubricated wear.

LOSS OF SEAL INTERFERENCE (LOSI)

Recovery From Bending provides an ability to test the rubber elasticity change for compounds immersed in different oils. It also helps to select the compound which works best in the required conditions for a given seal application. Seal follow-ability depends not only on rubber elasticity, and resistance of rubber to aging, but on other seal parameters as well.

It is well known that follow-ability changes when the radial load changes. Radial load could change due to a change in interference between the seal and the shaft, change of rubber stiffness or a spring size. To investigate the influence of the factors mentioned above, the Loss of Seal Interference method was designed. The principle of this method is to measure residual deformation after removing the seal from the shaft. The set of deformation indicates loss of elasticity in the same way as the rubber compression set and can be calculated using the following equation:

$$LOSI, \% = \frac{d_a - d_b}{d_m - d_b} \, x \, 100. \qquad (8)$$

where d_b = inside seal diameter before test
d_a = inside seal diameter after test
d_m = diameter of shaft or mandrel

The device used to measure LOSI is shown in Fig. 8. It consists of a holding fixture (Fig. 7A) and a metal container with oil which is installed on a heater with a magnetic stirrer (Fig. 7B).



Fig. 7 - Device used to measure the Loss of Seal Interference (LOSI).

Oil is aerated using a tank of compressed air. The aeration in this experiment was set up on the same oil degradation level as for the rubber immersion test in a glass test tube. Previously the aeration was established on the same level as in a passenger car engine under field conditions [17].

To process the test, the seals were put on the mandrel (Fig. 7A bottom) and inserted into the holding fixture. The holding fixture was dipped in the container of oil. Every 168 hours, the fixture was taken from the oil, the seals were taken off the mandrel and cleaned. The inside diameter and radial load were measured, and the seals were put back in the holding fixture and dipped in a container with fresh oil. The same nine compounds were tested which were described in the previous section. The test results are demonstrated in Figs. 8 and 9.

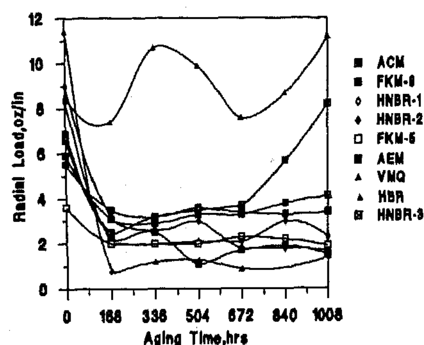


Fig. 8 - Radial load change after immersing seal on mandrel on LOSI device.

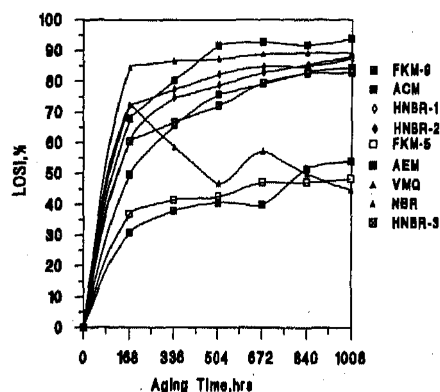


Fig. 9 - LOSI change after immersion in SAE 10w30 at 302°F (150°C) with aeration and oil change.

Fig. 8 shows that the radial load dropped dramatically after 168 hours of aging but did not change a lot for most compounds after that. Only the HNBR-1 compound after reduction of the radial load demonstrated an increase after 672 hours of aging. The NBR compound cannot be taken into account because the seals got hard and cracked at the used aging conditions. The Loss of Seal Interference (Fig. 9) demonstrates a much clearer picture. All seals have the same kinetics of LOSI change, the only exception is for the silicone compound.

The kinetics of LOSI change satisfactorily fits the following logarithmic equation:

$$L = a + b\, Ln\, (T^{1/2}) \qquad (9)$$

where L = LOSI, rate
T = aging time, hours
a,b = coefficients

The Heat Resistance Index from the kinetics of Loss of Seal Interference (HRI-i) can be calculated as:

$$HRI\text{-}i = \frac{1}{b} \qquad (10)$$

The HRI-i correlates well with the seal life (with correlation coefficient 0.984 (Fig. 10) which proves an ability to use the Loss of Seal Interference method not only for compound evaluation, but also to design and select optimum seal parameters.

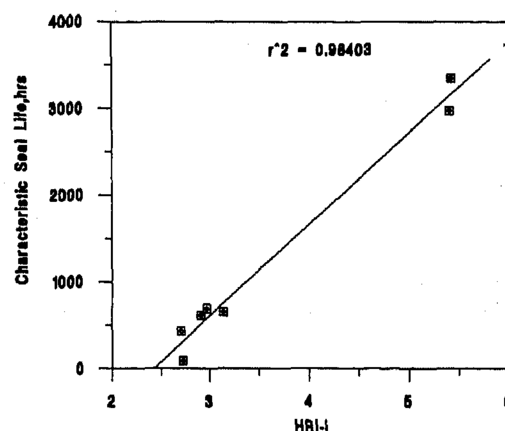


Fig. 10 - Correlation HRI-I with seal life.

The LOSI correlates also with Recovery From Bending (Fig. 11) which demonstrates the same character of elasticity change for slabs using RFB and seals using LOSI. It is necessary to point out that data obtained for the LOSI and RFB methods fit to one master curve with the exception of one AEM compound, although the AEM compound also shows satisfactory correlation RFB with LOSI.

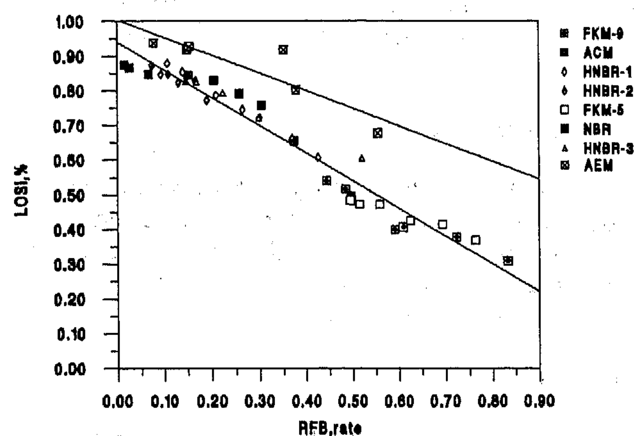


Fig. 11 - Recovery From Bending vs. Loss of Seal Interference.

DISCUSSION

The most important factors which are responsible for dynamic shaft seal life are seal design, type of elastomer compound, and testing conditions including type of oil, underlip temperature and shaft roughness. Even with the correct seal design and shaft roughness, the properties of the elastomer compound can provide the most valuable influence on seal life.

The elasticity of the elastomer compound provides the dynamic shaft seal with the ability to follow the shaft eccentricity and shaft roughness which is called follow-ability. Additional properties such as resistance to lubricated wear, resistance to swell in oil, resistance to create defects on the seal lip area like carbonization, etc., and pumping ability or the ability of the seal to pump oil from the air side to the oil side are the most important properties responsible for seal life.

In spite of many factors which influence the seal life, one of the most important is seal follow-ability. When rubber loses a certain amount of elasticity and the lip cannot follow the shaft, a gap is created between the lip and shaft, and this could be a cause of seal leakage. This raises a question of the size of the clearance causing seal leakage. According of Akoaka and Nitani [3], the occurrence of leakage is not determined by the gap alone. The leakage is affected by different internal and external parameters like oil pressure, viscosity and dynamic pressure. Shimoi [19] also demonstrated that it is possible to completely avoid seal leakage when a small gap occurs between the shaft and the lip, even if the oil viscosity is low.

This phenomena can be explained with pumping effect. If clearance is not very large, the oil could pump from the air side to the oil side and prevent seal leakage. The pumping effect was discovered by Jagger in 1957 [20]. Later, many investigations were conducted which introduced different theories. The historical reviews connected with the pumping mechanism are well known and broadly published [21 - 23].

Currently, it is well established that in most cases if a rubber lip seal is not able to pump oil from the air side to the oil side, the seal will instantly leak. Although there is no direct correlation between pumping rate or pump index with seal life, the pumping rate does improve service reliability [24]. The seal with high pumping ability could compensate the leakage which occurs due to the loss of elasticity or small rate of recovery.

In this paper, two methods are introduced that can provide the ability to select, create, or compare different compounds, and analyze the aggressiveness of oil and the seal design. The first method is Recovery From Bending (RFB) and the second is Loss of Seal Interference (LOSI). Using these two methods, it is possible to reduce the time for the seal bench testing and also to estimate the seal life. Fig. 12 shows a correlation of the Heat Resistance Index for different rubber compounds tested in different oils and different test conditions with seal life. One set of seals (#1) tested in SAE multigrade oil 10w30 at 302°F (150°C) and 3600 rpm with oil change every 168 hours (one week) (Fig. 2). The second set (#2) which was described in the previous paper [25] was tested in less aggressive conditions in monograde oil, SAE- 30 at 275°F (135°C) and 2165 rpm without oil change. The first curve fit into the equation:

$$Y = 544.1 + 87.6X \qquad r^2 = 0.959$$

The second fits into another equation

$$Y = 310.5 + 91.3X \qquad r^2 = 0.899$$

where Y = Seal life in hours
X = Heat Resistance Index calculated from kinetic of Recovery From Bending in suitable oil immersion test.

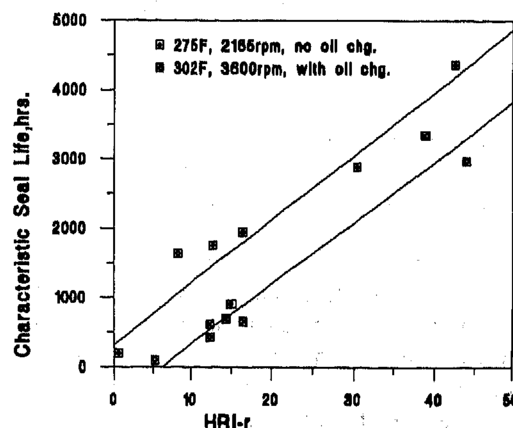


Fig. 12 - Correlation HRI-r with seal life.

Using an assumption that an average speed of a passenger car is about 35 miles per hour and the total mileage per year is about 12,000 miles, the following seal life prediction can be done:

Table 4
Estimation of Seal Life

Compound	Set #1 302°F, 3600 rpm, with oil change		
	HRI-r	Calculated Seal Life, Hrs	Seal Life Prediction, Yrs
NBR	5.3	-79	0
HNBR-2	12.4	542	1.6
ACM	16.3	883	2.6
FKM-5	44.0	3,310	9.7
FKM-9	38.8	2,854	8.3
AEM	12.5	551	1.6
HNBR-1	14.4	717	2.1
Compound	**Set #2 275°F, 2165 rpm, no oil change**		
	HRI-r	Calculated Seal Life, Hrs	Seal Life Prediction, Yrs
NBR	0.7	374	1.1
HNBR-2	8.3	1,068	3.1
ACM	16.3	1,799	5.2
FKM-5	42.5	4,190	12.2
FKM-9	30.4	3,086	9.0
AEM	15.0	1,680	4.9
HNBR-1	12.8	1,479	4.3

It is necessary to take into account that bench test conditions are very severe (20 hours on, 4 hours off). In real field conditions, an average passenger car runs no more than 1 to 2 hours per day and 22 - 23 hours in a static condition. In this case, the predicted seal life can be named as "minimum seal life in extreme conditions".

SUMMARY

Although the sealing mechanism is complicated and many factors can be the cause of seal leakage, one of the most important properties responsible for seal life is seal lip follow-ability. Follow-ability has a direct correlation with rubber elasticity - "the ability of rubber to sustain a large deformation without breaking and coming back to its original configuration when the deforming stress is removed." The Recovery from Bending and Loss of Seal Interference methods can be used to measure the change of elasticity during aging rubber or seals in oil. These methods, to a higher extent, reproduce the seal stress tension in the seal flex area; they do not require special equipment, provide high reproducibility and low variation, and finally, they are easy to use.

Recovery From Bending provides an ability to estimate the seal life and Loss of Seal Interference, to select the seal design, seal interference, spring size, and seal design.

ACKNOWLEDGMENTS

The author would like to gratefully acknowledge Alex Berdichevsky for his helpful discussion and comments in the preparation of this paper. Chicago Rawhide is also thanked for their permission in the publication of this presentation.

REFERENCES

1. Akaoka. J., "Effect of Wobbling or Jerking of Shaft on the Sealing Performance of an Oil Seal", Lubrication 2,6,273 (1957), Japan.

2. Akaoka J., Nitani A., "On the Following Property of the Oil Seal Lips (First Report)", Lubrication 22, 46 (1977).

3. Akaoka J., Nitani A., "The Following Property of the Oil Seal Lips (Second Report)", Lubrication 27, 6 (1982).

4. Prati E., "Dynamic Behavior of Radial Lip Seal Under Effect of Shaft Eccentricity", 10th International Conference on Fluid Sealing, Innsbruck, Austria, 1984.

5. Prati E., "A Theoretical-Experimental Method for Analyzing the Dynamic Behavior of Elastomeric Lip Seals", 129th Meeting Rubber Division, A.C.S., New York, 1986. Rubber Chemistry and Technology 60, 176-189, 1987.

6. Horve L., "Radial Lip Seal Dynamic Behavior", (University of Bologna), Internal Chicago Rawhide Report, No. 208, 1986.

7. Horve L., "Radial Lip Seal Dynamic Behavior", (University of Bologna), Internal Chicago Rawhide Report, No. 209, 1986.

8. Mims S., "Dynamic Behavior of Radial Lip Seals (A Followability Study)", Thesis of M.D., Armour College, Illinois Institute of Technology, 1988.

9. Stakenborg M.L., "On the Sealing and Lubrication Mechanism of Radial Lip Seals", Ph.D. Thesis, Eindhoven University of Technology, 1988.

10. Malanichev V.I., Chercky I.N., "Investigation of Workability of Rubber Seal at Low Temperatures", Kauchuk i Rezina 5, 24-26, 1984, Russian.

11. Dinzburg B.N., Keller R.W., Bond R., "Heat Resistance Evaluation of Rubber Compounds", Rubber World 1975, 28-37, 1988; Gummi Fasern Kunstoffe, 41, 8, 398-404, 1988.

12. Dinzburg B.N., U.S.A. Patent 5,178,017 (1993).

13. Dinzburg B.N., Bond R., "Stiffness Measurement for Evaluating Heat Resistance of Rubber Compounds", Rubber World 201, 4, 20-24, 1990.

14. Björk F., Dickman D., Stenberg B., "Long Term Studies of Rubber Materials by Dynamic Mechanical Stress Relaxation", Rubber Chemistry and Technology 62, 3, 387-425, 1989.

15. Isayev A.I., D.Cao, Dinzburg B.N., "Elastic Recovery of Rubber Vulcanizates at Very Short Times", Paper #47, Rubber Division ACS, Cleveland, OH, October, 1995.

16. Dinzburg B.N., "Investigation of the Effect of Aeration on Automotive Oils and Rubber Compounds", 142nd ACS Meeting, November, 1992, Paper #89.

17. Dinzburg B.N., "Influence of Oil Aeration on Dynamic Shaft Seals Under Laboratory and Field Conditions", SAE Paper 930535, 1993.

18. Dinzburg B.N., "Influence of Oil Additives on FKM and HNBR Compounds", Lubrication Engineering, 1995 (to be published).

19. Shimoi Y., "One Experiment Concerning the Sealing Mechanism of an Oil Seal", Nihon Dust Keeper, 1-5, Ginza, Chuo-Ku, Tokyo, 1962, Japan.

20. Jagger E.T., "Study of the Lubrication of Synthetic Rubber Rotary Shaft Seals", Proc. Conf. Lubric. Wear, 409, 1957.

21. Müller H.K., "Concepts of Seal Mechanism of Rubber Lip Type Rotary Shaft Seals", Proc. 11th Conference on Fluid Sealing BHRA, Paper K-1, 1987.

22. Horve L., "A Macroscopic View of the Sealing Phenomenon for Radial Lip Oil Seals", Proc. 11th Conference on Fluid Sealing BHRA, Paper F-2, 1987.

23. Johnston D.E., "Theoretical Analysis of a Pumping Mechanism Occurring between Relatively Moving Surfaces", 12th International Conference on Fluid Sealing, Paper H-3, 1989.

24 Horve L., "The Correlation of Rotary Shaft Radial Lip Seal Service Reliability and Pumping Ability to Wear Track Roughness and Microasperity Formation", SAE Paper 910530, 1991.

25. Dinzburg, B., "Rubber Elasticity - the Main Characteristic for Dynamic Shaft Seals", Paper #71, Rubber Division, American Chemical Society, October, 1995.

FIGURE CAPTIONS

Fig. 1 Comparison recovery after 0.02 seconds and 1 hour (3,600 seconds).

Fig. 2 Scheme of procedure to measure Recovery From Bending (RFB).

Fig. 3 Influence of volume swell on Recovery From Bending.

Fig. 4 Angle of recovery after Recovery From Bending.

Fig. 5 Kinetic of recovery change (with aeration and oil change).

Fig. 6 Correlation HRI-r with seal life.

Fig. 7 Device used to measure the Loss of Seal Interference (LOSI).

Fig. 8 Radial load change after immersing seal on mandrel on LOSI device.

Fig. 9 LOSI change after immersion in SAE 10w30 at 302°F (150°C) with aeration and oil change.

Fig. 10 Correlation HRI-I with seal life.

Fig. 11 Recovery From Bending vs. Loss of Seal Interference.

Fig. 12 Correlation HRI-r with seal life.

970552

Contamination Sensitivity of Automotive Components

Xiaojian Tao and Martin B. Treuhaft
Southwest Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

System contamination caused by contaminates or small particles built-in, self-generated, or inhaled from environment presents severe problems. The problems include but are not limited to the malfunctioning of valves, pumps, seals and injectors or lock-up of these components; increased wear of bearings, piston rings, and other friction components; and degradated machine performance. In general, system contamination changes a deterministic system into a stochastic system and shortens machinery service life. In this paper, these contamination problems are discussed in categories and associated analysis, testing and computer modeling methodologies are also discussed.

INTRODUCTION

In the past decades, many researchers [1] [2] have studied contamination related machinery behavior and developed theories and analytical tools in the academic arena. Contamination analysis and control have received more and more attention in industry in recent years [3,4]. People nowadays have come to realize that system contamination caused by contaminants or small particles built-in, self-generated, or inhaled from environment presents severe problems. For example, lubricant contamination changes the requirement of lubricating film thickness and increases friction component wear rates; air, hydraulic or fuel delivery systems contamination causes valves to stick, pumps to leak, seals to break, and injectors to jam; and electronic system contamination causes circuit boards to short, hard disks to fail, information to be lost and systems to malfunction. In general, system contamination changes a deterministic system into a stochastic system and shortens machinery service life. However, contamination analysis, testing, modeling and simulation methodologies will help alleviate or find remedies for these problems. This will help manufacturer and users to:

- Design To Allowable Contamination Levels with Cost Justification
- Obtain Proper Design Parameters with Contamination Consideration
- Define Key Contamination Sensitive Parameters and Components
- Obtain Realistic Service Life Estimates in a Short Period of Time
- Achieve Targeted Service Life

CONTAMINATION ANALYSIS

CONTAMINATION SENSITIVITY DEFINITIONS—An ideal system without contamination may be described by the following formula:

$$Q(t) = F(x,p,t) \qquad (1)$$

where

- t - Generalized time
- x - Operational parameters
- p - Operational commands
- Q(t) - Desired performance index, which is a function of x, p, and t.
- F - Generic functions

When a system becomes contaminated, its operational parameters change into stochastic or random processes such that

$$Q_s(t) = F(x + \Delta x(t), p, t) \qquad (2)$$

where

- $\Delta x(t)$ - Contamination caused fluctuation of operational parameters
- $Q_s(t)$ - Performance index of the contaminated system

Since contamination causes uncertainty in the operational parameters, making them become stochastic processes, the system performance index, $Q_S(t)$, also becomes stochastic. This implies that the performance of contaminated systems is best judged by statistical measures, such as mean value and variance, rather than by more traditional instantaneous measures.

Accordingly, one can define

$$S_p = \frac{E[Q_s(t)] - Q(t)}{Q(t)} \qquad (3)$$

and

$$S_m = \frac{E[(Q_s(t) - Q(t))^2]}{Q^2(t)} \qquad (4)$$

as contamination sensitivity for performance degradation and contamination sensitivity for motion malfunction, respectfully, where E [ ] is the ensemble average of the referenced argument in the brackets. As shown later, specific contamination sensitivities may be evaluated using these formulas. For example, contamination caused bearing, piston ring, and pump performance degradation may be evaluated by formula (3), whereas, automatic transmission valve stick may be evaluated using formula (4). On average, over time, an old contaminated automatic transmission may have the same performance as that of a non-contaminated transmission and equation (3) may give an indication of zero sensitivity. However, at any specific time, no matter how hard one presses the accelerator pedal, it is likely to take a recognizable period of time for the contaminated transmission to shift gears. This difference will definitely show contamination sensitivity when equation (4) is used.

CONTAMINATION CAUSED FAILURE—Contamination caused mechanical failures in general fall into two categories, (a) cumulative performance degradation and (b) motion or functional failure. In the first category, failure occurs in a progressive manner, whereas in the latter, failure occurs often in an unexpected or dramatic form and does not follow an exact repetitive pattern.

Cumulative Performance Degradation—There are two commonly observed failure modes in this category:

a. *Tribological or Wear Failure.* One of the major causes of this failure is due to the fact that contamination in a lubricant changes the lubr-

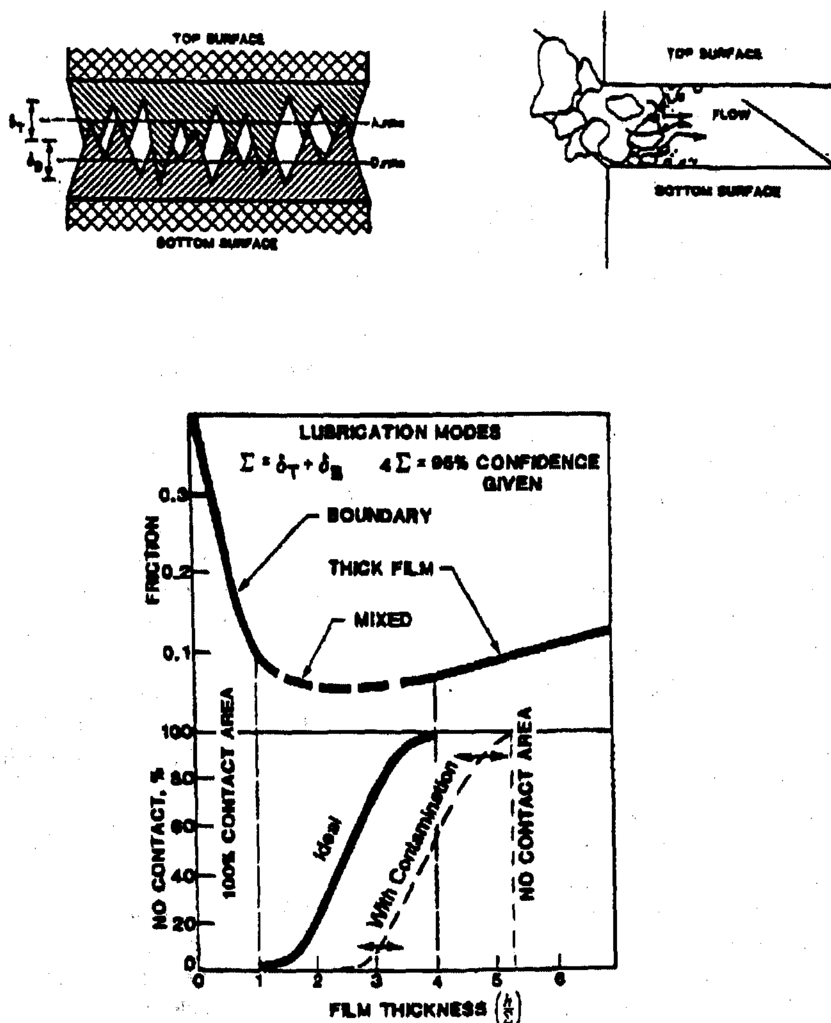
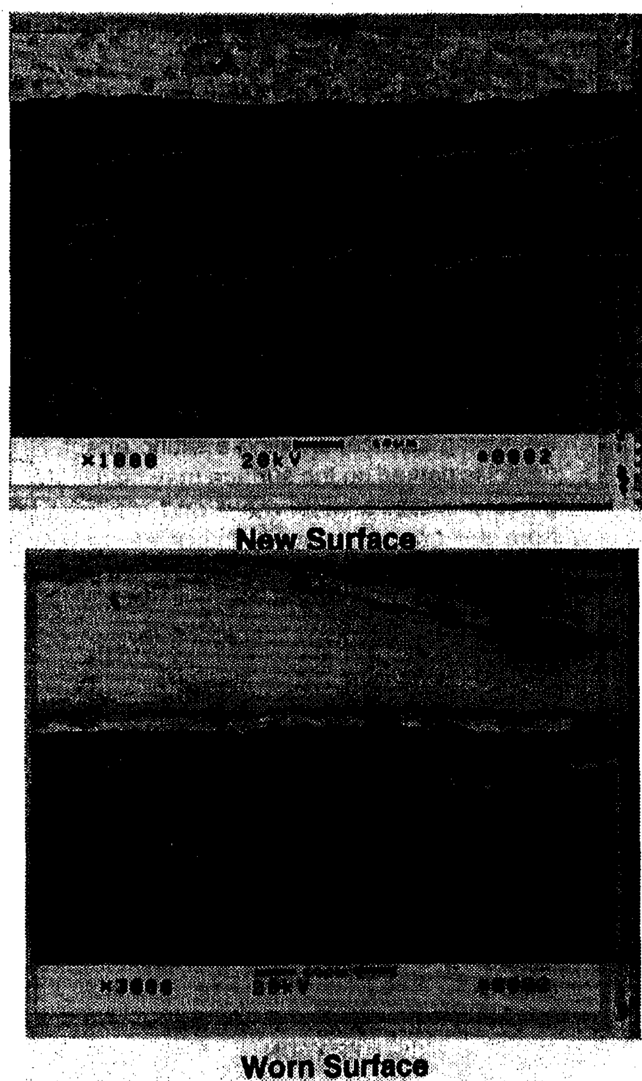


Fig 1. Contamination influence on film thickness

icating film thickness requirements. As shown in Figure 1, conventional lubrication film design uses the four sigma rule to achieve noncontact film thickness. This rule is based on the surface roughness of the contacting friction pair and assumes a statistical confidence level of 95%. Under ideal lubricant conditions, the required noncontact condition may be achieved by such a design. However, when contaminants come into play, the noncontact percentage boundary shown in this curve will shift and increased wear could occur. In addition, different surface treatments will, in general, either increase or decrease this wear process. Figure 2a shows a typical new wear surface and Figure 2b shows the same surface worn by contaminants.

b. *Premature Loading of Filtration System.* A filtration system is often designed to give a certain dust holding capacity when measured by some standard test dust, such as, AC and SAE Fine (FTD). System integrators often utilize this capacity value to design for a required service interval or in some cases, to design a filtration system that will last the whole system life span. These standard test dusts, however, may not represent local real-world dusts in terms of their geometrical shape, size distribution and density and their impact on dust loading [5]. The mismatch sometimes causes filters to load much quicker than expected, resulting in degradated system performance. Figure 3 shows a typical case in this category. The filter plugged in less than one third of its designed service life.



(b)

Fig 2. Typical wear surfaces

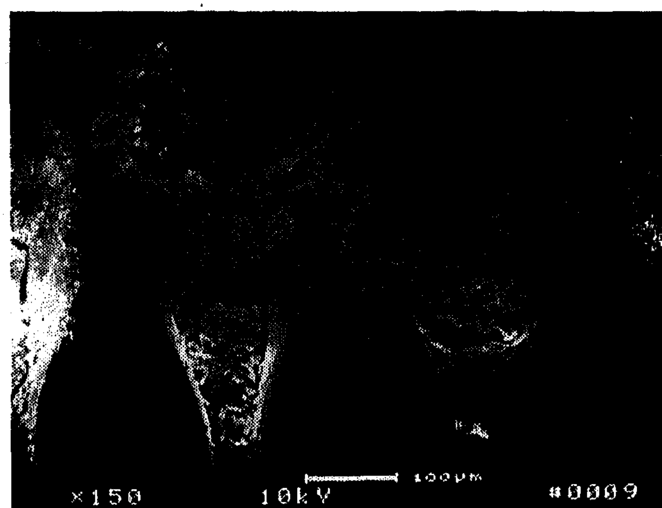


Fig 3. Filter premature loading up

Motion or Functional Failure—There are many different motion or functional failure modes caused by system contamination. While we will not list all these modes, we will list several typical, yet often overlooked motion failure modes in this section. These failure modes include, but are not limited to the following:

- Contaminant Lock
- Agglomeration
- Obliteration
- Hydraulic Lock
- Thermal Lock
- Magnetic Attraction

a. *Contaminant Lock.* Figure 4 shows three typical lock mechanisms caused by moving contaminants in the flow stream. These are known as static force locking, shear force locking, and dynamical force locking. Static force locking refers to the situation where a particle, having nominal size close to the clearance size, comes to the edge of the gap. Due to the sudden changes in pressure upstream and down stream of the particle, it is pushed into the gap causing a jamming effect. Shear force locking

Static Force Locking   Shearing Force Locking

$F = (P_1 - P_2)A_{particle}$ $F = (P_1 - P_2)A_{piston}$

$F \alpha MV^2$

Dynamic Force Locking

Fig 4. Three basic contaminant locking mechanisms

occurs when a moving spool is about to change from one position to another and a particle, having a larger nominal size than the clearance, comes to the middle of the closing gap, keeping the gap from closing because of the particle high shear strength. The dynamic force locking mechanism refers to the phenomenon exhibited by a group of small particles, having a nominal size generally smaller than the clearance, that behave like gas molecules in Brownian motion between the spool and the bore. Here, the pounding force of these small particles causes excessive dragging, which prevents the spool from moving from one direction to another.

b. *Agglomeration.* A very interesting real-world phenomenon is that many filters are designed to have specific nominal cut off sizes, yet contaminants in the passing stream are known to have size distributions below these levels. Intuitively, people think that contaminants or small particles below the cut off size will not be caught by the filter. However, in reality, these filters are often plugged by these fine particles. Why? One answer is agglomeration. Some filter pores are designed in such a way that when the flow stream, carrying the contaminants, passes through the pore, it has to split or splash, changing its flow direction many times before it passes completely through the filter. This repetitive re-direction strengthens the motion behavior of the carried particles in two aspects. First, it makes the dynamic size of the particles become much larger than their physical size. Second, it increases the chances that small particles will collide with larger particles. With the facilitation of some binding agents in the fluid and particle surface properties, many of these particles stick to each other after colliding. As this particle

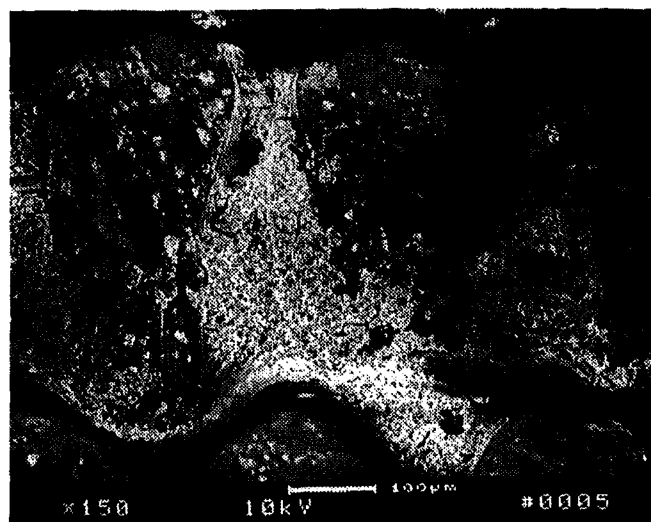


Fig 5. Typical field particle agglomeration pattern

"sticking" process grows, the cluster of particles become larger and larger, finally plugging the openings of the filter. This whole process is called agglomeration. Figure 5 shows a typical agglomeration pattern on a filter removed from field service. In this particular case, the filter had a nominal pore opening size of 70 microns, yet many particles agglomerated on the filter originally had a nominal size of around 5 microns.

c. *Obliteration.* Obliteration is the combined effect of particle silting and polar particle adhesion, as shown in Figure 6. Due to the low boundary layer velocity profile, particles getting to the layer cannot be carried by the flow and start to move towards the tubing surface, creating a silting effect. Some particles parked on the surface may have electrical polarity, which attracts more and more particles from the boundary layer. This leads to a silting surface, with the help of molecular adhesion forces. In some severe situations, obliteration blocks the incoming flow completely, causing system failure.

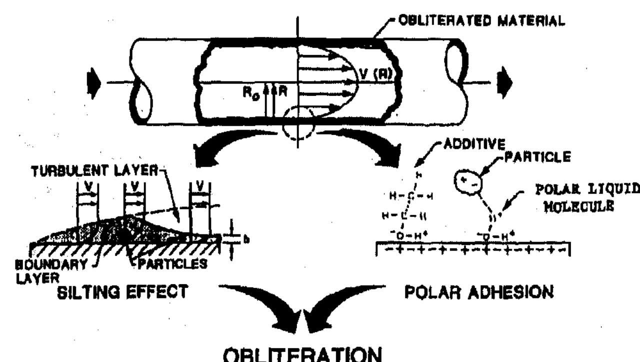
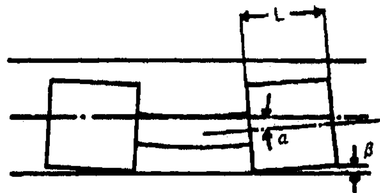
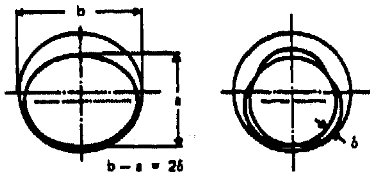


Fig 6. Obliteration working principle

d. *Hydraulic Lock.* Hydraulic moving parts may lock up by unbalanced hydraulic forces. This type of locking has the same apparent look as contaminant locking. Consequently, it often misleads people into using better filters to remove more contaminants (to get the system as clean as possible) at higher cost. However, in practice, even with super clean fluid, the locking phenomenon still occurs. Figure 7 gives the illustration. In virtually any manufacturing process, component dimensions are machined with certain tolerance. It is this tolerance that makes a spool valve have some tapered shape or two adjacent lands not to line up. As flow passes through these tolerance gaps, unbalanced forces create a static locking force or torque which eventually locks up the moving component. Fortunately, this problem often occurs in the product development stage and can be solved with a better design strategy.



(a) Hydraulic Lock from a Curved Axis Spool.



(b) Hydraulic Lock from an Out-Of-Round Spool-Bore Assembly.

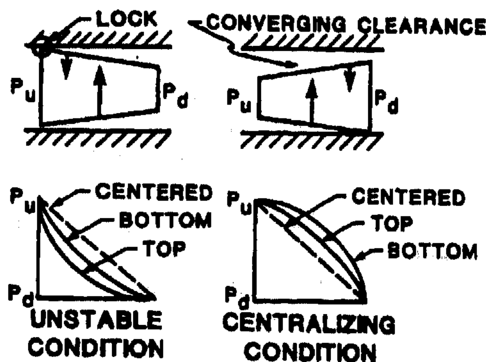


Fig 7. Typical hydraulic locking types

e. *Thermal Lock.* When a system has no or little contamination, moving parts, such as spool valves, work properly even if they experience a thermal cyclic environment. However, when a system is contaminated, particles landed in the gap area will seize the moving parts as clearances are reduced by thermal expansion effects. This phenomenon can become especially severe when two relative moving parts are made from different materials, such as when the bore is made of aluminum and the spool is made of steel. This implies that when a system is known to have to work in a high/low temperature environment, it is important to consider the combined effects of thermal expansion and contamination during the design process.

f. *Magnetic Attraction.* Many hydraulic actuators have electrical or electronic driving devices. These devices generate magnetic fields which will attract ferrous particles, such as iron particles. The attracted particles migrate to the gaps between the moving parts as the electromagnetic devices energize or deenergize. Solenoid valves are typical examples of these types of devices. As particles are attracted into the valves, the armature will seize up.

The above listed failure modes may mutually influence each other. For example, magnetic attraction associated with thermal expansion may produce very serious contamination locking problems. Obliteration and agglomeration may choke system flow, degrading system performance. However, for a particular system, experience shows that one or two key contamination failure modes will typically dominate. As long as these key failure modes are overcome, systems contamination problems will generally be under control.

CONTAMINATION SENSITIVITY ANALYSIS— Contamination sensitivity analysis is a very important step toward achieving total system contamination control and should be considered in all the stages of machine development and usage, i.e. contamination sensitivity analysis should be performed when a machine is

- In the design stage
- In the prototyping stage and
- In operating working condition

With respect to the machine development stage, the corresponding contamination analysis approaches are applicable:

- Preventive design analysis
- Product development analysis
- Contamination diagnostic analysis

Preventive design contamination analysis refers to contamination analysis performed on machines which are undergoing design. For instance, many systems have some components which are likely to have potential

contamination problems. Designers at this stage should select proper filtration systems to protect these components. If possible, designers may build mathematical models to simulate different component arrangements to evaluate the system's anti-contamination performance in virtual reality. In cases where computer modeling and simulations are too complicated, designer may use established databases to assist the analysis.

Product development contamination analysis refers to situations where prototype products have already been manufactured and their contamination sensitivity is to be investigated. At this stage, contamination sensitivity of the prototype products can be evaluated using certain specified or standardized contamination sensitivity test procedures. Although test procedures may differ from each other, they have the following aspects in common:

a. They all vary contamination concentration levels or size distributions from light or small to heavy or large.
b. They all use specific contaminants or standard test dusts such as SAE Fine Test Dust.

Through contamination sensitivity testing, product contamination sensitivity can be defined in terms of contamination concentration levels and particle size distributions. As long as system contamination concentration levels are controlled below the concentration level given by the sensitivity test, and the size distribution is also controlled within the sensitivity test's specified range, the system is not likely to experience contamination problems.

In many cases, knowing the contamination sensitivity of a product is not enough. People want to know how long the product will last under normal working conditions, and whether its real life span will meet the required design life or not. A direct approach to answering these questions is to operate a representative group of the product under its working condition for an extended period until all stop working. However, with today's heavy competition among manufacturers, and with constraints of cost effectiveness, this approach may not be the best among all possibilities. Instead, the accelerated life testing method may be used to investigate longevity problems of a product. In performing accelerated life testing, special attention must be given to both the chemical composition and concentration level of the contaminate so that the failure mode induced during testing will truly represent the field failure mode. Otherwise, accelerated life testing may not produce meaningful results.

Contamination diagnostic analysis refers to the situation where a real system has failed to work properly during field operation due to contamination problems. However, the failure mode is not known. Conducting a diagnostic contamination failure mode analysis on all possible causes, especially on those mentioned in the earlier sections of this paper, will usually show that one or two failure modes dominate. Once these dominant modes are found, the following steps can be used to confirm and remedy them. To confirm a field failure mode in general means to reproduce the failure mode under controlled laboratory conditions. This requires reproduction of the contaminants that caused the field failure and the utilization of these reproduced stimulants in conducting laboratory testing. As soon as the field failure modes are reproduced through laboratory testing , the confirming process is finished. The remaining step is to find a solution which solves the problem.

Figure 8 summarizes contamination sensitivity analysis in a generalized flow chart. As indicated in this chart, in order to achieve targeted longevity, contamination analysis should be conducted alternatively between components and systems in all design, prototyping and field operation stages, utilizing methodologies such as contamination sensitivity testing, accelerated life testing, failure mode analysis and reproducing, and mathematic modeling and computer simulations. In the following section, some case studies are presented.

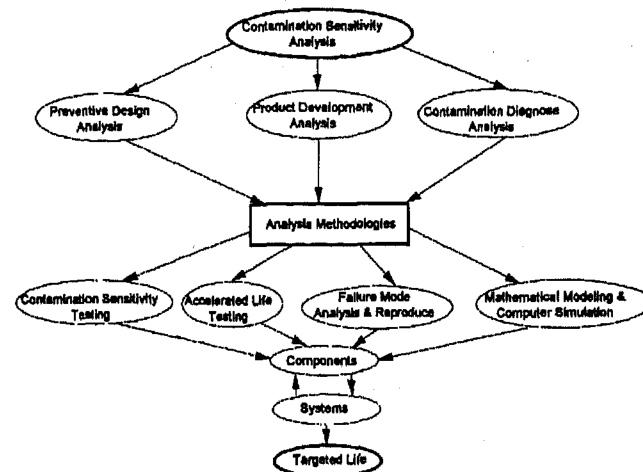


Fig 8. Contamination sensitivity analysis flow chart

CONTAMINATION ANALYSIS CASE STUDY— Contamination sensitivity analysis may best be illustrated through case study examples. In this section, seven specially selected case studies are presented to show the importance and benefit of contamination sensitivity analysis. These cases are listed as follows:

- Dust Ingression Contamination Study
- Lube Oil Additive Sensitivity Study
- Shift Valve Contamination Lock Study
- System Failure Mode Diagnosis
- Contamination Sensitivity Experiment Study
- Accelerated Life Testing
- Contamination Mathematical Modeling and Computer Simulation

Case 1. Dust Ingression Contamination Study. The objective of this study was to investigate contamination sensitivity of engine piston rings upon exposure to two

different dust cuts inhaled through the engine air intake [6]. The engine used in this study was a 5.9 liter, six cylinder, turbocharged industrial-type engine. The engine was operated at 2500 rpm, full load, for about 260 minutes. The first dust cut, a 2-4 µm dust fractionated from AC Fine dust, was then injected into the intake air stream at a concentration of 0.27 mg/ft^3 air, while the engine was operating at the above noted test condition. After 30 minutes dust injection was stopped. When engine operating time reached 350 minutes, a second dust cut, 4-10 µm dust, was injected into the air intake with the same dust concentration level. Again, dust injection was stopped at the end of 30 minutes, while the engine operated under the same profile for another 60 minutes. Ring wear was measured in real-time throughout testing, using two calibrated radioactive isotope tracers, Fe-59 and Cr-51. Figures 9 and 10 show the experimental results. Fig. 9 shows cumulative total ring wear, chrome wear and the chrome to total wear ratio. Places where different dust cuts were turned on and off are also shown. Fig. 10 shows the wear rates for the total ring wear and chrome wear. The curves in Fig. 10 are calculated from the derivatives of the cumulative wear curves shown in Fig. 9. Using formula (3), one may easily calculate the contamination sensitivity values for these two different dust cuts as follows:

Total Wear Sensitivity:

$$S_{T,\ 2\text{-}4\mu} = \frac{\Delta Q_{T,2\text{-}4\mu}}{Q_T} = \frac{25}{13} = 1.92$$

$$S_{T,\ 4\text{-}10\mu} = \frac{\Delta Q_{T,4\text{-}10\mu}}{Q_T} = \frac{52}{13} = 4.00$$

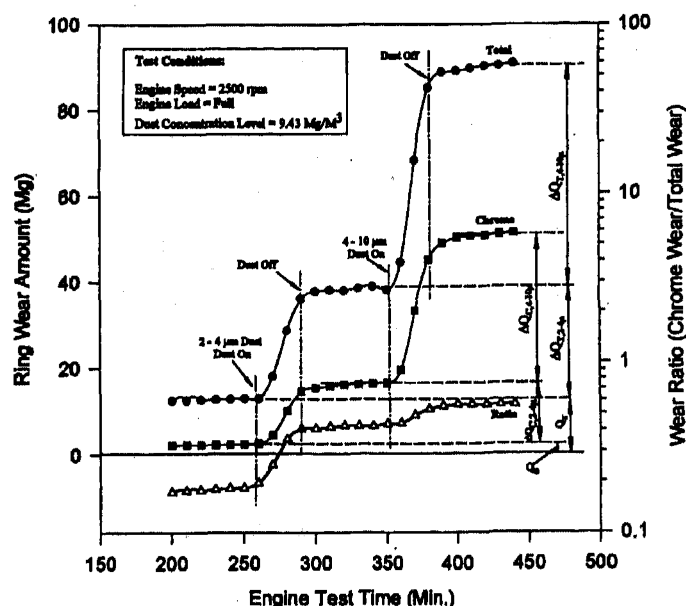


Fig 9. Dust ingression test - cumulative wear & wear ratio

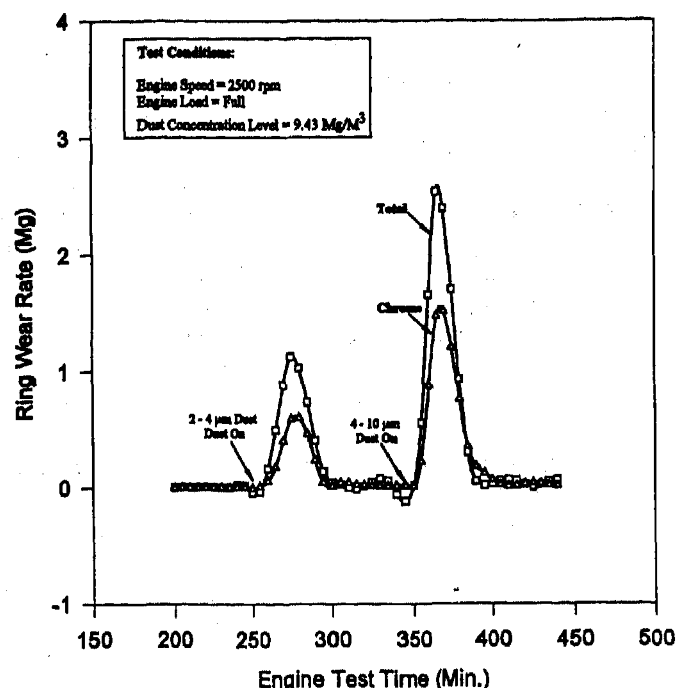


Fig 10. Dust ingression test - ring wear rate

Chrome Wear Sensitivity

$$S_{c,\ 2-4\mu} = 5.96$$

$$S_{c,\ 4-10\mu} = 14.78$$

Examining the results, one may conclude that for this particular engine:

a. Ring wear is more sensitive to the dust cut of 4-10 µm than to a cut of 2-4 µm,
b. The wear ratio increases as the dust cut changes from zero to 2-4 µm and then to 4-10 µm. Since chrome coating is concentrated in the top ring, this ratio increase implies that top ring wear is very sensitive to dust ingression.
c. Ring wear rate changes dramatically as dust turns on and off. The peak wear rate with 4-10 µm dust is more than twice as high as that of the 2-4 µm dust.

This case study shows how important a good air filter is to protect an engine from dust ingression.

Case 2 - Lube Oil Additive Sensitivity Study. Each year chemists from different petrochemical companies develop varieties of different additive packages to improve certain lubricating properties of fuel and lubricants. What is the effectiveness of these packages? People often run bench tests to answer this question. Due to the undefined correlation between bench testing and field operation interpretation of the bench test results can be difficult and

sometimes misleading. However, treating additives as specially formulated contaminants and utilizing radioactive tracer real-time wear measurement techniques, one can perform contamination sensitivity analysis to screen different fuels, lubricants and additive packages and to show their effectiveness with respect to actual engine operation. Figure 11 shows sensitivity analysis results of four different lubricant additives on engine wear. All tests were performed using a four cylinder 2.3 liter engine. The major wear trending curve in this figure is obtained through alternatively testing a reference oil between groups of additives. Each point along this curve represents cumulative wear at the end of an identical four-hour engine test. A representative history of the four-hour test is given in Fig 12. The insert in Fig. 11 shows the sensitivity analysis results calculated using equation (3). Here, one can easily see that additive I is better than additive IV. However, casually comparing the wear data along the wear curve, one may reach the opposite conclusion, since the amount of wear for additive I is higher than that of additive IV. Fig 12 shows the influence of ambient cold start, speed, and load on the wear. In fact, contamination sensitivity analysis can be performed on every point along the histogram curve to examine a variety of properties of additive packages, such as:

- Additive Sensitivity to Cold/Hot Start
- Additive Sensitivity to Speed Change
- Additive Sensitivity to Load Change
- Additive Sensitivity to Steady State/Dynamic Operations (for instance, Urban/Highway Driving Cycles)
- Additive Sensitivity to Lubricant Carry-Over-Effects

These and many other properties are very important to the success of good lubricants. The methodologies given in this case study may offer an effective, time and cost savings approach to the evaluation of fuels, lubricants, and fuel and lubricant additive packages.

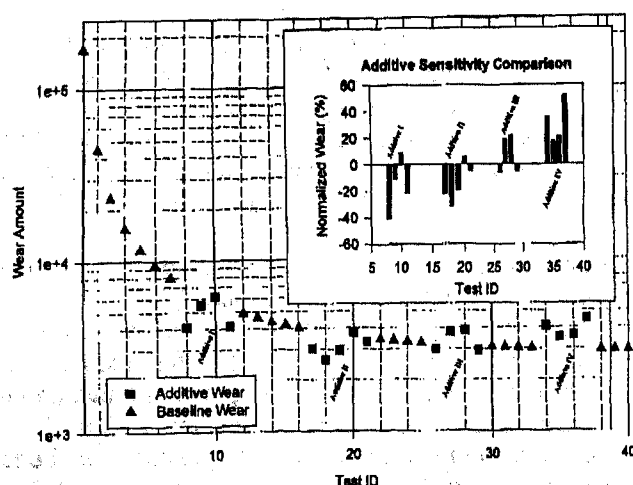


Fig 11. Lube oil additive sensitivity test results

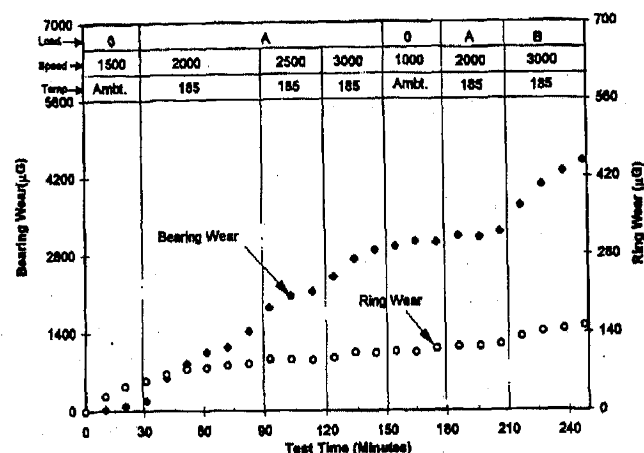


Fig 12. Individual wear measurement histogram

Case 3 - Shift Valve Contamination Lock Study. The objective of this study was to investigate the contamination impact on spool valve shift response time. As shown in Fig. 13, a spring loaded spool shift valve configuration was selected for this study. During the experimental study, two different tolerance groups were used. One group had minimum tolerance, the other had maximum tolerance. The valves were first tested under clean fluid (less than 5 mg/L) conditions, and then with fluid contaminated with 0-200 μm dust at a concentration of 250 mg/L. During each test, the valve spool was pressurized and depressurized, with identical dwell time intervals and the valve response time was measured. Figures 14 and 15 show typical experiment

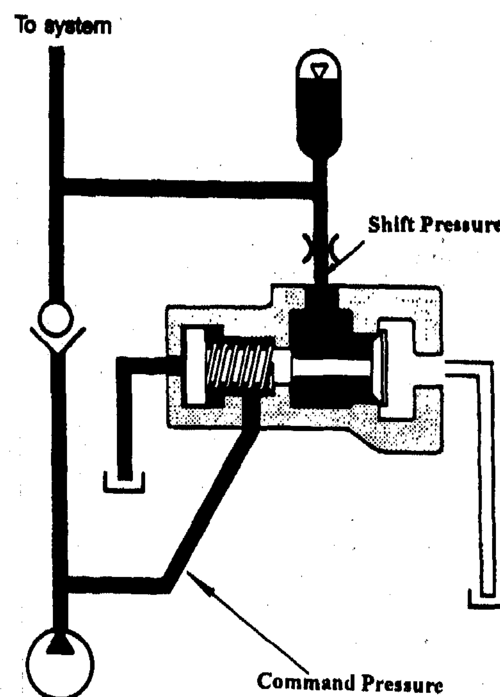


Fig 13. Shift valve contamination lock study

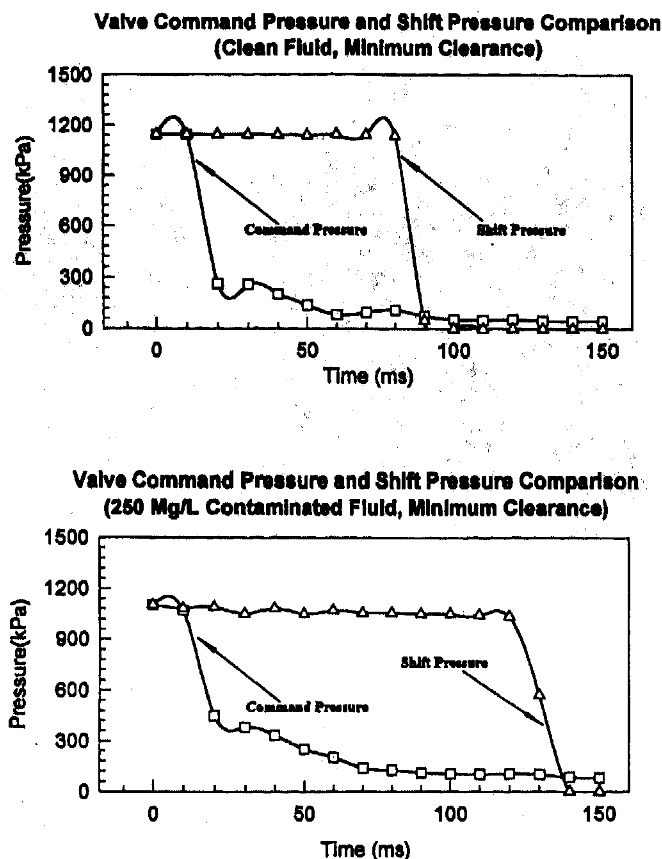


Fig 14. Valve shift pressure signals of minimum clearance

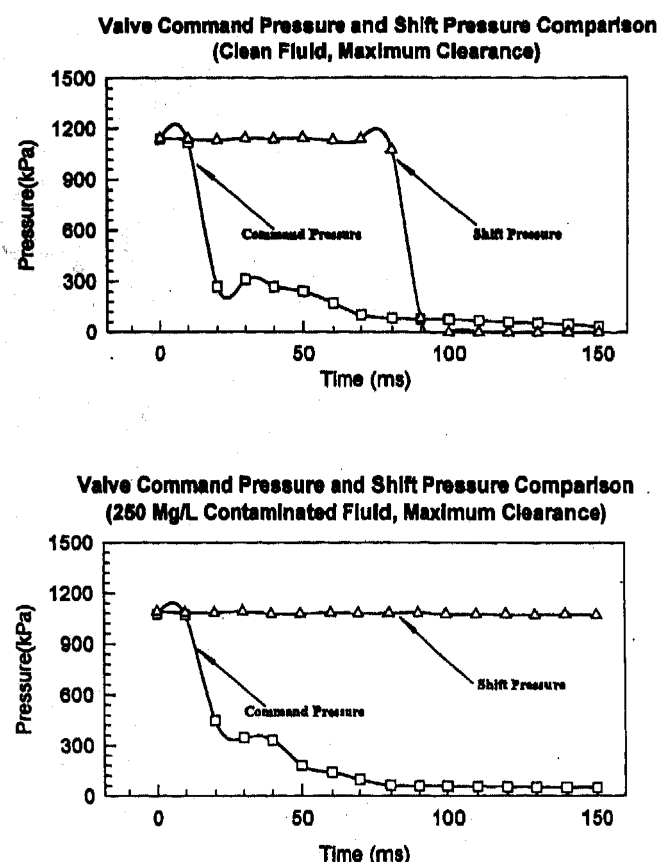


Fig 15. Valve shift pressure signals of maximum clearance

results for valves with minimum and maximum clearance, respectively. For this particular valve assembly, one may conclude that:

a. Under clean fluid conditions, both minimum and maximum clearance valves have similar response time.
b. Under the specified contaminated fluid conditions, both minimum and maximum clearance valves experience some contamination sensitivity, and valves with maximum clearance are more sensitive than valves with minimum clearance.

Valves with minimum clearance shifted within the specified duration time (150 ms) under dirty fluid conditions, while valves with maximum clearance did not shift or "lock up" throughout the specified time duration. In practice, lock-up may not be permanent; the valves may eventually shift, but with a much longer response time. Hydraulic valves, pneumatic valves, and automatic transmission valve bodies will experience similar conditions as shown in this case study. A typical example is that one can definitely feel the "gear jumping phenomenon" when he operates a vehicle with an old or contaminated automatic transmission. Each year there is much transmission warranty repair due to valve sticking has caused by fluid contamination. However, this problem can be solved through contamination sensitivity analysis in the design and prototyping stages.

Case 4 - System Failure Mode Diagnosis. A liquid transfer system was designed to last at least 2000 hours. However, the system stopped working properly at less than one third of its designed life. This system is a two-stage pumping system. The first stage pump has a 30 micron nominal depth media inlet filter and the second stage pump has a 70 micron nominal reversed Dutch weave screen inlet filter. Preliminary investigation indicated that the first stage was normal and that the second stage pump was in good condition. Logic indicated that the second stage inlet filter was the cause of the problems. A SEM photograph of this filter (Fig. 16) confirmed that the screen's filter pores were plugged. This raised an interesting question, namely, why didn't the first stage filter, which has much smaller pores, plug instead of the second stage inlet filter, which has larger nominal pores? It was suspected the particle agglomeration caused the problem. To confirm this as the failure mode, two different experiments were conducted. The first experiment introduced standard Arizona Fine Test Dust directly into the second stage filter. This generated a system malfunction which seemed to mimic that caused in the field. However, further study of SEM photographs of the contaminated filter (see Figure 17) showed that the lab failure did not, in fact, mimic the field failure. A second experiment was conducted using a mix of contaminant simulants which were constituted from field

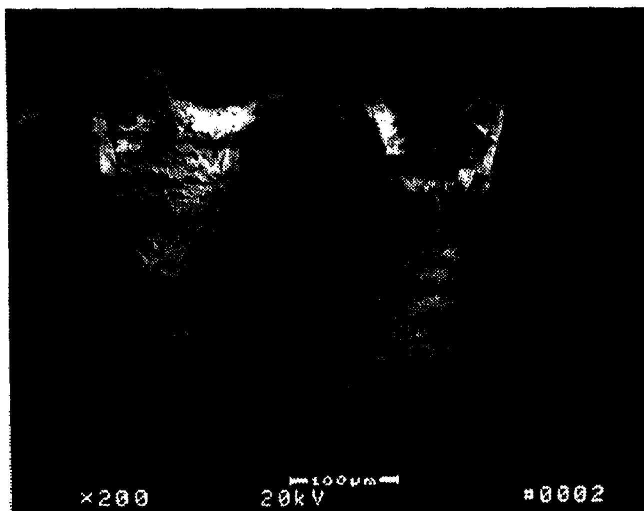


Fig 16. Field failure mode

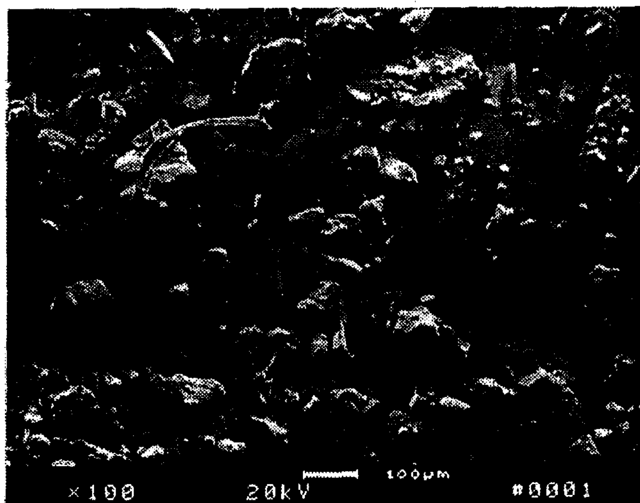


Fig 17. Simulated failure mode usling AC FINE test dust

contaminants. This contaminant mix was introduced to the first stage, with a calculated acceleration dosage to simulate true field operations. This generated a system malfunction which looked like those being experienced in the field. SEM photomicrographs of the second stage inlet filter (see Fig. 18) strongly resembled SEMs of the field filter shown in Fig. 16. These experiments confirmed that agglomeration of the fine particles was the main plugging mechanism, not dust loading.

Case 5 - Contamination Sensitivity Experimental Study. The objective of this study was to investigate the contamination handling ability for a group of turbine pumps. These pumps were installed in a test reservoir, in which the test fluid was circulated with proper speed and agitation to maintain certain level of suspended contaminants that were specifically formulated to simulate field operations.



Fig 18. Simulated failure mode using specially formulated field contaminants

Throughout testing, contamination concentration levels were checked and adjusted at specified time intervals to compensate for particles lost in the test pump assemblies. These pumps were tested for 65 hours and the results are shown in Fig. 19. The predicted data band, estimated average flow rate trend and its predicted band with 95% confidence level are also shown. Using Least Squares Curve fitting technique, one may obtain the flow trending formula:

$$Q = Q_o - Kqt$$

or

$$\frac{Q}{Q_o} = 1 - \frac{Kqt}{Q_o}$$

This formula characterizes contamination caused flow degradation, which is very important for both manufacturers and users. For manufacturers, this formula may be utilized to design pump contamination tolerance. For users, it may be used to evaluate different suppliers and to determine proper pump service intervals.

Case 6 - Accelerated Life Testing. Accelerated Life Testing is a very effective test method to predict system service life in a relatively short period of time. There are three commonly used approaches. The most frequently used is the so called S-N curve approach, which requires several contamination (stress) levels, from high to low, to stress the test articles to failure at each level in order to produce a representative S-N curve, as shown in Fig. 20. Article life is estimated by extending the low stress portions of the S-N curve. A second approach is a two-level contamination stress test to establish a high stress life vs.

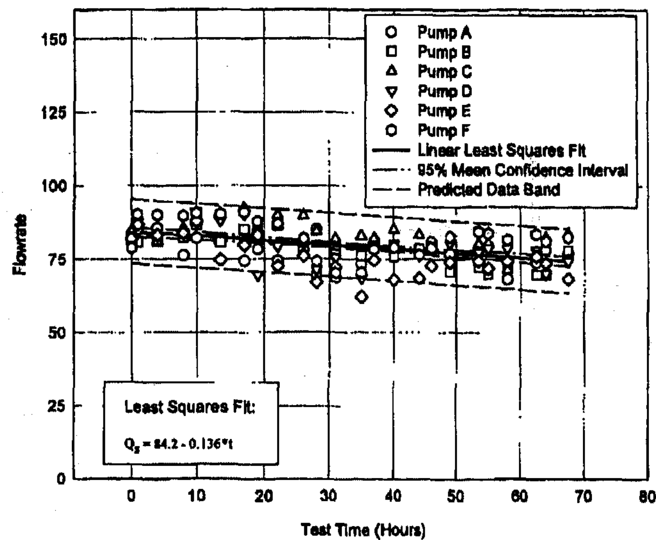


Fig 19. Contamination sensitivity experiment study - test results

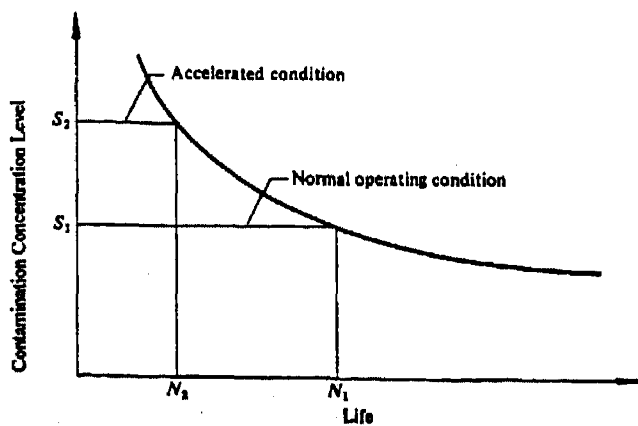


Fig 20. Accelerated life testing - S-N curve approach

low stress life curve. As shown in Fig. 21, the test articles are first tested to failure at a high contamination level to establish point A. Test articles from the same batch are then tested at the high contamination level to about 60-90% of failure. Testing is then continued at low concentration to failure to establish point B in the figure. Connecting points A and B and extending the line to the horizontal axes gives point C, the test article service life.

Sometimes people like to know a component's expected performance history under certain contamination levels in advance. In this situation, the approach of a single acceleration factor is used i.e., the articles are tested only at a specific accelerated concentration level. At the end of testing, test results are converted to predicted life by utilizing some known S-N relationship and mathematical calculations. Fig. 22 shows the accelerated life testing results of this approach. During the test, six pumps of each

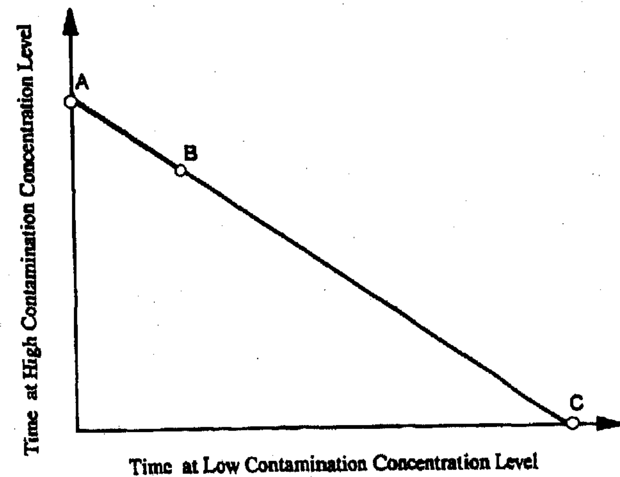


Fig 21. Accelerated life testing - two-point approach

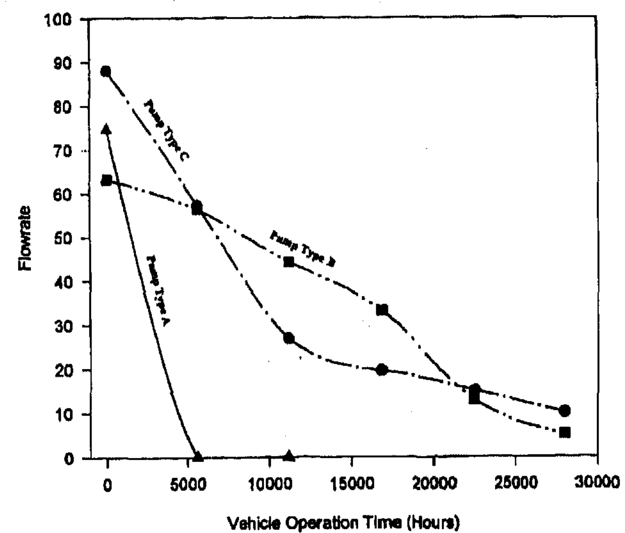


Fig 22. Accelerated life testing results - single level approach

type were tested under the same test conditions, using one specific contamination level, for 1000 hours. Test results were converted to predict service performance history shown in the figure. Examining the curves provides important information:

a. How to select a pump type to meet a targeted service life in a cost effective manner (if pump price is known).
b. How to improve each pump type to maximize its performance and service life.
c. How to utilize pump residual life to optimize total system performance.

Case 7 - Contamination Mathematical Modeling and Computer Simulation. A non-contaminated system, whether linear or nonlinear, may be described by deter-

ministic system governing equations. For example, a motion related system may be described by a second order differential equation in which the spring ratio and damping factors are constant. However, when the system is contaminated, the spring ratio and damping factor are no longer constant. For instance, when a shear force jamming particle comes to the gap between two relative moving components, the spring ratio will dramatically increase, whereas, when the particle leaves the gap, the spring ratio will restore. The same is true for the damping factor. A cluster of small particles, pounding the moving parts, will make the damping factor become a random process. This, in turn, will make the deterministic systems become a stochastic system. To model such a system, stochastic system theories [7] should be used. Fig. 23 shows a turbine motor-pump assembly and its deterministic and stochastic mathematical models. Figure 24 shows the preliminary simulation results. It is obvious that without contamination, pump speed becomes constant once the pump achieves steady state. However, this is not the case for a contaminated system in which the pump speed fluctuates even after stationary average speed has been reached. Additional performance results, such as flow degradations, versus the contamination level, may be obtained through further modeling and simulation effort.

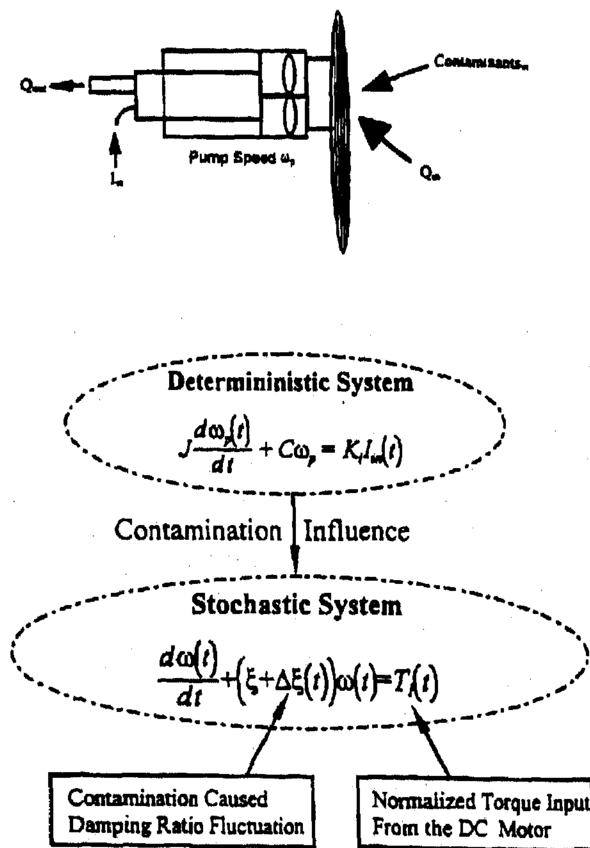


Fig 23. Fundamental mathematical model of contamination systems

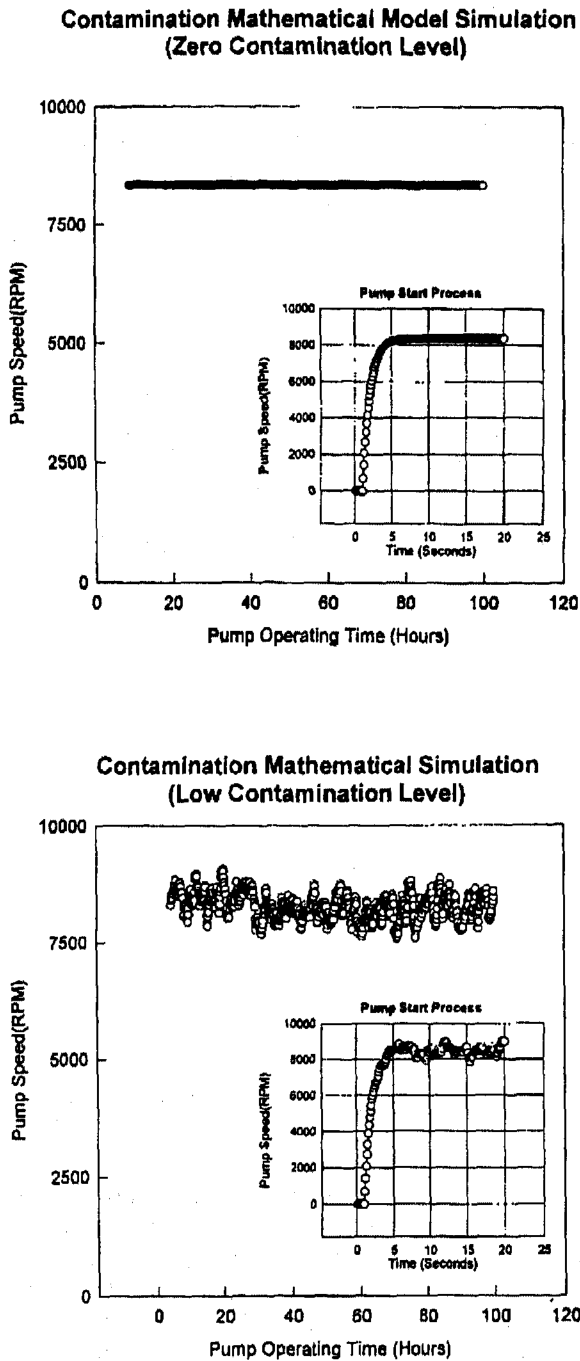


Fig 24. Preliminary contamination mathematical model simulation results

CLOSURE

Contamination sensitivity analysis methodologies, including testing, analysis modeling and simulation technologies presented in this paper, offer a systematic approach towards system contamination control. Failure

modes analysis presented in Section 2.2 may help pinpoint some system failure modes that occurred in field operations, but are often neglected. Case studies presented in Section 2.4 cover several areas of tribology, lubrication, filtration and contamination related issues and interconnections between these disciplines. Dust ingression studies may help to understand the importance of filtration systems and the role of wear resistance surface treatment. Additive sensitivity studies may help lubricant developers investigate lubrication properties of different additives and base stocks or to screen these products in a real-time fashion. Valve lock studies may offer an insight into malfunctioning components. Failure mode case studies may help to understand some filtration phenomenon which look impossible by conventional calculations, but occur during field operations. Contamination sensitivity and accelerated life testing may offer tools to predict machinery service life in a short time frame. Modeling and simulation studies may let designers work in virtual reality to investigate contamination related issues. In summary, proper utilization of contamination control tools will assist simple and complex components and systems to achieve their targeted service life.

REFERENCES

[1] E. C. Fitch, **Fluid Contamination Control**, FES, Inc., Stillwater, 1988.

[2] John K. Beddow, **Particulate Science and Technology**, Chemical Publishing Co., Inc., New York, 1980.

[3] Vickers Guide to Systematic Contamination Control," Vickers, Inc., 1992.

[4] Contamination Control and Filtration Fundamentals," Pall Industrial Hydraulics Co., 1994.

[5] "Comparison Between Real-Life Dust Samples and Standardized Test Dusts," Karen Kohl and Gary Bessee, Proceedings International Filtration Conference, San Antonio, Texas, July 1996.

[6] "Engine Wear Experiments," Martin B. Treuhaft, SwRI Internal Research Final Report, 1989.

[7] "Methodologies for Nonlinear Systems Subjected to Stochastic Parametric and External Excitations," X. Tao, Ph.D. Thesis, Oklahoma State University, December 1990.

970556

Filter Performance Requirements for Engine Air Induction Systems

Neville J. Bugli
Ford Motor Company

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Automotive engine air intake filters are constantly being challenged to deliver higher filtration performances. The need to protect the engine from abrasive contaminants is ever increasing to achieve longer engine life and improved engine performance. This paper discusses some of the key issues affecting engine filter performance.

Currently, the SAE J726 procedure is used to test and evaluate engine air induction filters (AIF). The advantages and limitations of this procedure are discussed. Based on this procedure current engine air filtration technologies are also compared.

Engine protection requirements are also discussed relating to contaminant size and concentrations. To design robust engine air induction filters, the particle size and concentration ingested by the engine should be controlled. In addition to the overall mass (gravimetric) efficiency of the filters, the fractional size efficiency of these filters should also be measured. Fractional size efficiencies would quantify the contaminant size/concentration ingested by the engine which is related to engine wear, oil consumption, power loss etc..

Field evaluations are presented which capture the real time performance of engine air induction filters in dusty environments. Field evaluations should be mandatory to design effective and robust engine air intake filters.

New methods and improvements are briefly recommended for the SAE J726 procedure in quantifying air induction filters. The proposals are to include fractional size efficiencies in addition to the mass efficiency currently measured. Future trends in engine air induction filter systems are also discussed which may benefit from the new test proposals.

NOMENCLATURE

AIS	- Air Induction Systems
AIF	- Air Induction Filters
MAFS	- Mass Air Flow Sensor
TSP	- Total Suspended Particles (ambient contaminant concentrations)
APG	- Arizona Proving Grounds

INTRODUCTION

Designing robust engine air induction filters is always a challenge. Engine air filters currently designed are required to deliver higher performance levels in smaller package designs. Further, higher filtration performances are required to reduce engine wear and extend engine durability. The engine air induction filter has to survive and perform in a variety of environments and driving conditions. For this paper we will restrict our study and analysis to engine air filtration systems used on passenger cars and light truck applications. Engineers designing air induction filters need to know the different parameters and conditions the air filter is going to be exposed. Packaging constraints also need to be met and satisfied. Packaging is probably a top priority when designing engine air induction systems and in many instances has a dramatic effect on filter performance.

Understanding engine filter performance requirements is mandatory in optimizing air induction system (AIS) designs. The filter system design is especially important to manage flow and flow distribution through the filter. In addition to meeting filtration requirements, the engine AIS filtration also meets and exceeds other requirements. In this paper we briefly

cover some design parameters for engine air induction filters.

We also need to understand the current laboratory evaluation methods to qualify engine AIF. The current SAE J726 method is briefly discussed with its advantages and limitations. Based on the SAE J726, benchmark data are presented on filtration performances for the most popular filter media technologies used today. Performance comparisons include dust holding capacities, initial efficiencies and overall efficiencies.

This paper also discusses engine protection requirements to reduce wear. Typical engine design clearances are reviewed and its impact on engine AIF performances. Studies show that engines need to be protected from abrasive dusts down to 1μm in size (1, 2, 8). Controlling the particle size and concentrations ingested by the engine reduces engine wear and extends engine life/durability.

To design robust engine AIF, we need to understand the field environment as it exists in the real world. The author feels that it should be mandatory for a design engineer to evaluate engine AIF in the field. This paper discusses some field results measured at the Ford proving grounds in Arizona (APG). Efforts are ongoing to evaluate more engine AIF in the field. With majority of the automotive companies being globally oriented, we need to understand what's out there in the rest of the world.

Based on some of the filter performance requirements and expectations, recommendations are being suggested to improve the current SAE J726 standard. The recommendations include;

- Quantifying the dust concentrations ingested by engine
- Identify suitable test aerosol
- Measure fractional size efficiency (by number & mass)
- Simple to use

Future trends and developments in engine AIF are also discussed. The new designs may also benefit from the new recommended test procedures.

ENGINE 'AIF' SYSTEM DESIGN

Over the years, AIS filter designs have dramatically changed to meet and exceed performance parameters. Air induction filtration systems that are designed to protect the engines have to meet and/or contribute to the following key performance requirements:

1. Filtration efficiencies (initial and overall)
2. Dust/contaminant removal capacity
3. Overall sound attenuation;
   - ⇒ In passenger cabin
   - ⇒ Drive by noise requirements
4. Flow management (restriction)
5. Overall Emissions
6. Service requirements

Overall higher design and performance expectations are demanded from engine air induction systems. The design of engine AIF and filtration systems have to meet and satisfy a variety of parameters;

- ♦ packaging environment
- ♦ design compactness
- ♦ higher engine flow rates
- ♦ higher media face velocities
- ♦ higher filtration performances
- ♦ lower head loss (pressure drops)
- ♦ lower costs
- ♦ mass air flow sensor requirements-emissions
- ♦ higher underhood temperatures
- ♦ extended service life & durability
- ♦ generic designs
- ♦ reduce parts complexities
- ♦ recyclability

Excellent related information on engine air induction filters and filtration system designs is provided in references 1, 2 & 19. The engine AIF systems has to be <u>tuned in</u> not only to meet filtration, but other significant requirements. For this paper we will only discuss some filtration performance requirements for engine AIF.

CURRENT FILTRATION PERFORMANCE LEVELS FOR 'AIF'

A variety of air filter media and filter designs are available today to deliver the required filtration performances. Air induction filtration systems should be optimized to provide maximum engine protection from abrasive contaminants. It is widely and well accepted that ***" Engine Air Inlet is the Major Source of Harmful Contaminants"***.

Abrasive contaminants ingested by the engine cause irreversible wear and damage to the engine. Engine wear and protection requirements will be discussed later.

There has always been a debate on the level of AIF performance required to adequately protect the engine. One of the key parameters is the contaminant removal efficiency and how high should it be. Previously

for AIF, the overall gravimetric efficiency would be specified and well accepted by the end user. Today, in addition to the overall efficiency, the end user also specifies a initial efficiency requirement for engine intake filters or AIF. It has been well demonstrated that a clean filter media exhibits a much lower collection efficiency compared to a loaded filter (1,2,31). Even small amounts of contaminant loading dramatically improves the collection efficiency of the filter. Engine designers are increasingly concerned about the amount and type of contaminant that could be ingested by the engine. Nevertheless, the problem of engine wear as it relates to abrasive contaminants (type, size, shape, concentrations etc.) is complex and quantitative analysis is always challenging.

CURRENT TEST METHOD - We currently use the SAE J726 procedure to evaluate engine air induction filters. This test procedure is popular and followed in North and South America. Other standards followed in Japan and Europe are JIS-D-1612 and ISO-5011. All of these standards are very similar and may yield comparable results. There is also an ongoing effort to commonize the SAE J726 with the ISO standard.

Recently, the SAE J726 procedure went through significant and useful revisions. Although not perfect, the SAE J726 has been well accepted as the industry standard for testing and evaluating engine AIF. Some of the key features and advantages of the SAE J726 procedure are;

- Simple
- Standard test dusts used
- Will provide restriction and head loss
- Will measure dust loading/capacity characteristics
- Will measure dust mass/gravimetric efficiency
- Accelerated test
- Provide relative and competitive data
- Covers a wide range of applications

The SAE J726 procedure may have some limitations. Some of them are noted below;

- Static test - The vibration effects are absent.
- Real flow simulation/duty cycle - Although the SAE J726 has a variable flow cycle, it may or may not represent real world driving conditions.
- Standard tests dusts - Real world dusts may be different. Studies have shown that dust/dirt distributions vary depending on geographic locations (1, 34).
- Contaminant size data - There are no provisions for measuring the contaminant size that is presented to the filter and ingested by the engine.
- Contaminant concentrations - There are no requirements or acceptable contaminant concentrations limits for the dust ingested. Ingested air cleanliness is important to protect engines from wear (1, 2, 3).
- Environmental effects absent - Temperature effects are not taken into account. Although temperature cycling may not have a significant effect. On the other hand the humidity may have an effect on the AIF performance depending on the type of filter media used.
- Driving conditions absent - Effect of different driving and ambient conditions are not taken into account. Driving conditions are important and have to be understood before designing engine intake filters (1,2,3,19,31).
- Service life requirements are not addressed.

It is generally agreed that the nature of the above limitations are quite complex and quantitative determinations may not be easy. Nevertheless, they play a very important role in designing engine AIF.

COMPETITIVE EVALUATIONS - For this paper we investigated the filtration performance of various O.E. air induction filters. For simplicity of comparisons the various filtration systems tested are grouped in three categories based on the type of filter media used. The three most common media categories are;

a) Dry paper media
b) Treated paper media
c) Synthetic felt media

The dry paper media evaluated were generally constructed of cellulose fibers. Some dry media used a blend of cellulose/synthetic fibers and some had a phenolic resin applied to the fibers to improve the stiffness and durability of the media.

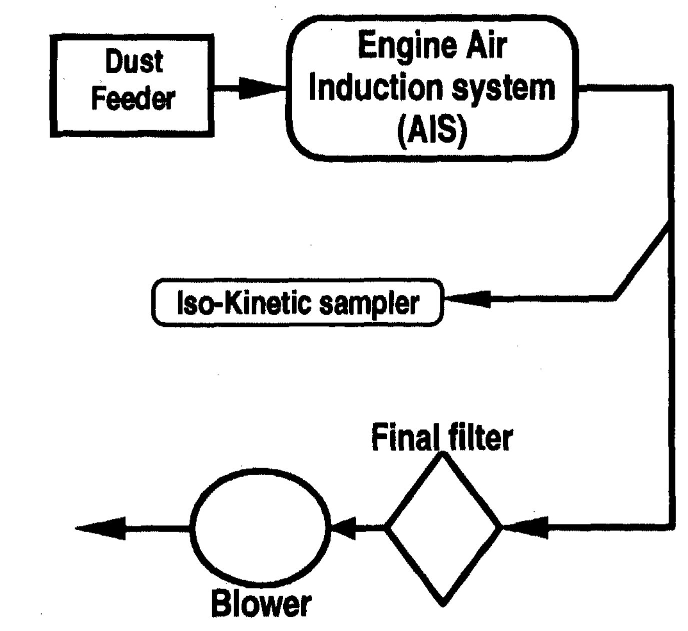
The treated paper media evaluated had a similar media fiber mix as the dry paper media. The treatment was mainly mineral or synthetic oils.

The synthetic media evaluated was generally a multilayered structure (about 2 to 3 layers). The multilayered media was arranged to give a gradient density effect (coarser to finer fibers) to capture contaminants within the media. The media construction ranged from cellulose/synthetic to 100% synthetic fibers (mostly polyester). Needled felt or spunbonded construction was commonly used. All synthetic media evaluated were dry.

The information on the filtration performances for the above AIF was analyzed from a competitive benchmark database. The competitive benchmark data includes vehicles from model year 1993 to present. The competitive analysis presented in this paper includes some of the latest filtration technologies used for engine AIF.

Complete air induction systems were evaluated using the SAE J726 procedure with one exception. Instead of using the absolute filter downstream we used iso-kinetic sampling method to calculate the gravimetric efficiencies. Figure 1 shows a schematic sketch of the test set up.

Figure 1: Schematic Test Setup



By using iso-kinetic sampling we were able to improve our measurement sensitivity and repeatability. All AIS were evaluated under constant flow conditions. Further, all systems were flowed at their maximum flow rate which was calculated based on engine red line RPM. Both SAE coarse and SAE fine test dusts were used. Initial gravimetric efficiency was calculated after feeding 20 GMs of test dust. The test was terminated at a restriction rise of about 2.5kPa beyond initial restriction. The dust holding capacity and overall gravimetric efficiency was also measured at test termination. Overall, the test data has been compiled for about 70+ vehicles. The competitive comparisons for the three different filtration technologies evaluated are presented below.

For the analysis presented, one-on-one comparisons are difficult as each air induction system was unique. Normalized data have not been presented purposely in order to capture the effect of air cleaner design. The main purpose for the analysis was to relatively distinguish the three technologies used and evaluate general performance levels.

Dust Capacity Comparisons - The dust capacity is a measure of how long the air filter will survive before service is required. As the data will show the overall dust capacities measured were in the range of about 40 GMs to about 450 GMs depending on the type of test dust used. Further, almost all of the air filters evaluated had a recommended service life of 30K miles under normal driving conditions regardless of the type of media used.

The dust capacity measurements are presented in figure 2. The figure shows the average values and the range (minimum & maximum values).

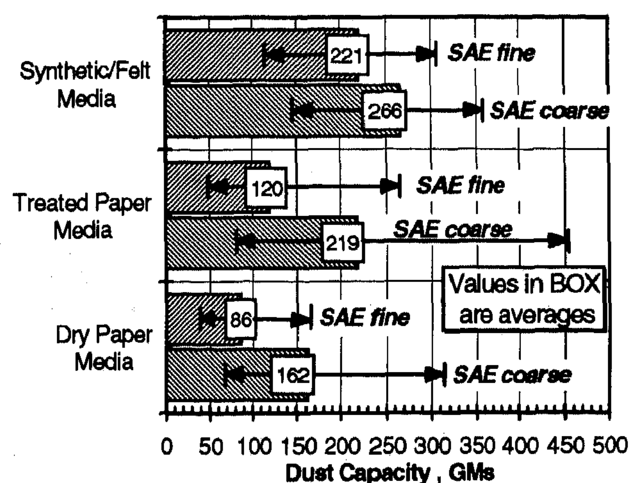


Figure 2:Typical Dust Capacity Performance Levels of AIF. Both Average and Ranges are shown.

On an average for both dry and treated paper media the dust capacity measured using SAE fine test dust was about half the dust capacity measured using when SAE coarse test dust (1,32). The synthetic media on an average showed only a 20% reduction in dust capacity when using SAE fine compared to SAE coarse. The synthetic media due to its depth loading and gradient density characteristics may be less sensitive to dust size distributions. On the other hand the dust loading capacity for the cellulose paper media depends on the formation of a dust cake on its surface. Studies have shown that the dust cake formation is strongly dependent on the face velocity and type of dust size distributions.

The data also shows that on an average the synthetic media when using SAE coarse dust collected about 22% more dust when compared to treated paper

media technology and about 64% more dust when compared to dry media. Similarly, using SAE fine dust the synthetic media collected about 85% more dust compared to treated paper media technology and about 2.5 times more dust when compared to dry media.

Overall, the synthetic media does show a competitive edge in dust holding capacity especially using SAE fine dust when compared to the other two. It could be argued as to which dust (SAE coarse or SAE fine) may be more representative of field conditions. Field evaluations performed by the author and other researchers indicate that SAE fine test dust may be more appropriate to evaluate engine air induction filters.

For normal driving conditions a dust holding capacity of about 100 GMs is generally recommended for a service life of 30K miles. Based on some of the field data the author has analyzed, this may be a very conservative requirement (1). The data shows that the majority of the filters evaluated will meet or significantly exceed this expectation. Field data analysis will be presented at a later date.

Initial Efficiency Comparisons - Initial efficiency is the key filtration performance measure to determine how well the air filter is protecting the engine (1,2). It has been well demonstrated that the initial efficiency of the filter is significantly lower when the filter is new and clean. The efficiency generally builds up after the first 5 - 15% of its useful life (loading) (31,35). Studies have shown that there is a very strong relation between filter removal efficiency and engine wear (2,3,31,35).

The initial efficiency measurements are presented in figure 3. The figure shows the average and the overall range for each media type. The initial efficiencies were measured after feeding 20 GMs of test dust.

When using SAE coarse dust, the treated paper media may show the best initial efficiency when compared to both dry and synthetic media. On an average using SAE coarse dust the treated paper media measured about 4 times lower dust penetration compared to dry paper media and about 2 times lower dust penetration compared to synthetic media. This is significant in terms of protecting the engine. When using SAE fine dust both the treated and synthetic media showed similar performance levels. On an average using SAE fine dust the treated paper and synthetic media measured about 2 times lower dust penetration when compared to dry media.

On an average the dry paper and synthetic media showed little differences in initial efficiency when using SAE coarse or SAE fine test dusts. On the other hand the initial efficiency for the treated paper media shows a strong dependence on the type of test dust used.

The data also shows that a majority of the dry paper media evaluated had significantly lower initial efficiencies. This is a concern as such lower initial efficiencies will induce excessive engine wear. The same was also true for treated media and to a lesser extent for synthetic media. For the dry paper media, efficiencies as low as 91% to 94% were measured using SAE coarse and fine test dust. On the other hand the lowest initial efficiencies measured for the treated and

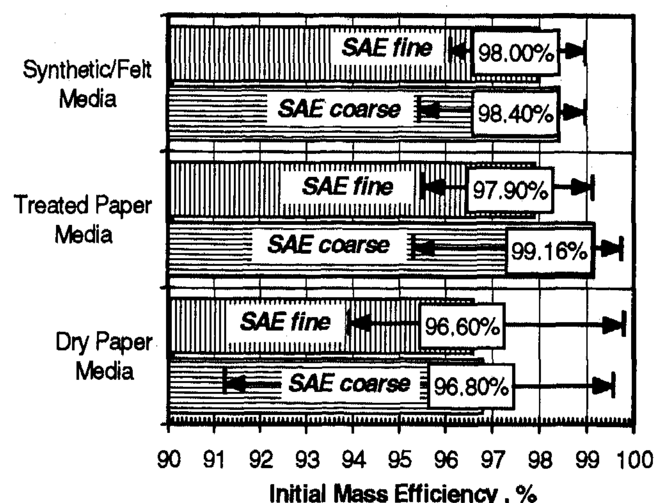


Figure 3: Typical Initial Efficiency Performance Levels of AIF. Both Average and Ranges are shown. Values in box are averages.

synthetic media was in the range of about 95.5% - 96.5% depending on the type of dust.

It is also interesting to note that the dry media when using SAE fine dust showed a higher initial efficiency performance range compared to when using SAE coarse dust. This may happen when the filter media are exposed to high face velocities. At such high face velocities the fine dust builds up relatively compact dust cake compared to coarse dust (35). The build up of the compact dust cake helps in enhancing the filter performance.

For normal driving conditions it is recommended for engine AIF to have a initial gravimetric efficiency in the range of about 98%+ to adequately protect the engine. Based on this requirement (assuming SAE test dust to be field representative) a majority of the filters evaluated did not meet the performance criteria and may not protect the engine. This was true for all three media evaluated. For off- road or dusty environments none of these filters may adequately protect the engine.

Overall, the treated paper media performed better than the dry or synthetic media.

Overall (final) Efficiency Comparisons - The overall efficiency is a measure of how well the air filter will protect the engine throughout its service life. The overall efficiency needs to be high to protect for re-entrainment, bounce, etc. under high dust loading (1,2). Generally the overall efficiency is significantly higher than the initial efficiency and is strongly dependent on dust loading characteristics of the media.

The measured overall efficiencies are shown in figure 4. On an average both the treated paper and synthetic media showed higher overall efficiency levels when compared to the dry paper media. Here also (on an average) the dry paper showed little differences in overall efficiency levels when using SAE coarse or SAE fine test dusts. This was similarly true for the synthetic media.

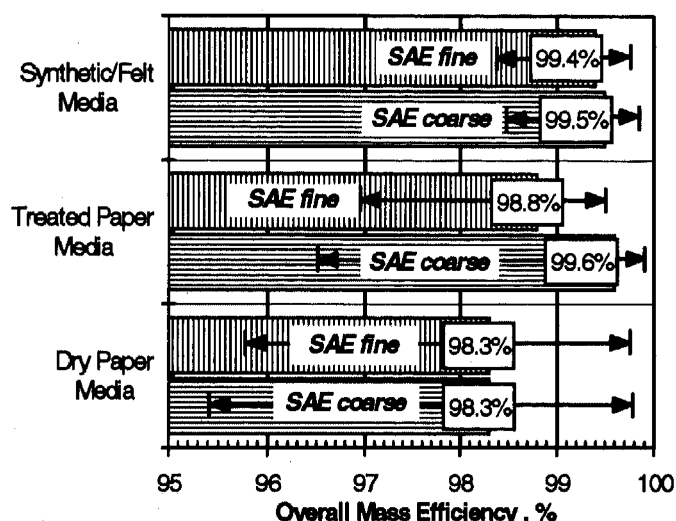


Figure 4: Typical Overall (final) Efficiency Performance Levels of AIF. Both Average and Ranges are shown. Values in box are averages.

For both the dry and treated paper media the overall penetration (1-overall efficiency) values dropped by a factor of ~2 when compared to initial penetration (1-initial efficiency) values. Similarly, for the synthetic media the overall penetration dropped by a factor of ~3.3 when compared to initial penetration.

The spread or the range of data collected also reduces significantly when comparing overall efficiencies to initial efficiencies. This was very prominent for the synthetic media.

The recommended overall efficiency for normal driving conditions is generally 99%+ to adequately protect the engine. Based on this criteria we see that a majority of the filters evaluated may not meet this requirement. This was especially true for the dry and treated paper media.

Overall, the synthetic media seems to perform only slightly better than the treated paper media. On the other hand, both the treated paper and synthetic media performed significantly better than the dry paper media.

Media Face Velocity Comparisons - There is a strong dependence of the media face velocity to its filtration performance. The face velocity of the media affects;

- filtration system restriction
- dust removal efficiency
- dust cake formation
- dust capacity and life

The overall average media face velocities and the corresponding spread (range) are shown in figure 5. The media face velocity was calculated using the following equation;

$$\text{media face velocity} = \frac{\text{total media area}}{\text{rated engine maximum flow}}$$

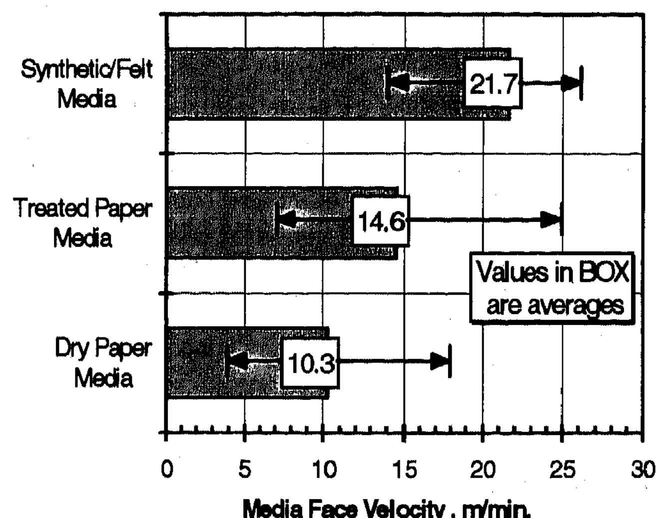


Figure 5: Typical Media Face Velocities Measured For AIF. Both Average and Ranges are shown.

Measurements show that the synthetic media operated at the highest face velocities. By comparison (on an average) the synthetic media operated at about twice the face velocity when compared to dry paper media. Similarly, the synthetic media operated at about 1.5 times higher face velocity when compared to treated paper media.

The average face velocity measured for the dry paper media was about 10.3 m/min. (~34 fpm) and for the treated paper media was about 14.6 m/min. (~48 fpm). The average face velocity measured for the synthetic media was about 21.7 m/min. (~71 fpm). On the other hand the media face velocities for heavy duty applications (medium/heavy duty trucks, trailers, off-road vehicles, agricultural applications etc.) are generally in the range of about 0.9 - 3.0 m/min. (3.0 - 10.0 fpm). By comparison the media face velocities for passenger car/light truck engine AIF about 4 - 16 times higher compared to heavy duty applications. One of the main reasons for this difference is the packaging constraints and environment.

Studies have also shown that filter media are dramatically affected by their face velocity. There is a critical velocity beyond which the filter media should not be used (19). Determining the critical velocity is important in order to effectively design air filters(19).

Discussions - For fibrous filter media the pressure drop rise is directly related to the amount of dust collected within the media and the formation of the dust cake (32). The pressure drop of the media remains low when dust is being collected in the media regime (32). This may be the reason why the synthetic media evaluated showed a relatively high dust loading per unit area for the same terminal conditions. Another way to enhance the dust holding capacity without any pressure drop penalty is to apply a treatment on the media (1).

Typically, we find that the initial and overall filtration efficiency levels for passenger car and light truck air intake systems are significantly lower when compared to heavy duty applications. For heavy duty air intake filters (highway trucks, agricultural vehicles, mining vehicles etc.) the initial efficiencies are in the range of about 99.5% to 99.9%+ and overall efficiencies are in the range of 99.9% to 99.99%+. In addition, majority of the heavy duty air intake filtration systems are also equipped with a pre-separator device. The pre-separator may be stand alone or be an integral part of the filtration system. Such pre-separators may have dust removal efficiencies in the range of about 70%-90% depending on flow rates.

The reasons for these performance differences (efficiency) are not very clear. In our opinion the passenger car/ light truck engine air filtration systems should have the higher efficiency levels to provide better engine protection. One reason for the high efficiency levels of heavy duty filters may be due to their lower media face velocities. Because of the lower face velocities a tighter media (lower permeability) could be used without any appreciable pressure drop penalties. Also the service intervals on heavy duty applications are significantly longer compared to light duty.

ENGINE PROTECTION

Dirt is the most common cause of high metal wear in engines (13). Dirt ingested by the engine eventually ends up in the oil as silica. The recommended maximum amount of silica or dirt in the oil should be less than 5ppm for 4 cylinder engines and less than 10ppm for 6 cylinder engines, when the engine is operated for 50hrs or 2500 miles (13). Efficient combustion and well filtered air are essential to protect engines from wear and also extend oil drain intervals (12).

The key function of a good engine AIF is to provide optimum protection from abrasive contaminants. In real world conditions it could be very challenging for the engine air filter to meet and exceed protection requirements. Selection of proper air and oil filters is very critical for engines designed today (2).

Instead of just focusing only on the engine air filter, engineers and designers should focus more on the overall filtration system to protect the engine. The two key filtration systems for engine protection are the engine intake and the lube oil systems. We may need to look at the interactions of the air and lube filtration systems for optimum engine protection. There is a certain synergy between the design of oil and engine air filters. Because what passes through the engine air filter system also affects the lube oil filter system.

It is beyond the scope of this paper to discuss lube oil filter system or its interaction with the engine intake filtration system. Although occasional references have been made as appropriate. Barris (2) also talks about a 'Total Filtration' approach for heavy duty filters and is an excellent paper.

A well designed air induction filtration system will deliver the following;

- Remove and reduce abrasive contaminants
- Reduce and control the ingested dust concentration levels within acceptable limits
- Control and remove contaminant sizes most harmful to the engine
- Reduce overall engine wear
- Reduce oil contamination
- Reduce blow-by conditions
- Control emissions
- Increase engine life and durability
- Increase lube oil service interval
- Reduce overheating of the engine
- Smooth engine performance

Fodor (3,4) in his study found that about 80% of all engine failures were attributed to engine wear. It is absolutely necessary to have improved filtration for air and oil filters to reduce engine wear (4).

Both, contaminant concentration and size ingested by the engine are important to reduce wear and

improve engine durability and life. It has been demonstrated that by using high efficiency air and oil filters, engine wear can be significantly reduced (16). Studies have also shown that engine life is a function of induction air cleanliness (19). It has also been established that cleanliness of an engine is a pre-requisite for problem free operation (7). Further, the piston ring cleanliness is important to ensure a free piston ring movement in clean ring grooves to ensure proper control of emissions and oil consumption (7).

Unfiltered air can induce significantly wear in an engine generally referred to as dusting out (17). Dirty air intake is the main cause today for wear on piston rings, pistons, cylinder liners, valves, bearings etc. (17).

Engine wear was reduced by about 70% going from a 40 micron to a 15 micron lube oil filter because it reduce the number of abrasive particles by a factor of 2 to 3 (8).

Today, there is a growing trend to deliver engines claiming to have extended service life and durability up to 150K miles or 10 years. In order to achieve this the engines have to be well protected from abrasive contaminants, oil additives etc. and also have better and cleaner manufacturing processes.

CONCENTRATION REQUIREMENTS - Controlling the amount of dust ingested by the engine is very critical for engine protection. In real world (field) conditions, the ambient dust concentrations can vary from < 0.1mg/m^3 to > 1000 mg/m^3 depending on the environment and driving conditions (1,2,5,18). For normal (typical) driving conditions the dust concentrations may be in the range of about 0.1 mg/m^3 to 0.5mg/m^3. In addition, the dust concentrations actually seen by the filtration system may also depend on the inlet tube location and design underhood. Higher dust concentrations also require higher efficiency levels to adequately protect the engine. Often the engine AIF are designed for normal to typical driving conditions. In a dusty environment such filters may not adequately protect the engine. This causes excessive wear and irreversible damage to the engine. Fodor (3) indicated that the air cleanliness ingested by the engine should be < 0.01mg/m^3 to reduce engine wear.

Studies have also indicated that on highways and paved roads diesel engine exhaust or soot is one of the dominant source of airborne particulates (5). Typical average ambient dust concentrations (TSP) can range from 0.03 -0. 3 mg/m^3 (5).

Figure 6 shows a plot of ambient dust concentration levels (mg/ m^3) and the required air filter mass efficiency levels. The required efficiency (initial efficiency) was calculated to meet a minimum downstream dust concentration of 0.01mg/m^3 ingested by the engine. The plot also shows three different ranges;

1. Normal driving conditions (0.1 - 1 mg/ m^3)
2. Dusty Conditions (0.5 - 50 mg/ m^3)
3. Very Dusty Conditions (10 - 1000+ mg/m^3)

For normal driving conditions like rural highways, urban highways, country roads, residential areas etc., in order to meet or exceed the downstream criteria we may require the engine air filter to perform in the range of about <95% to 99.0%+. For an average concentration of about 0.5 mg/ m^3 the required initial efficiency may be ~98%. This is probably a more conservative approach.

Similarly, for dusty conditions like congested urban highways , congested city roads , industrial areas , unpaved roads etc. the required efficiency may

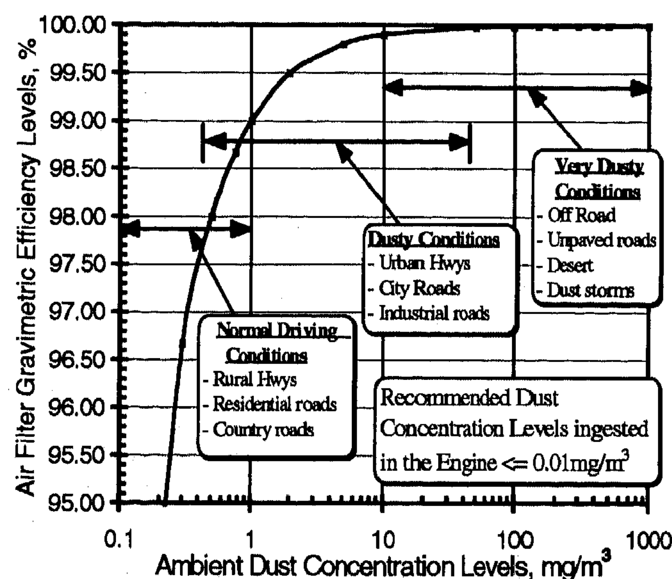


Figure 6: Calculated Engine AIF Gravimetric Efficiencies (initial) based on Ambient Dust Concentration Levels.

be in the range of about 99.0% to 99.98%+. This type of driving would be encountered in congested and busy cities and towns. Assuming an average concentration of about 1.0 mg/ m^3, the initial efficiencies required would be about 99%+.

For very dusty conditions like dust storms, mining operations, agricultural areas etc. the required efficiency levels may be in the range of about 99.98% to 99.999%+.

CONTAMINANT SIZE REQUIREMENTS - It is very critical to remove or control the different contaminant sizes being ingested by the engine. It is also critical to determine the contaminant sizes most harmful to the engine. Many studies have been performed and a variety of technical data is available.

Maximum wear is caused by particles in the size range of about 5 - 15 µm (6, 11). Piston oil ring wear is most sensitive to very small particles (6). It was also determined that dirt and abrasive particles equal to or less than the oil film thickness separating the various sliding or rotating parts will cause excessive wear (13). Even real fine particles (< 1 µm) may cause excessive wear, if their concentrations are high enough to agglomerate (13).

On of the most common wear mechanisms observed on piston ring and cylinder bore is abrasion wear caused by hard particles in the oil film originating from the combustion zone (14). Particles having diameters equal to or larger than the oil film thickness will make contact with the surfaces, initiating wear and excessive friction loads (9). Further, the wear debris adds to the contamination present in the oil causing additional wear on sliding/rotating surfaces (9).

The oil film thickness for any engine may be unique by design and performance requirements. Studies have shown that measuring various oil film thickness may indicate the different sizes of contaminants that may cause excessive wear.

Typical oil film thickness and clearances found in engines are listed in table 1.

The data from table 1 shows that abrasive particles down to 1µm or even smaller can cause engine wear, especially between piston ring and cylinder wall clearances. Investigations have also shown that during the power stroke or high load conditions the piston ring - cylinder wall gap could be reduced to almost zero (10, 30). This tells us that it is important and necessary to control or remove abrasive particles down to very fine sizes. This has also been supported by experimental measurements and investigations (1, 6, 8, 11, 15). The air filters designed for engine AIF may need to remove or reduce fine particles down to 1µm.

Figure 7 shows a typical efficiency curve for a engine AIF. The fraction size efficiency for the air filter was measured using SAE ultrafine test dust. The procedure used was similar to the SAE J1669. The filter media used was a typical cellulose fiber paper media generally used for engine intake filtration. The initial clean efficiencies for the filter have been plotted at three different face velocities of 3.5 m/min., 7.0 m/min. & 10.5 m/min.. The particle size efficiency was also measured after loading the filter using SAE fine test dust (~ 115 GMs loading).

Piston skirt - cylinder wall /liner	~ 25 µm - 100 µm *
oil film thickness between piston ring - cylinder wall	< 1µm - 15 µm **
piston - pin bosses	~ 8 µm - 15 µm ***
other bearing film thickness	~ 2 µm - 22 µm ****

*** ref. (33)**
**** ref. (20, 21, 22, 23, 24, 25, 26, 27, 28, 30)**
***** ref. (33)**
****** ref. (8, 29)**

Table 1: Typical Oil Film Thickness and Clearances Found in Engines.

Data in figure 7 shows that the removal efficiency for 1µm particle was measured in the range of about 40% - 70% depending on the face velocity. The higher face velocities also measured higher removal efficiencies. This is because of higher inertial and interception parameters at higher face velocities. For a 1µm particle size the initial efficiency measured at 10.5m/min. was about 70%. For a 5µm particle size, the

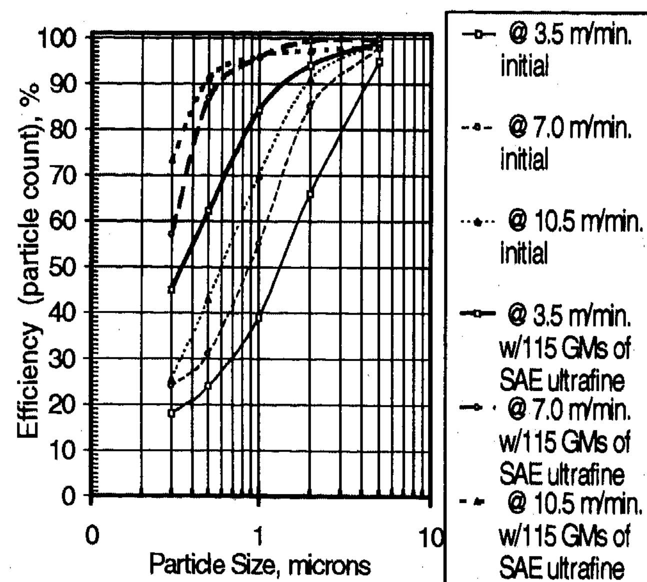


Figure 7: Typical Fractional Efficiency v/s Particle Size for a Dry Paper Media at Various Face Velocities with & without Loading.

the initial collection efficiencies measured were in the range of about 95% - 98% depending on the face

velocity. On the other hand for a 0.3μm particle size the face velocity had little effect on the clean efficiency of the filter.

There is also evidence that at the high face velocity of 10.5m/min., re-entrainment and bounce may be dominant due to a slight drop in efficiencies measured for particles > 1μm.

The filter after being loaded with 115 GMs of SAE ultrafine showed a significant increase in the collection efficiencies. This was especially true for the particle sizes in the range of about 0.3μm to 3.0μm. Beyond 3.0μm size the loading had a marginal effect on the removal efficiency. After loading the filter the particle removal efficiencies measured were in excess of 90%+ for 0.6μm and above. Overall the data shows that the clean filter does allow a significant amount of dust < 5.0μm to penetrate and be ingested by the engine.

Overall after loading , the dust penetration significantly reduces and is lower by a factor of 5 - 10 when compared to initial penetration.

FIELD EVALUATIONS

Field evaluations are mandatory to design and develop robust engine air induction filtration systems. There is definitely a need to understand relationships between laboratory and real world conditions , which is generally a missing link.

Field evaluations are generally complex involving dynamic changes. A variety of driving and environmental conditions will be encountered in field evaluations. Instruments and equipment used in the field need to be robust and should be adaptable to changing conditions. In our opinion such measurements do generate useful data and can be an excellent development tool.

Contaminants generated in the field will vary with the environmental and geographic conditions (34). Contaminants generated in the field will vary in ;

- size distributions
- various shapes
- wide concentration ranges
- material properties
- sources and source generation
- contaminant types

We generally evaluate engine air filters using dry standard test dusts. This may or may not represent field conditions, as contaminants in the field vary. A variety of field contaminants / aerosols potentially ingested by the engine AIF are listed below in table 2 (36, 37);

Types	Typical size ranges
solids - abrasives - dirt - road dust	~0.001 - 100μm +
sooty type aerosols	~0.001 - 1 μm
oils - sticky hydrocarbons	~0.01 - 10 μm
fine metal debris	~0.01 - 10 μm
industrial dusts - generally fine	~0.001 -100 μm
agricultural dusts	< 1.0 - 1000 μm
atmospheric dust	< 0.01 - 20 μm
mining dust	~0.1 - 1000 μm
sand	< 100 - 1000 μm+
water - rain	~50 - 10000 μm
salt crystals	~0.01 - 1.0 μm
snow	large
mud	NA

Table 2: Typical Contaminant Size Distributions

FIELD MEASUREMENTS - Engine air induction filter were evaluated at Arizona Proving Grounds (APG). The main objectives of the field measurements were;

- Capture the real time performance of engine air filters.
- Measure the dust concentrations and size distributions challenged to the engine air filter.
- Measure the dust concentrations and size distributions ingested by the engine.

The tests performed at APG were accelerated tests to evaluate the performance of engine AIF. It is beyond the scope of the paper to discuss details of the measurement techniques used to perform these measurements. Nevertheless some interesting data have been presented on the performance of engine intake air filters.

Test Setup - To capture the real time mass efficiency of the air cleaners, photometers were used upstream and downstream of the filter. By using photometers we were able to improve the sensitivity of measurements. The samplers were designed to be iso-kinetic, but due to the variable engine flow iso-kinetic sampling was not practical. The sampling line losses were kept to a minimum by shortening the lines. Absolute membrane filters were used to collect dust and later analyzed under a electron microscope for

calculating size distributions. All the tests were performed on the dirt track where a lead truck was used to generate a cloud of dust. The test vehicle followed the lead truck always surrounded by a cloud of dust.

Results & Discussions - Figure 8 shows a real time curve of measured mass efficiency as a function of time. Two averaged curves are shown, one measured at APG and the other was generated in the lab using the same photometers. Both curves have been plotted on the same time scale. It is very interesting to observe the low initial efficiencies measured at APG and on the bench. Initial efficiencies as low as about 65 %-70% were measured at APG. The lab measurements measured initial efficiencies in the range of about 50% - 60% . These efficiencies are significantly lower when compared to the initial efficiencies in figure 3.

The air filters were loaded up to a terminal restriction rise of ~2.5kPa. We also see that with loading the mass efficiency quickly builds up as expected. The filter tested in the lab loads up faster compared to APG due to higher dust concentrations used. Further, the filter in the lab was evaluated at maximum engine flow conditions. In comparison the filter tested at APG was exposed to much lower variable flow rates. For both

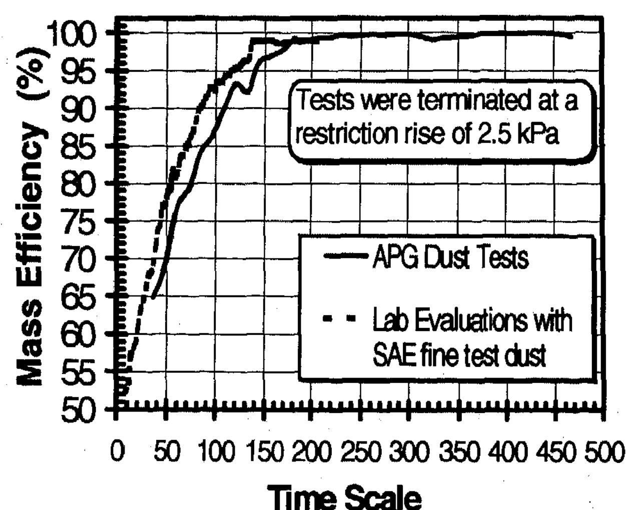


Figure 8: Typical Mass Efficiency of a Dry Paper Filter with Dust Loading at APG (Field Evaluation).

tests, the final efficiencies were measured in the range of about 98% - 99%. The upstream concentration measured during the APG tests were in the range of about 50 - 100 mg/ m^3. The downstream concentrations measured for the APG test was about 11mg/m^3 initially. With loading these dust concentrations significantly reduced to about 0.015 mg/ m^3. These numbers on an average translate to about 85%+ initial mass efficiency and about 99.9%+ final efficiency. APG tests showed that the air filter evaluated had a low initial efficiency and may not adequately protect the engine.

Figure 9 shows the measured upstream dust size distribution (number distribution) that was challenged to the air filter. Majority of the dust particles were measured in the range of about << 1.0µm - 50µm. Dust particles as large as 80µm were also measured. Data also shows that more than 95% of the dust particles were in the range of << 1.0µm - 15µm (particle sizes most harmful to the engine).

Similarly figure 10 shows the downstream dust concentration ingested by the engine. Majority of the particles ingested by the engine were in the range of << 1.0µm - 5.0µm. The filter effectively removed everything above 10µm. On the other hand more than 95% of the dust particles ingested by the engine were in the range of << 1µm - 4µm. Fraction size efficiency measurements in figure 7 also show similar results, correlating to an extent the field and lab results . These particle sizes can potentially cause abrasive wear in the engine.

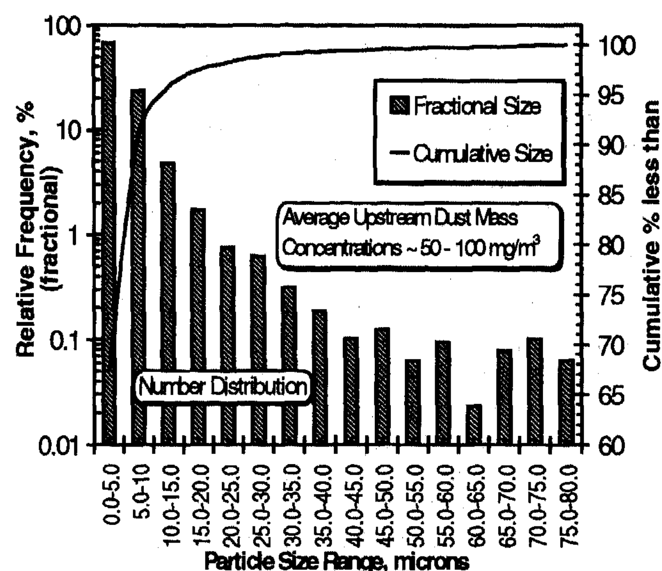


Figure 9: Typical Particle Size Distribution Measured Upstream of the AIF.

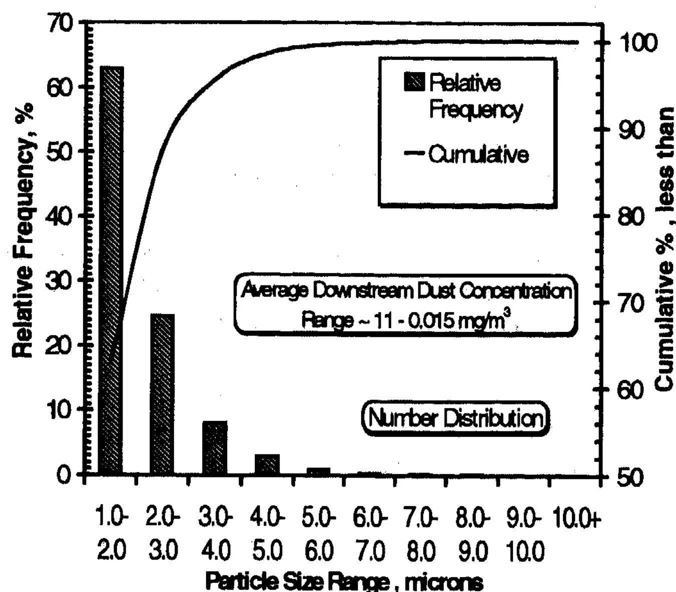


Figure 10: Typical Particle Size Distribution Measured Downstream of the AIF (ingested by the engine).

PROPOSED IMPROVEMENTS AND RECOMMENDATIONS

Studies and field evaluations have clearly established that in order to protect the engines and reduce engine wear, the ingested contaminant concentrations and size distributions need to be controlled. The current SAE J726 procedure does not provide any information on dust concentrations or size distributions. Gravimetric or mass efficiencies currently measured may not be adequate to understand and design effective engine air induction filters.

There is a definite need to quantify downstream conditions by defining acceptable limits on ingested contaminants. Additional studies may be required to selectively quantify what particle sizes are most harmful to the engine. Further, we also need to know the acceptable concentration limits of these harmful particles sizes that may be ingested by the engine. The following improvements are recommend that may add value in designing robust engine air induction filters;

◊ **Identify a Representative Test Aerosol/Dust** - We currently use the SAE coarse, fine, medium and ultrafine tests dusts to evaluate filters. Studies have shown that in real life there may be a distinct BI-modal aerosol distribution present (18, 38). One study showed that the a BI-modal distribution exist with a mass mean radii of about 4μm and 45μm (18).

◊ **Specify Downstream Conditions** - In order to protect the engine, the downstream conditions of the AIF (acceptable ranges) may need to be specified. The downstream conditions may include both concentration requirements and/or size distributions.

◊ **Measure Fractional Size Efficiency** - There is a need to measure how well the engine AIF removes particles that are most harmful to the engine. Fractional efficiency requirements may be specified in terms of number or mass.

◊ **Flow rates** - There may be a need to identify realistic flow rates to evaluate engine air filters for specific engine operations. Many a times the air filters are evaluated at maximum engine rated flows. This may be unrealistic. Using the SAE variable flow cycle to evaluate engine air filters may yield more meaningful data. This may be especially true for filtration systems incorporating inertia separation devices. At low flows (realistic) these separation devices operate with significantly lower efficiency levels.

Some key factors affecting the fractional efficiency of engine air filters are also briefly summarized below;

- type of contaminants - dirt, sand, soot etc.
- contaminant concentration levels
- driving conditions
- contaminant size distributions
- interaction of media with environmental conditions
    - heat
    - humidity
    - water
    - salt
    - snow
    - mud splashes
- type of filter media used
    - paper type - surface loading
    - multilayered - gradient density - depth filtration
    - porous depth media
    - true depth filters
- media face velocity
- pleat design - size and shape
- filter media treatments - oils etc.
- air filter design
- air cleaner housing design - inlet/outlet configurations
- effects of vibrations on filter media
- shelf life requirements

- effect of pre-separation (inertia type) devices
- others

FUTURE TRENDS

Design of current and future engine air induction systems are being driven by customer needs and wants. Engine air intake filtration being an integral part of the air induction system will also need to meet customer requirements. Future air induction filtration designs will be challenged to meet new designs, packaging constraints, new technologies, performances, manufacturing and cost targets. Some future challenges may include;

- Finer filtration may be required. This will apply to both engine air and lube oil filters.
- Extended service life on both engine and lube oil filters. Service life greater than 100K miles may be achievable on engine air induction filters - AIF.
- Packaging constraints are driving designs to be very compact. Compact canister type filter designs may be the way to go.
- Filter media designs may have to handle higher face velocities.
- Higher underhood reliability and robustness to meet temperature and durability requirements.
- Systems integration involving mass air flow sensors, resonators, various other sensors - ACT...., battery cool box, etc..
- Generic designs to save costs
- Overall system cost savings
- Green filters / recyclability
- Provide higher value to the customer.

SUMMARY

- Higher initial collection efficiencies are required to reduce engine wear.
- Lab and field evaluations show that the initial collection efficiencies of engine air filters may be significantly lower than expected. This may cause excessive engine wear. Efficiencies as low as 60% were measured.
- For normal to dusty driving conditions initial gravimetric (by mass) efficiencies in the range of about 98% to 99.5% may be acceptable.
- Higher initial gravimetric efficiencies in excess of 99.9%+ may be required for very dusty environments.
- The overall/final efficiencies should be higher than the initial efficiencies. Lower overall/final efficiencies would indicate re-entrainment problems.
- Quantifying the engine air filter performance by measuring the fractional size efficiency may add value. This may be added as an option in the SAE J726 procedure.
- Based on oil film thickness and bearing clearances abrasive particles in the size range of < 1.0μm to 20μm may cause excessive engine wear.
- Field and lab evaluations show that for a typical engine air filter there is a significant penetration of particles in the range of < 1.0 μm - 5 μm. Such fine particles may be harmful to the engine.
- Understanding ingested contaminant size distributions, mass concentrations and related engine wear is important.
- The dry paper filters that were evaluated may or may not provide adequate engine protection.
- Overall, treated and synthetic media filters may provide better engine protection.
- The standard SAE fine test dust may be more representative of field conditions
- We need to understand the synergy between the engine air and lube oil filters.

REFERENCES

1. N. Bugli, R. Puckett, and V. Lanier, " Filtration Challenges and Conical Filter Development for Engine Air Induction Systems, " SAE technical paper 950941, Detroit, February 27 - March 2, 1995.

2. M. Barris, " Total Filtration™: The Influence of Filter Selection on Engine Wear, Emissions, and Performance," SAE technical paper 952557, Fuels & Lubricants meeting and exposition, Toronto, October 16-19, 1995.

3. J. Fodor, " Improving Utilization of Potential I. C. Engine Life by Filtration," TRIBOLOGY International 12, pp. 127-129, June 1979.

4. J. Fodor, " Improving the Economy of I.C. Engines by Controlling the Contaminants Through Filtration," World Filtration Congress III, pp. 707-711, 1982.

5. William. R. Pierson and Wanda W. Brachaczek, " Particulate Matter Associated with Vehicles on the

Road," SAE technical paper 760039, February 23-27, 1976.

6. F. L. Milder, A. J. Armini and G. W. Jones, " The Use of Surface Layer Activation Wear Monitoring for Filter Design and Evaluation ," SAE technical paper 810329.

7. Rudolf Thom, Karl Kollman, Wolfgang Warnecke and Mike Frend, " Extended Oil Drain Intervals: Conservation of Resources or Reduction of Engine Life," SAE technical paper 951035, February 27 - March 2, 1995.

8. David R. Staley, " Correlating Lube Oil Filtration Efficiencies with Engine Wear," SAE technical paper 881825. Truck & Bus meeting & exposition, November 7 - 10, 1988.

9. Brian W. Schwandt and Barry M. Verdegan," Influence of Lube Oil Filter Performance on Engine Wear in City Buses," SAE technical paper 902238, Truck & Bus meeting & exposition, October 29 - November 1, 1990.

10. L. L. Ting and J. E. Mayer, JR," Piston Ring Lubrication and Cylinder Bore Wear Analysis, Part I - Theory," Journal of Lubrication Technology (ASME transactions), pp. 305-314, 1974.

11. Mark J. Beauchamp," Application of High Efficiency Oil Filtration to Heavy Duty Diesel Bus Engines," SAE technical paper 920927, 43rd annual Earthmoving Industry Conference, April 7-8, 1992.

12. G. W. Holmes and R. Overton," Experiences with Extended Oil Drain Intervals in Diesel Truck engines," SAE technical paper 780958, International Fuels & Lubricants meeting, November 13-16, 1978.

13. Sacramento Sky Ranch Inc.," Engine Contamination and Wear," private notes.

14. T. S. Eyre, K. K. Dutta and F. A. Davis, " Characterization and simulation of wear occurring in cylinder bore of the internal combustion engine," Proceedings of the Institute of Mechanical Engineers - 894029, MEP-331, Second International Conference, September 19-20, 1989.

15. J. J. Truhan and C. B. Covington, " Effect of Filtration on Top Ring Face Wear in Heavy Duty Diesel Engines," CEC EL18 (coordinating European council), Fourth International Symposium on the Performance Evaluation of Automotive Fuels and Lubricants, Birmingham, UK, May 5 -7, 1993.

16. G. W. Jones and J. G. Eleftherakis, " Correlating Engine Wear with Filter Multipass Testing," SAE technical paper 952555, Fuels & Lubricants meeting and exposition, Toronto, October 16-19, 1995.

17. Robert N. Brady, " Modern Diesel Technology," handbook, Prentice Hall Inc. publication, ISBN: 0-13-288382-1, pp. 557-570, 1996.

18. R. G. Pinnick, G. Fernandez, B. D. Hinds, C. W. Bruce, R. W. Schaefer and J. D. Pendelton, " Dust Generated by Vehicular Traffic on Unpaved Roadways: Sizes and Infrared Extinction Characteristics," Aerosol Science and Technology 4:99-121, 1985.

19. Peter E. Sherburn, " Air Cleaner Design - Present and Future," SAE technical paper 690007, International Automotive Engineering Congress, Detroit, January 13-17, 1969.

20. Masaaki Takiguchi, Hajime Ando, Takahiro Takimoto, Akiyoshi Uratsuka, " Characteristics of friction and lubrication of two-ring piston," JSAE technical paper 9541362, JSAE Review 17 (1996) 11-16.

21. Yutaro WAKURI, Mitsuhiro SOEJIMA, Toyoharu TANIGUCHI, " On the Oil Film Behavior of Piston Rings (Correction of effective pressure region of oil film)," Bulletin of JSME, Vol. 21, No. 152, Paper No. 152-17, February 1978, pp. 295-302.

22. Jeffrey P. Lux and David, " Lubricant Film Thickness Measurements in a Diesel Engine Piston Ring Zone," Journal of the Society of Tribologists and Lubrication Engineers, Vol. 47, 5, May 1991, pp. 353-364.

23. Yoichi Watanabe, Akihito Uratsuka, Masaaki Takiguchi, Masaru Hiruma, Tsuneo Someya and Masaki Yamada, "Effect of Piston Ring Design on Oil Thickness," JSAE technical paper 9433696, Proceedings 943, Spring Convention , Yokohama, Kanagawa, Japan, May 1994

24. Shoichi Furuhama, Chikashi Asahi, Masaru Hiruma, " Measurement of Piston Ring Oil Film Thickness in an Operating Engine," technical paper at the ASME/ASLE Lubrication Conference, Washington D.C., October 5-7 1982, Preprint No. 82-LC-6C-1.

25. Keiichiro Shin, Yukio Tateishi and Shoichi Furuhama, " Measurement of Oil-Film-Thickness between Piston Ring and Cylinder," SAE technical paper 830068, International Congress and Exposition, Detroit, February 28 - March 4, 1983.

26. Dana E. Richardson and Gary L. Borman, " Measuring engine oil films," Automotive Engineering, Vol. 100, pp. 15-18, January 1992.

27. I. Sherrington, E. H. Smith, " Experimental methods for measuring the oil-film thickness between the piston ring and cylinder-wall of internal combustion engines," Tribology International, Vol. 18, N 6, December 1985, pp. 315-320.

28. Dana E. Richardson and Gary L. Borman, " Theoretical and Experimental Investigations of Oil Films for Applications to Piston Ring Lubrication," SAE technical paper 922341, SP-936, International Fuels & Lubricants Meeting, San Francisco, October 19-22, 1992.

29. David R. Staley, " Correlating Lube Oil Filtration Efficiencies with Engine Wear," SAE technical paper 881825, Truck and Bus Meeting and Exposition, Indianapolis, November 7-10, 1988.

30. D. G. Allen, B. R. Dudley, V. Middleton and D. A. Parker, " Prediction of Piston Ring-Cylinder Bore Oil Film Thickness In Two Particular Engines and Correlation with Experimental Evidence," Proc. I. Mechanical Engineering, technical paper C73/75, 1975, pp. 107-124.

31. Ken Conti, " Air Filters - Contractors Maintenance Notebook," Highway & Heavy Construction , August 1985.

32. Declan C. Walsh, " Recent Advances in the Understanding of Fibrous Filter Behaviour under Solid Particle Load," Filtration and Separation Journal, Vol. 33, No. 6: 206th issue, June 1996, pp. 501-506.

33. William H. Crouse and Donald L. Anglin, " Automotive Engines, " Seventh Edition, ISBN 0-07-014957-7, 1992.

34. G. B. Bessee and K. B. Kohl, " Comparison Between Real-Life Dust Samples and Standardized Test Dusts," SAE technical paper 940322 and SP-1040, International Congress and Exposition, Detroit, Michigan, February 28 - March 3, 1994, pp. 45-55.

35. Tad Jaroszczyk, John Wake and Mike J. Connor, " Factors Affecting the Performance of Engine Air Filters," Journal of Engineering for Gas Turbines and Power, Volume 115, October 1993, pp. 693-699.

36. Christopher Dikenson, " Filters and Filtration Handbook," 3rd edition, Elsevier Advance Technology, 1994, ISBN 1 85617 078 0.

37. William C. Hinds, " Aerosol Technology," 1982, ISBN 0-471-08726-2.

38. Kenneth L. Rubow, Benjamin Y. H. Liu and Jae-Keun Lee, " A Bi-Modal Test Aerosol for Filter Loading Studies," Proceedings of the American Filtration and Separations Society, Advances in Filtration and Separation Technology, Volume 9, 1995, pp. 98 - 109.

970557

The Paper-Tire Concept: A Way to Optimize Tire Force and Moment Properties

Dieterich J. Schuring
Consulting

Wolfgang Pelz
University of Akron

Marion G. Pottinger
Smithers Scientific Services, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A Paper-Tire is a virtual product defined as a set of force and moment equations that are entirely determined by well defined and widely recognized tire characteristics (such as cornering stiffness, peak location, and slide/peak ratio). The vehicle dynamicist may use the Paper-Tire concept to study the effects of tires with various hypothetical characteristics on vehicle behavior. If the dynamicist discovers a set of characteristics yielding a desired vehicle response, he could then ask tire manufacturers to attempt to develop a tire specification with the preferred characteristics. (There is, of course, no guarantee that tire manufacturers can develop a practical tire with the preferred characteristics.)

The paper explains the general principles of the Paper-Tire concept with the help of the BNPS model (derived from the 'Magic Formula'). A limited set of examples for tire cornering and braking performance are provided at a single load; further possible developments are indicated.

INTRODUCTION

In his daily modeling work, a vehicle dynamicist is often faced with the problem of optimizing the performance of a vehicle component, for example in the suspension or the steering system. This task is greatly eased if the component's performance equation can be expressed in terms of widely known and accepted parameters.

This paper facilitates the optimization process for tires, specifically their cornering and braking/driving capabilities. The optimization process is based on a model containing only universally recognized tire descriptors, such as cornering stiffness and slide/peak ratio; no additional information is necessary to describe the desired tire performance function. By varying the numerical values of these well defined and generally accepted descriptors rather than the values of some physically unspecified quantities (e.g., the constants of polynomial expressions), the effect of various tire characteristics on vehicle behavior can be mapped out rather quickly.

Since the numerical values of the performance descriptors are constructed 'on paper', the model is called the 'Paper-Tire'. By varying any or all of the independent descriptors the vehicle dynamicist is in a position to study their potential effects on vehicle behavior and select tire characteristics that yield the best vehicle response. The next step would be to ask tire manufacturers to develop tires with characteristics that come as close as possible to the selected values.

THE BASIC EQUATION

Of all the different models describing the cornering or braking/driving behavior of tires (physical models, black-box models, etc.), the Magic Formula seemed to be best suited for our purposes. The Magic Formula was first proposed in 1987 by Bakker, Nyborg, and Pacejka [1] and later altered and expanded in various ways. A thorough discussion was offered by Schuring, Pelz, and Pottinger [2], who also suggested a version of the Magic Formula that is suitable for personal computer processing (the BNPS model).

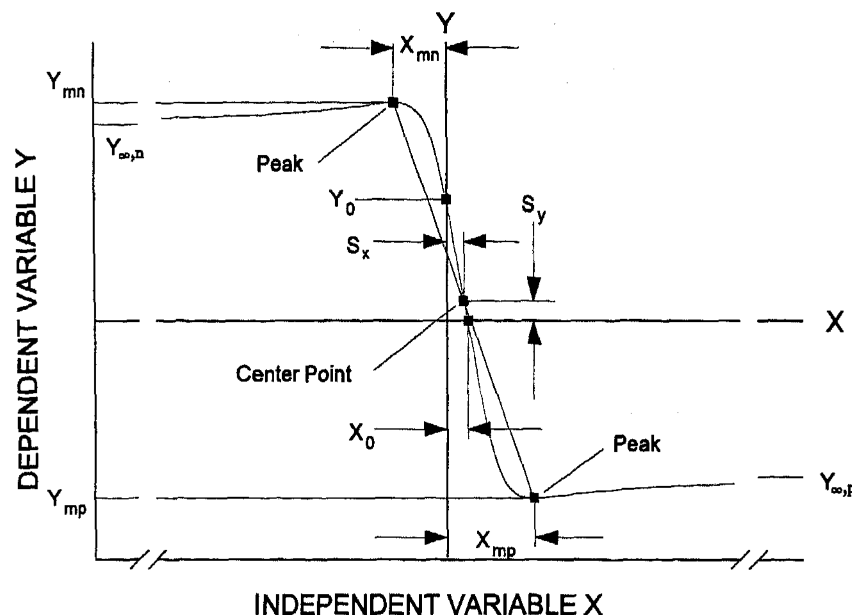


Figure 1. Typical Magic Formula Function, Y,(X)

The Magic Formula makes extensive use of trigonometric functions. Here it is expressed in the following form:

$$Y = S_y + D \sin[C \tan^{-1}\{B(X - S_x)(1-E) + E \tan^{-1} B(X - S_x)\}], \quad (1)$$

where the dependent variable Y represents any quantifiable tire response, such as a force or moment; and the independent variable X is a tire service variable, such as slip angle, inclination angle, or longitudinal slip.

The Magic Formula formula contains six constants -- S_x, S_y, B, C, D, and E (called MF constants in the following).

By varying these constants, a large variety of Y(X) functions with common and uncommon curve shapes can be generated [2]. However, all these curves regardless of their shape are composed of two symmetrical branches spreading from a center point (or center of gravity) toward positive and negative X-values, respectively. The coordinates of the center point are $X = S_x$ and $Y = S_y$ as given in Eq. (1) and indicated in Fig. 1, which shows a typical Y(X) relation between lateral force and slip angle produced by Eq. (1). Symmetry is provided by the fact that a 180 degree rotation of one of the two branches around the center point brings it in congruence with the other branch.

Restrictions- In this paper, the use of Eq. (1) will be restricted to S-shaped curves as shown in Fig. 1 with two peaks, one for the positive branch and a corresponding one for the negative branch. Each is followed by a smooth tail reaching a fixed value at infinite X values, as indicated in Fig 1. Furthermore, the center point of the two branches is situated near the origin of Y(X). These curves suit our purposes because they fully satisfy the formula's intended application to cornering and braking/driving tire forces and moments.

These conditions are realized if two restrictions are satisfied. First we demand that

$$1 < C < 3 \quad (2)$$

The mathematical background leading to this restriction is given in Ref [2].

Second, calling for a smooth S-type path for Y(X), as discussed, we require that

$$1 > E > -(1 + 0.5 \cdot C^2) \quad (3)$$

This relation is also discussed in [2].

Fig. 2 outlines the practical field of C and E given by Eqs. (2) and (3). For instance, for C = 0.9, E may range from 1 to -1.41; for C = 1.5, from 1 to -2.13; for C = 2.4, from 1 to -3.88; etc. Outside this range, Eq. 1 begins to describe rather uncommon shapes ('warp', 'hump', 'saddle') that are of no interest here.

The lower limit given in Eq. (3) can be generously overstepped without causing large errors, however. For instance, if the lower limit of E is selected at $E \approx -1.8(1 + 0.5 \cdot C^2)$ rather than at $E \approx -(1 + 0.5 \cdot C^2)$, a slight irregularity ('warp') of the slope dY/dX in the vicinity of the center point will occur with a maximum error of only 5 %.

CONVERSIONS

None of the six MF constants is a direct part of the commonly accepted nomenclature for tire force and moment performance. Hence the task at hand is to convert the six constants S_x, S_y, B, C, D, and E into six other, well-recognized and graphically demonstrable tire

performance descriptors. These six well-recognized descriptors must not only be independent from each other, their values must also cover the whole range of the performance function Y(X). These requirement are satisfied by :

1. The slope dY/dX ,S_t,at X = 0, also called stiffness (first constant). The stiffness governs the cornering and braking behavior at small slip values.

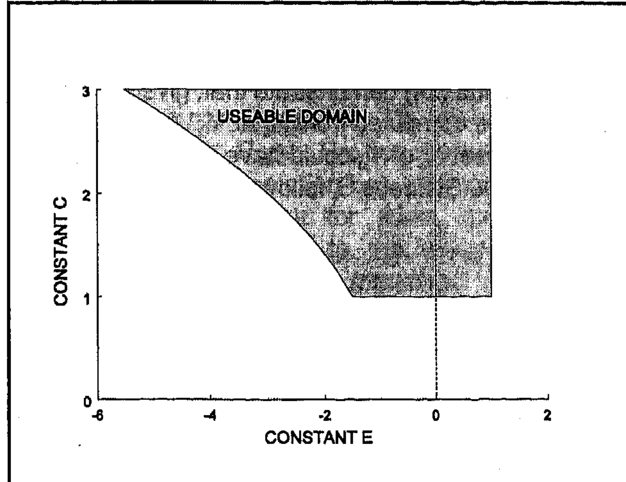


Figure 2. Field of Application for Constants C and E

2. The coordinates of the two peaks (four more constants). The peak values affect the cornering and braking behavior at larger slip values. The four coordinates of the two peaks are Y_{mp} and X_{mp} for the positive branch of X (index p); and Y_{mn} and X_{mn} for the negative branch of X (index n). Index 'm' stands for 'max' or 'min'.

3. The slide/peak (S/P) ratio (the sixth constant); that is the ratio of the ordinate at large X-values and the ordinate of the peak. This ratio is important for both cornering and braking at large slip values. The S/P ratio can be specified for either the positive or the negative branch of X. For the positive branch it is

$$R_p = Y_{\infty,p}/Y_{mp} \quad (4a)$$

for the negative branch it is

$$R_n = Y_{\infty,n}/Y_{mn}. \quad (4b)$$

Index ∞ stands for very large X-values. The S/P ratios of the positive and the negative branch are not independent from each other. Hence only one of the two can be picked arbitrarily.

The task is now to express the six constants of Eq. (1) by the six tire performance descriptors just given.

Center Point - Since the performance curve Y(X) is symmetrical, a straight line between the two peaks intersects the curve at the center point (see Fig. 1).

$$S_x= 0.5(X_{mp}+ X_{mn}) \quad ((5))$$

$$S_y= 0.5(Y_{mp}+ Y_{mn}) \quad ((6))$$

Hence by selecting the maximum and the minimum point of the desired Y(X) function, we have the coordinates of the center point. (Note: Equation numbers in double parentheses (( )) will be used later for final conversions.)

<u>Constant D</u> - Constant D is derived from the max/min values of Y. D is positive for a positive slope (stiffness) and negative for a negative slope (stiffness). The max/min values follow from the fact that the argument of sin [ ] in Eq. (1) cannot be larger than +1 at X > S_x, and -1 for X < S_x.

The max/min values of Y are therefore

$$Y_{mp}= S_y+ D, \quad \text{and} \quad Y_{mn}= S_y -D \quad (7)$$

Consequently,

$$D = 0.5(Y_{mp}- Y_{mn}) \quad ((8))$$

I.e., the constant D is half the vertical distance between the two peaks. Again, by selecting the max/min points of the desired Y(X) function, constant D can be easily computed.

<u>Constant C</u> - Constant C is determined from the S/P ratio, R, defined by Eqs. (4a) and (4b). The desired Y values at very large X-values are derived from Eq. (1) at X → (±)∞. The result is

$$Y_{\infty,p}= S_y+ D\cdot\sin(0.5\cdot\pi\cdot C), \quad \text{and}$$
$$Y_{\infty,n}= S_y- D\cdot\sin(0.5\pi\cdot\cdot C) \quad (9)$$

provided that E < 1, as stipulated in Eq. (3). These relations together with Eqs. (7) substituted into Eqs. (4a) and (4b) yield

$$\sin(0.5\cdot\pi\cdot C)=(1/D)[R_p(S_y+D)-S_y]=-(1/D)[R_n(S_y-D)-S_y] \quad (10)$$

If we abbreviate

$$R^*\equiv (1/D)[R_p(S_y + D)- S_y]= -(1/D)[R_n(S_y - D) - S_y] \quad (11)$$

we have

$$\sin(0.5\cdot\pi\cdot C) = R^* \quad (12)$$

so that as shown in the APPENDIX

$$C = 2[1 - (1/\pi) \sin^{-1}R^*] \quad ((13))$$

From this equation, constant C can be computed as a function of given values of R_p (or R_n), D, and S_y. If, as is mostly the case, S_y<<D, Eq. (11) simplifies to

$$R^*\approx R_p = R_n \equiv R \quad (14)$$

so that

$$C \approx 2[1 - (1/\pi) \sin^{-1}R] \quad (15)$$

Eq. (13) is valid for R*-values (or Eq. (15) for R-values) between +1 and -1, so that C values extend from 1 to 3.

Constant B - Constant B is derived from tire stiffness defined as the slope S_t= dY/dX at X = 0, as noted. To alleviate the differentiation of Eq. (1), we make use of our initial premise that the center point is in close proximity to the origin [X,Y = 0,0] so that, for the purpose of differentiation, $S_x = S_y \approx 0$. The first derivative dY/dX at X=0 then yields the stiffness

$$S_t \approx B \cdot C \cdot D \qquad (16)$$

With a numerical value of stiffness S_t selected, and with C and D computed by Eqs. (13) and (8), respectively, we determine constant B as

$$B = S_t/(C \cdot D) \qquad ((17))$$

Note that stiffness S_t may be positive or negative. It is positive if D is positive; and negative if D is negative. Hence B is always positive (as is C).

Constant E - Constant E is found at dY/dX = 0, as follows. Eq. (1) can be re-written as

$$y = D \cdot \sin(C \cdot z) \qquad (18)$$

with

$$z = \tan^{-1}\{B \cdot x(1 - E) + E \cdot \tan^{-1}(B \cdot x)\} \qquad (19)$$

where, temporarily,

$x = X - Sx$ and $y = Y - S_y$.

To find the max/min values of y, we impose

$$y' = z' \cdot C \cdot D \cdot \cos(C \cdot z) = 0 \qquad (20)$$

This equation is satisfied if either z' =dz/dx= 0 or cos(C·z) = 0 (C = 0 or D = 0 are trivial). A short investigation reveals that only the case of cos(C·z) = 0 need be considered here.

Hence we have (C · z) = (±) π/2 or z = (±) π/(2·C), so that substituting in Eq. (19)

$$(\pm)\pi/(2 \cdot C) = \tan^{-1}\{B \cdot x_m(1-E) + E \cdot \tan^{-1}(B \cdot x_m)\} \qquad (21)$$

The variable $x_m = X_m - S_x$ refers to the X-value of the peak for either the positive or the negative branch of X. From Eq 21 follows that

$$E = \frac{(\pm)\pi \tan[(/(2 \cdot C)] - B(X_m - S_x)}{\tan^{-1}[B(X_m - S_x)] - B(X_m - S_x)} \qquad ((22))$$

The positive sign stands for the positive X branch where $X_m = X_{mp}$; the negative sign, for the negative X branch where $X_m = X_{mn}$. In either case, the numerical value of E is the same.

Offsets X_0 and Y_0 To compute pull forces and pull moments, the offsets of Y at X = 0 and of X at Y = 0 must be known. Fig. 1 shows that the X-offset is (nearly)

$$X_o = S_x - (S_y/S_t) \qquad (23)$$

and the Y-offset

$$Y_o = S_y - (S_x \cdot S_t) \qquad (24)$$

Eqs. (23) and (24) demonstrate that, given S_x and S_y, stiffness S_t can be computed from X_o or Y_o. Hence instead of stiffness S_t, offset X_o or Y_o could have been specified as independent input. Stiffness S_t follows then from S_x, S_y, X_o or Y_o as

$$S_t = (S_y - Y_o)/S_x \qquad (25)$$

or

$$S_t = S_y/(S_x - X_o) \qquad (26)$$

COMPUTATIONS

We begin by choosing numeric values of the coordinates of the two peaks: for the positive branch, they are X_{mp} and Y_{mp}; for the negative branch they are X_{mn} and Y_{mn}. Also, numerical values of both stiffness, S_t, and S/P ratio, R, are selected.

These six performance descriptors are then converted into the six original MF constants in several steps. As a first step, constants S_x, S_y, and D are computed from the four peak coordinates, according to Eqs. (5), (6), and (8), respectively. Also, constant C is calculated from the S/P ratio R by Eq. (13) or (15).

As a next step, constant B is derived from stiffness, S_t, and the previously determined constants C and D by Eq. (17). Lastly, constant E is computed from B, C, S_x, and X_{mp} (or X_{mn}) by Eq. (22).

Constant E must be checked against the permissible range given by Eq. (3). If E exceeds the lower end of the permitted range by more than, say, a factor of 1.8, new values of the six performance descriptors must be chosen that would bring E into compliance. Finally, if needed, offsets X_o and Y_o are computed from S_x, S_y, and S_t, according to Eqs. (23) and (24).

We realize that the choice of these six performance descriptors is somewhat arbitrary; six other independent descriptors could have been selected. For instance, instead of stiffness S_t, offset X_o or Y_o could have been specified. Stiffness S_t would then be calculated from S_x, S_y, and X_o (or Y_o) according to Eqs. (25) or (26), respectively.

All these steps are easily combined into a Paper-Tire program. The program accepts six independent performance descriptors as input (stiffness, four peak coordinates, and S/P ratio), and delivers six MF constants (two center point coordinates; plus B, C, D, and E) as output. From these constants the desired Y(X) function can be computed.

SOME EXAMPLES

The few examples given here are arbitrarily restricted to the relation between lateral force and slip angle at a single load. Following the same Paper-Tire concept, we could evolve similar examples for other performance functions such as the relation between longitudinal force and slip ratio; or between aligning moment and slip angle; or indeed between any two variables whose relation can be represented by the general S-shaped curve described earlier.

Standard Curve - To demonstrate various possibilities of the Paper-Tire concept, we will start with an actually measured relation between lateral force and slip angle and consider it for some reason in need of modification, according to the wishes of the vehicle dynamicist. The actually measured curve, called the standard curve, is depicted in Fig. 3; its physical parameters are

stiffness S_t= - 80,000 N/rad
positive peak values Y_{mp}= -5,500 N; X_{mp}= 0.18 rad
negative peak values Y_{mn}= 5,000 N; X_{mn}= -0.16 rad
S/P ratio R = 0.9

From these six physical parameters, the six MF constants were computed by the Paper-Tire program and the MF function Eq. (1) calculated and drawn Fig. 3. The six MF constants of the standard curve are
S_x= 0.01 rad; S_y= -250 N; B = 11.84 1/rad; C = 1.29; D = -5250 N; E = -0.801

We also find the two offsets as X_o= 0.00687 rad; Y_o=550 N.

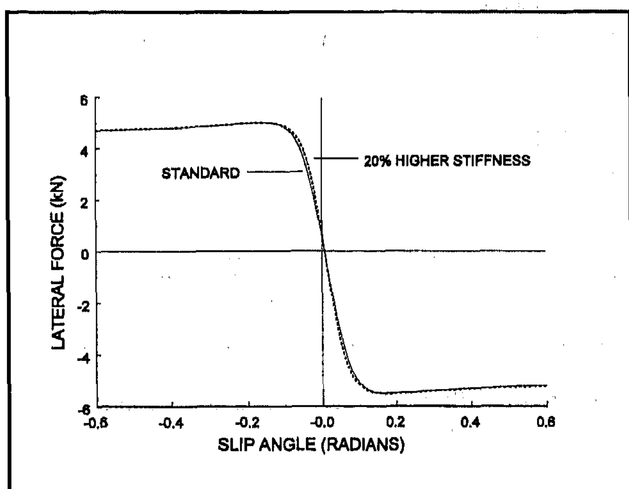


Figure 3. Effect of a 20% Increase in Cornering Stiffness

Change of Stiffness - We will increase the stiffness S_t of the standard curve by 20% from -80,000 N/rad to -96,000 N/rad; the peak coordinates and the S/P ratio are kept the same. How are the MF constants affected by this change?

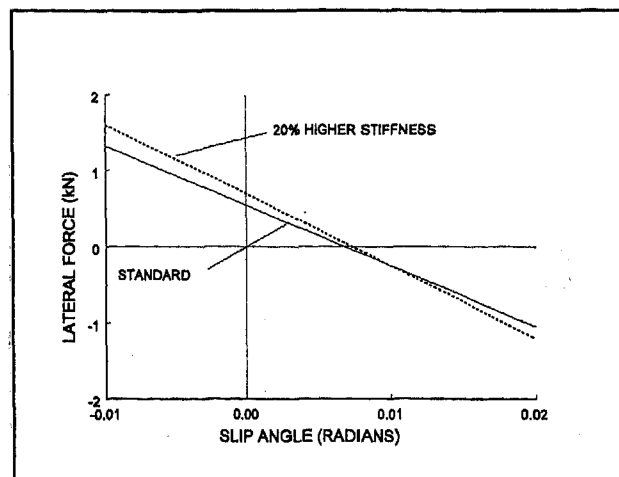


Figure 4. Effect of 20% Increase in Cornering Stiffness at Small Slip Angles

Exercising the Paper-Tire program, we find a new B value of B = 14.21 1/rad; and a new E value of E = -0.261. The latter number is well within the permissible range of E between 1 and $-(1+0.5{\cdot}C^2)$ = -1.83. The other four MF constants remain unchanged. The result is shown and compared with the standard curve in Figs. 3 and 4. We note that the two offsets are increased – X_o by 8% from 0.00687 rad to 0.0074 rad; and Y_0 by 29% from 550 N to 710 N.

Change of Peak Forces - We might find it desirable to increase the peak lateral forces of the standard curve by 10%; i.e., from -5500 N to -6050 N for the positive peak, and from 5000 N to 5500 N for the negative peak. The slip angle location of the peaks should remain unchanged, and so should stiffness and S/P ratio.

The Paper-Tire program shows that the increases in the peak forces entails three modifications to the MF constants: S_y expands by 10% from -250 N to -275 N; B drops by 9% from 11.84 1/rad to 10.76 1/rad; and D grows by 10% from -5250 N to -5775 N. We also note that the two offsets are diminished by 5% – X_o from 0.00687 rad to 0.00656 rad, and Y_o from 550 N to 525 N. The Y(X) function computed from these new MF constants is compared with the standard curve inFig. 5.

Change of Slip Angle at Peak Forces - Instead of changing the magnitude of the peak forces we will change their slip angles, here by 100%. Hence X_{mp} increases from 0.18 rad to 0.36 rad; and X_{mn} from -0.16 rad to -0.32 rad. All other physical parameters are kept the same. Result: S_x grows by 100% from 0.01 rad to 0.02 rad; and E jumps from -0.801 to +0.478. The latter value is safely below the upper limit of 1. Fig. 6 shows an overall comparison of the new curve with the standard curve; Fig. 7 highlights the area around the positive peak.

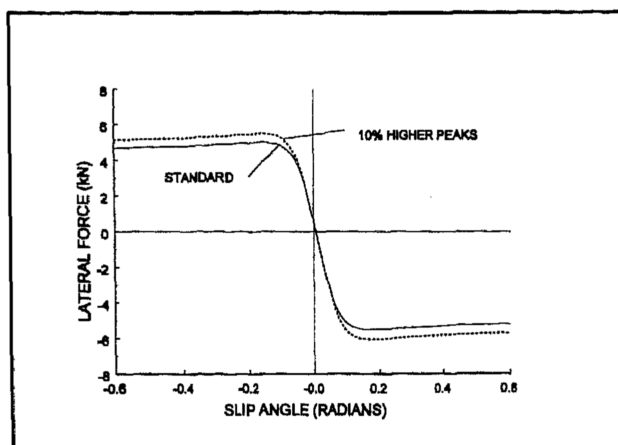


Figure 5. Effect of a 10% Increase in Peak Lateral Force

Changes of the S/P Ratio - We might find it advantageous to modify the S/P ratio of the standard curve downwards by 20% from 0.9 to 0.72; peaks and stiffness remain unchanged. The Paper-Tire program responds withchanges of C by 16 % from 1.29 to 1.49; of B by 14% from 11.84 1/rad to 10.24 1/rad; and of E from -0.801 to -0.036 (well within limits). The resulting new function is compared with the standard function in Fig. 8.

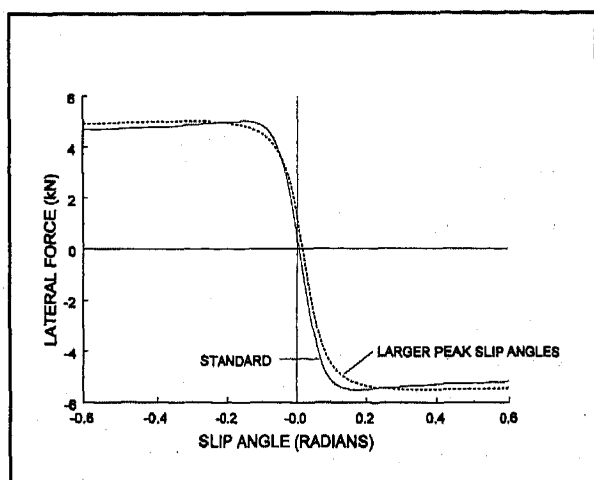


Figure 6. Effect of a 100% Increase in the Slip Angle at Peak Lateral Force

Changes of All Performance Characteristics - As a last example, the effect of simultaneous changes of all physical parameters is demonstrated. We have dropped stiffness S_t by 15% (from -80,000 N/rad to -68,000 N/rad); reduced both the positive and negative peak forces by 10 % (Y_{mp} from -5500 N to -4950 N, and Y_{mn} from 5000 N to 4500 N); increased the two corresponding slip angles by 20% (X_{mp} from 0.18 rad to 0.216 rad, and X_{mn} from -0.16 rad to -0.192 rad); and elevated the S/P ratio R by 5% from 0.90 to 0.95. (Reminder: at R = 1, the peak values recede into the region of very large slip angles.)

As a result, all six MF constants are affected. S_x is increased by 40% from 0.010 to 0.014; S_y dimishes by 10% from -250 N to -225 N; C is reduced by 7% from 1.29 to 1.20; B slightly increased by 1% from 11.84 1/rad to 11.97 1/rad; D reduced by 10% from -5250 N to -4725 N; and E changed a little from -0.801 to -0.965. The resulting new function is compared with the standard function in Fig 9, and, in some detail, in Fig 10.

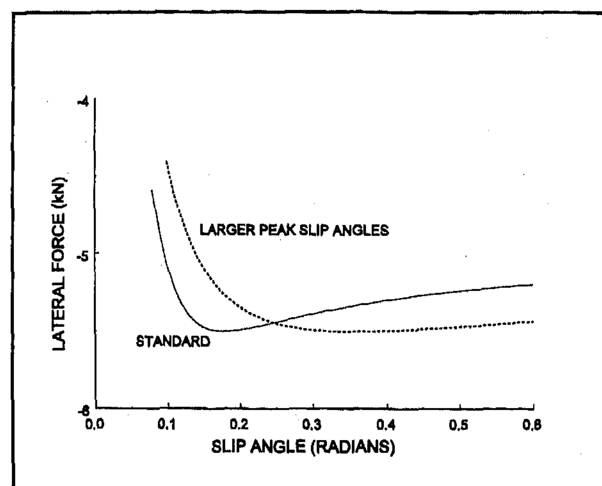


Figure 7 Peak Detail for an Increase in the Slip Angle at the Peak Lateral Force

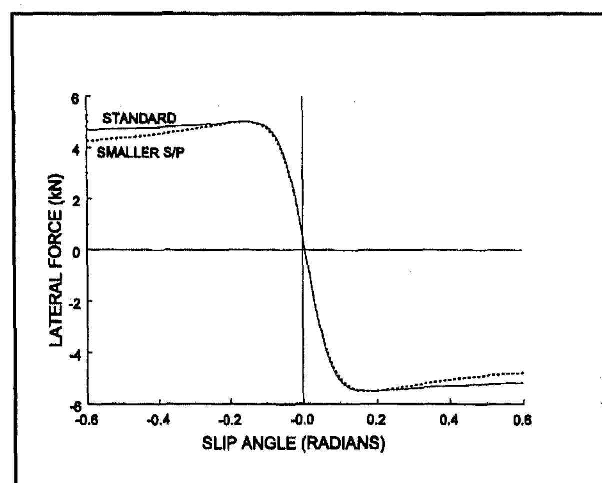


Figure 8 Effect of a 20% Reduction in Slide/Peak Ratio

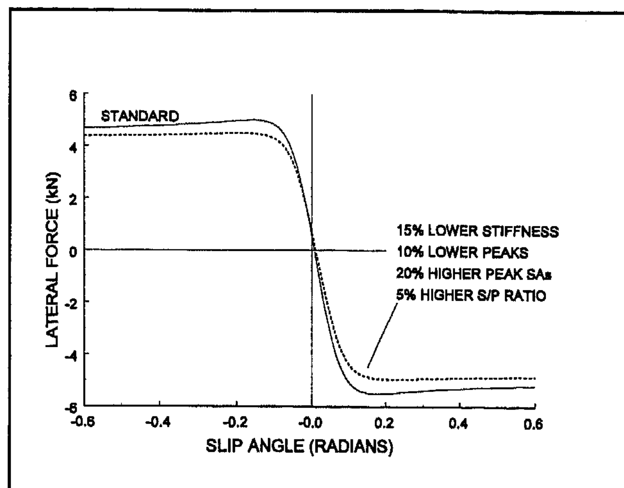


Figure 9 Effect of Making Six Independent Changes at Once

DISCUSSION

The main purpose of the preceding remarks was to demonstrate the principle of the Paper-Tire concept and give some samples of its usefulness and versatility. There appears to be no other tire model presently available that would combine the accuracy of the Magic Formula (accuracy with respect to fitting physical test data) with the convenience and ease offered by the use of generally accepted and physically meaningful performance descriptors.

The Paper-Tire concept is by no means restricted to lateral force/slip angle relations, as exercised here for illustration purposes; any other relation exhibiting the desired features (i.e.,S-shape, and symmetry with respect to a center point located closely to the origin) would be amenable to the Paper-Tire concept. In this general sense, the foregoing must be considered a small part of a potentially much larger body of equations that would include other performance relations such as, the longitudinal force/slip ratio relation; the overturning moment/slip angle relation; and the relation between longitudinal and lateral force for combined braking/driving [3]. For all these and other, similar performance relations, suitable performance descriptors are available that could take the place of the MF constants, so that a total picture of tire performance would evolve based exclusively on conventional performance characteristics, and hence, could be easily modified and optimized.

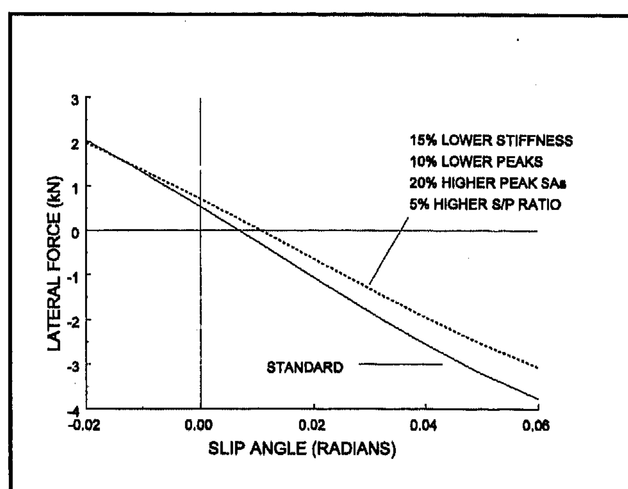


Figure 10 Positive Slip Angle Detail for a Combination of Six Independent Changes

This paper is only the first step toward that goal. Two areas of further development are foreseen: asymmetrical performance curves, and load effects.

Asymmetry - Eq. (1) is not able to directly handle asymmetrical cases where there is no center point and hence imperfect congruence exists between the positive and the negative branches. A case in point is the relation between longitudinal force and slip ratio: the relations for braking (negative slip ratios) and for driving (positive slip ratios) are often different. A proven way out of this is to develop two MF functions -- one for braking and one for driving. The driving function ignores all data for negative slip ratios and covers only data for positive slip ratios; vice versa, the braking function ignores all data for positive slip ratios and covers only data for negative slip ratio. The 'ignored' branches are then dropped and the remaining branches fused together near the origin with identical slopes. Details of this procedure have already been worked out and are available for further processing.

Load Effects - Another problem that must be addressed before universal application of the Paper-Tire Concept becomes feasible is posed by the effect of tire load on pertinent performance descriptors. Vehicle dynamics is inevitably bound up with tire load changes, which in turn influence the numerical values of the descriptors involved. The solution to this problem is rather straightforward. Each of the performance descriptors used in the Paper- Tire Concept, such as cornering stiffness and peak forces, must be expressed as a function of load, so that at any given load a set of numerical descriptor values used as an input to the Paper-Tire program would evolve. The load functions of a given Paper-Tire could be based either on actual data or on hypothetical load-descriptor relations. The latter would provide an additional degree of freedom for determining the 'perfect' tire for the task at hand.

REFERENCES

[1] E. Bakker, L. Nyborg, H.B. Pacejka, 'Tyre Modeling for Use in Vehicle Dynamics Studies', SAE Paper 870421 (1987)

[2] D.J. Schuring, W. Pelz, M.G. Pottinger, 'The BNPS Model -- An Automated Implementation of the "Magic Formula" Concept', SAE Paper 931909 (1993)

[3] D.J. SCHURING, W. PELZ, M.G. POTTINGER,'A Model for Combined Tire Cornering and Braking Forces', SAE Paper 960180

APPENDIX

Eq. (12) in the derivation is repeated as Eq. (A1).

$$\sin(0.5\pi C) = R^* \qquad \text{(A1)}$$

By solving for C, we obtain

$$C = (2/\pi)\sin^{-1}R^* \qquad \text{(A2)}$$

Since the principal values of $\sin^{-1}$ lie in the range$[-\pi/2,\pi/2]$, C satisfies the relation $-1 \leq C \leq 1$. However, Eq. (2) requires that $1 < C < 3$.

We can resolve this discrepancy by considering the portion of the sine curve given below. The value C_1 corresponding to the ordinate value R^* is the principal value as given in Eq. (A2). However there are an infinite number of values of C that satisfy equation A1, including the point C_2. If both C_1 and C_2 satisfy equation A1 for the same value R^* and $-1 < C_1 < 1$, then we have the $1 < C_2 < 3$. Because of the symmetry of the sine curve around the vertical line $C = 1$, we have the $C_1 - 0 = 2 - C_2$. Some elementary algebra produces Eq. (A3) which is also Eq. (13) in the paper.

$$C = 2[1 - (1/\pi)\sin^{-1}R^*] \qquad \text{(A3)}$$

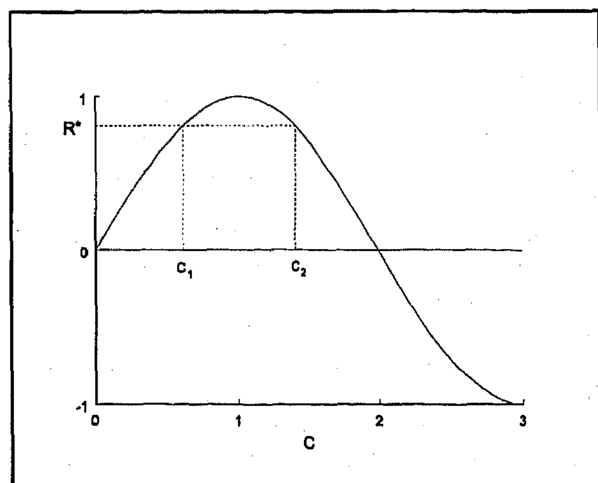


Figure A-1 R* vs. C Wave

970560

Variable Dynamic Testbed Vehicle: Dynamics Analysis

Allan Y. Lee* and Alan T. Marriott
California Institute of Technology

Nhan T. Le
University of California

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

The Variable Dynamic Testbed Vehicle (VDTV) concept has been proposed as a tool to evaluate collision avoidance systems and to perform driving-related human factors research. The goal of this study is to analytically investigate to what extent a VDTV with adjustable front and rear anti-roll bar stiffnesses, programmable damping rates, and four-wheel-steering can emulate the lateral dynamics of a broad range of passenger vehicles. Using a selected compact-sized automobile as a baseline, our study indicated this baseline vehicle can be controlled to emulate the lateral response characteristics (including the vehicle's understeer coefficient and the 90% lateral acceleration rise time in a J-turn maneuver) of a fleet of production vehicles, from low to high lateral acceleration conditions. Also, the roll gradient of the baselined vehicle can be altered via changes made to the torsional stiffnesses of the front and/or rear anti-roll bars to emulate the roll stiffnesses of a fleet of production vehicles.

Introduction

To study the correlation between vehicle response characteristics and driver commands relative to crash avoidance, the National Highway Traffic Safety Administration's Office of Crash Avoidance Research (OCAR) has at its disposal a comprehensive set of tools and facilities. These include the Vehicle Research and Test Center, and the (currently being developed) National Advanced Driving Simulator. To augment these tools and facilities, OCAR has defined its concept of a Variable Dynamic Testbed Vehicle (VDTV).[1] This vehicle will be capable of emulating a broad range of automobile dynamic characteristics, allowing it to be used in development of collision avoidance systems, and conducting of driving-related human factors research, among other applications.

Vehicles with "programmable" response characteristics have been proposed and developed in the past. In the 1970's, an experimental vehicle, called Variable Response Vehicle, was developed by the General Motors Corporation for vehicle handling research.[2] It had independent electro-hydraulic controlled front and rear steering actuators and a front steering feel system. These active systems enabled it to emulate a variety of directional control characteristics. In the 1990's, a similar research vehicle, called Simulator Vehicle, was developed by the Nissan Motor Company.[3] Both yaw rate and lateral acceleration response characteristics of this vehicle were varied independently. It was used to study the relation between driver's perception and vehicle handling quality.

To emulate both the lateral and longitudinal response characteristics of a broad range of vehicles, the "mechanical" steering, suspension, and braking sub-systems of a "passive" vehicle must

*To whom all correspondence should be sent.

cles, the "mechanical" steering, suspension, and braking sub-systems of a "passive" vehicle must all be made "programmable". With regard to emulating the lateral response characteristics of vehicles, an earlier study[4] indicated that the VDTV must have the following active sub-systems: (1) steering: steer-by-wire, programmable steering feel as well as four-wheel-steering, and (2) suspension: semi-active suspension as well as variable front and rear anti-roll bar systems. Other active sub-systems considered in Reference 4 (such as the brake-by-wire and throttle-by-wire systems) were not included in this study.

Equipped with the above mentioned actively controlled systems, the lateral response characteristics of the VDTV can be conveniently altered via the governing control algorithms. However, it was not clear what range of production vehicles could be emulated by such a variable dynamic vehicle. One objective of the dynamics analysis was to gain a quantitative understanding on the "emulability" of such a variable dynamic vehicle. The second objective was to generate quantitative information for the functional requirements document[14] that accompanied the VDTV Request for Proposal.

Scope and Approach of Dynamics Analysis

The scope and approach taken in the dynamics analysis are as follows:

1. A vehicle dynamics simulation program, called Vehicle Dynamic Analysis, Nonlinear (VDANL) was selected as the simulation tool to perform all the vehicle dynamic computations.

2. The VDANL program has parameter files for about twenty passenger vehicle models. Only five models were selected to represent a broad range of production passenger vehicles.

3. Three performance metrics were selected to characterize both the steady-state and transient lateral responses of these production vehicle models in "representative" cornering maneuvers.

4. One of the five models studied was selected as the baseline VDTV. However, to account for the added weights of the data acquisition system, four-wheel-steering system, etc., several vehicle and tire parameters of the selected vehicle model were modified accordingly.

5. Two sensitivity analyses were made to assess to what degree vehicle performance metrics selected in step (3) are influenced by the following vehicle parameters: (a) the torsional stiffnessess of the front and rear anti-roll bars, and (b) the damping rates of the suspension shock absorbers.

6. A simulation study was made to assess to what extent the understeer coefficient and speed of lateral response of the five production vehicles selected in step (2) can be emulated by a four-wheel-steered VDTV.

7. The Consumers Union obstacle avoidance course (to be described latter) was used to objectively evaluate the handling qualities of passenger vehicles during emergency double lane change maneuvers. The performance of the baseline VDTV in making double lane change maneuvers using different combinations of tires and four wheel steering control algorithm were compared.

Results obtained from these seven steps are given in the following sub-sections.

Vehicle Dynamics Simulation Program

A vehicle dynamics simulation program, developed by Systems Technology, Incorporated, called "Vehicle Dynamics Analysis, Non-Linear" (VDANL) was used in this research. This program was originally developed to study the performance of vehicle/driver systems in a variety of driving scenarios and conditions, and to study vehicle lateral control and stability.[5]

A signal flow diagram of the program VDANL is depicted in Figure 1 (from Reference 6). The program VDANL has a total of seventeen degrees-of-freedom (DOFs). The block labeled "Vehicle Dynamics" (in Figure 1) has six DOFs

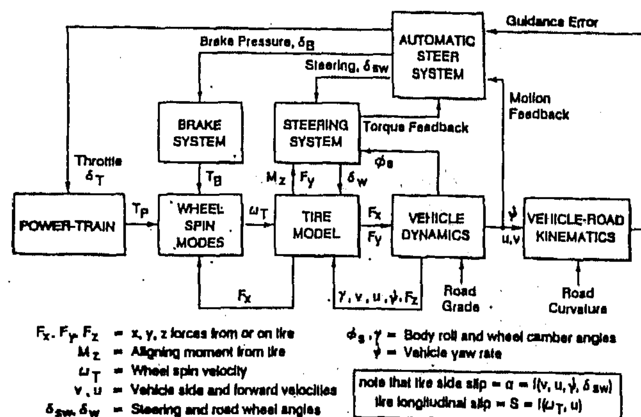


Figure 1: Flow diagram of the VDANL vehicle dynamic simulation program

for the sprung mass and three DOFs each for the front and rear unsprung masses. In the block labeled "Tire Dynamics," four DOFs for the rotational speeds of the four wheels are included. Finally, one degree of freedom each is contained in the blocks labeled "Steering System", "Brake System", and "Power Train." Representative vehicle and tire parameters that must be supplied for a typical simulation run are tabulated in Table 1 for the five passenger vehicle models that were selected in step (2).

The suspension system modeled in VDANL exerts forces on three lumped masses. Two unsprung masses represent the front and rear axles with tires, and one sprung mass represents the vehicle's body. Rather than model each component of the suspension individually, VDANL uses composite characteristics of all components to calculate overall suspension dynamics. This method produces less complicated calculations and a significant reduction in data size. The simulation program VDANL also includes an option of using both front and rear anti-roll bars to provide extra auxillary roll stiffness.

The simulation program VDANL uses a comprehensive tire model that includes the effects of road surface interaction and saturation limit. This assures accurate representations of real driving conditions, from low-g to limit lateral maneuvers, and for combined lateral and longitudinal maneuvers (for example, brake-in-a-turn). The program VDANL uses tire parameters defined by the Calspan Corporation (cf. Table 1) that contain parameters for cornering stiffness, camber stiffness, lateral and circumferential friction, aligning torque, as well as overturning moment.

Besides the above mentioned models, the simulation program VDANL also provides a "closed-loop" driving capability by using the block labeled "Automatic Steer System." This block contains internal algorithms that a human driver uses to generate the required steering, braking, and throttle commands. It can also represent how an autonomous steering system generates the needed steering commands. Furthermore, an "Open" module option provides an interface between a user-supplied "subroutine" and the main program. In this research, we used this option to implement four-wheel-steering control algorithms.

Data predicted by VDANL have been extensively validated via road tests for a wide variety of vehicles and driving conditions.[6–9] Results obtained using the simulation program were found to represent measured vehicle responses quite well for most vehicles and maneuvers.

Selected Production Vehicle Models

Five production passenger vehicle models were selected in this study to represent a fleet of passenger vehicles for the VDTV to "emulate." These represented a wide spectrum in vehicle weight: from "small," "compact," "mid-sized," to "full-sized" passenger vehicles. Additionally, these five vehicle models were selected because they span wide ranges in both wheelbase ratio and track width ratio. The wheelbase ratio, the ratio between the vehicle's wheelbase and its center of gravity (c.g.) height $(L/h_{c.g.})$, is strongly related to the amount of weight transfer between the rear and front wheels during acceleration/deceleration maneuvers. Track width ratio, the ratio between the vehicle's halved track width and c.g. height $(tw/2h_{c.g.})$, is strongly related to the amount of weight transfer between the inside and outside wheels during lateral cornering ma-

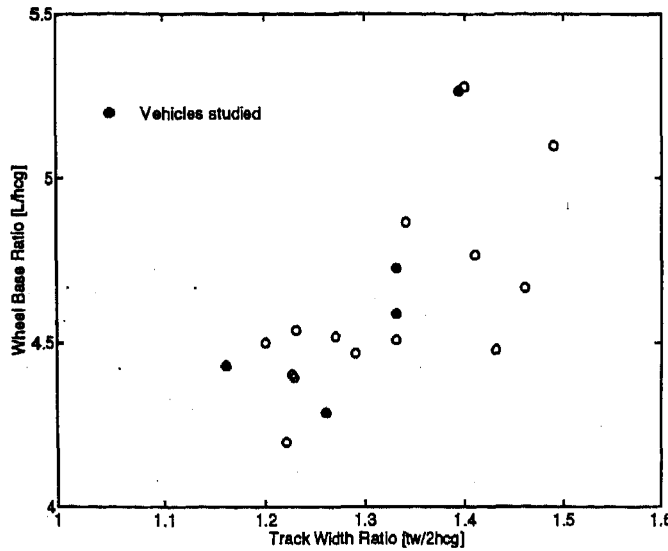


Figure 2: Vehicle's wheelbase ratios versus trackwidth ratios

A plot of wheelbase ratios versus track width ratios for five production vehicles is shown in Figure 2. In that figure, the "o"s denote the wheelbase and track width ratios combinations of production vehicle models with parameter files available from the VDANL program. The "•"s represent the five selected passenger vehicle models. Clearly, the selected vehicles span good ranges in both the wheel base ratios and the track width ratios. Estimated values of other vehicle and tire parameters are tabulated in Table 1. See Appendix E of Reference 10 for the estimated values of vehicle and tire parameters that are not given in Table 1.

Steady-state and Transient Lateral Response Characteristics of Passenger Vehicles

Using the simulation program VDANL, and estimates of vehicle and tire parameters given in Table 1, three steady-state and transient lateral response performance metrics of the five selected passenger vehicles were computed. Results are given in the following sections.

The steady-state handling quality of a vehicle can be characterized by its understeer coefficient (K_u, in deg/g). The understeer coefficient, defined in sub-section 9.4.7 of Reference 11, can be understood using the following steady-state cornering equilibrium equation:

$$\delta_T = \delta_{SW}/N_S = \overbrace{57.3L/R}^{Ackerman\ angle} + K_u a_y\,, \qquad (1)$$

where,

δ_T = tire angle (deg),
δ_{SW} = steering wheel angle (deg),
N_S = steering ratio (-),
L = wheel base (m),
R = turning radius (m),
K_u = understeer coefficient (deg/g),
a_y = lateral acceleration at vehicle's c.g. (g).

From this equation, the understeer coefficient of a vehicle can be determined with a "turn circle" maneuver. In this maneuver, the forward speed of the vehicle is kept constant at, for example, 80 km/h. The steering wheel angle is increased slowly at an uniform rate of about 5 deg/sec until the limit lateral acceleration of the vehicle is reached. A cross plot between the lateral acceleration of the vehicle's c.g. on the X-axis and the steering wheel angle on the Y-axis can then be generated. From this plot, the understeer coefficient is determined as follows:

$$K_u = \frac{1}{N_S}\frac{\partial \delta_{SW}}{\partial a_y} - 57.3\frac{gL}{U^2}\,. \qquad (2)$$

The first term on the right hand side of equation (2) can be derived from the instantaneous slope of the a_y versus δ_{SW} plot at a given lateral acceleration. The variable "U" in the second term represents the forward speed of the vehicle. That term can be easily computed if we know the wheelbase of the vehicle as well as the speed at which that simulation run was performed. Results obtained from these turn circle maneuvers are given in two cross plots: vehicle's lateral acceleration versus the steering wheel angle (Figure 3), and lateral acceleration versus the understeer coefficient (Figure 4), for the five selected passenger vehicles. From Figure 4, we note that the understeer coefficients of these vehicles remain unchanged when the lateral acceleration of the vehicle is lower than about 0.3 g. However, these understeer coefficients increase drastically as their respective limit lateral acceleration levels are approached. These results are fairly representative for most passenger vehicles.

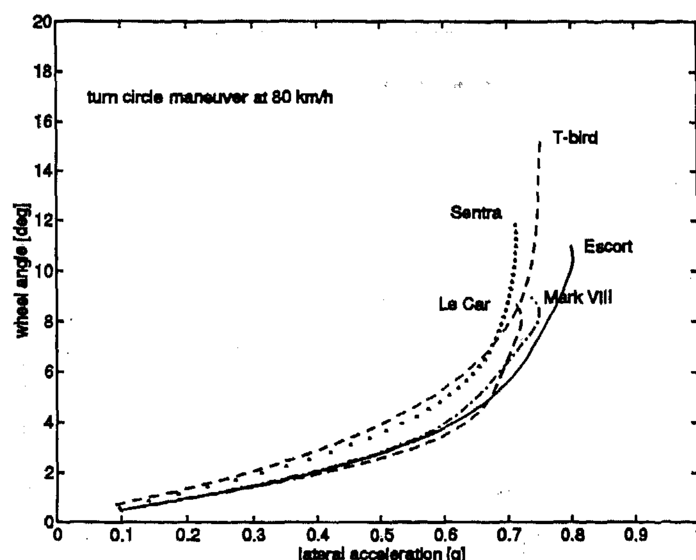


Figure 3: Turn circle maneuver, steering wheel angle versus lateral acceleration

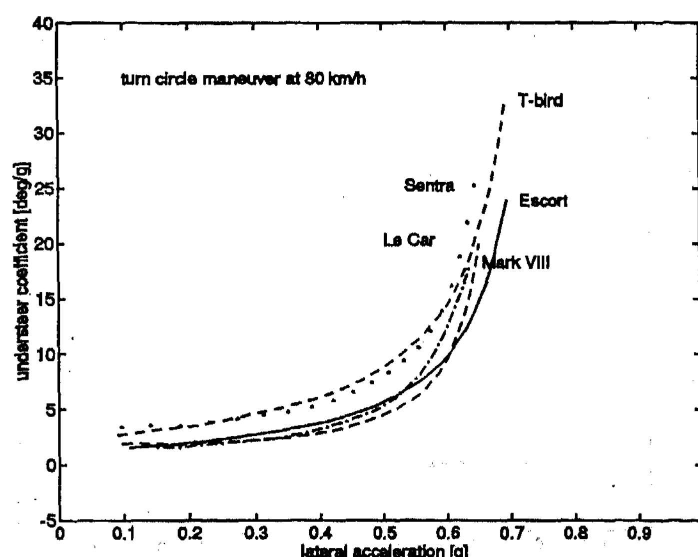


Figure 4: Turn circle maneuver, understeer coefficient versus lateral acceleration

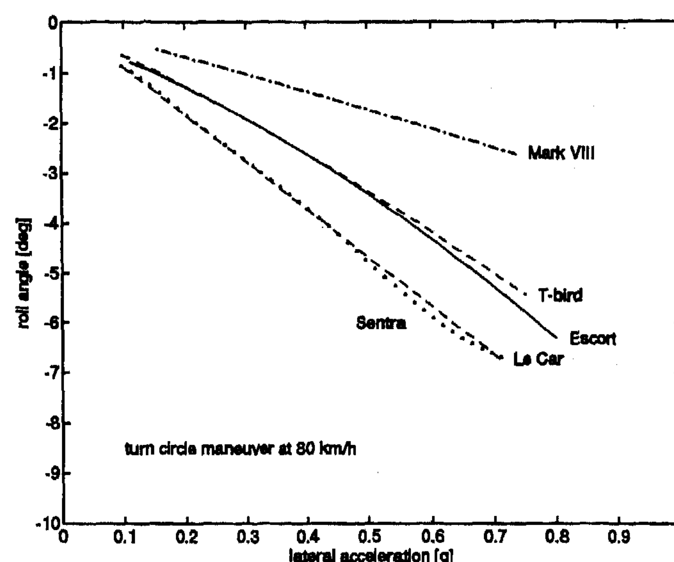


Figure 5: Turn circle maneuver, roll angle versus lateral acceleration

Another performance metric commonly used to characterize the steady-state vehicle handling quality of a vehicle is its control sensitivity (in g/deg). The control sensitivity of a vehicle at a given forward speed is also known as its steering sensitivity or lateral acceleration gain. In equation (1), the vehicle control sensitivity is given by a_y/δ_{SW}. The relation between the vehicle's control sensitivity and its understeer coefficient is given by:

$$\frac{a_y}{\delta_{SW}} = \frac{1}{N_S}[\ K_u + 57.3\frac{gL}{U^2}\ ]^{-1}. \qquad (3)$$

Since the vehicle's control sensitivity can be computed using the understeer coefficient, only the understeer coefficient is used in this study.

The roll gradient of a vehicle can also be determined using data obtained from the turn circle maneuver. A vehicle's roll gradient, defined in sub-section 9.4.19 of Reference 11, can be determined from a cross plot between the lateral acceleration of the vehicle's c.g. and the vehicle's roll angle. Results obtained for the five passenger vehicles are depicted in Figure 5 wherein we note that the roll gradients of full-size vehicles tend to be small, while the reverse is true for small and compact vehicles.

The transient lateral response characteristics of the vehicles are compared using the "90% rise

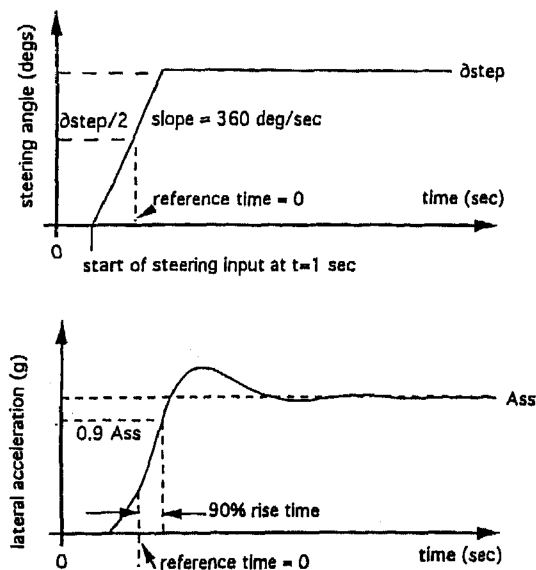


Figure 6: Steering wheel angle and vehicle's lateral acceleration during a J-turn maneuver

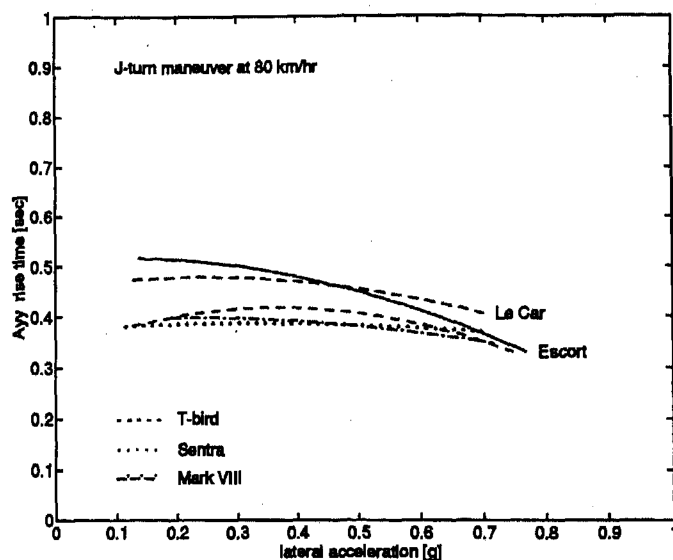


Figure 7: J-turn maneuver, 90% rise time versus lateral acceleration

times" of their lateral acceleration responses. The 90% rise time is a measure of the vehicle's "speed of response" when it is subjected to a "step" steering wheel command. Since a true "step" is physically impossible, the steering command is ramped to its steady-state value at an uniform rate of 360 degrees per second. The resultant maneuver is commonly called a J-turn maneuver. The 90% rise time is defined as the time it takes the vehicle's lateral acceleration to reach 90% of its steady-state value, as measured from the time the steering command reaches 50% of its steady-state value. See Figure 6 and Reference 12 for further details. Plots of the 90% rise time versus the vehicle's lateral acceleration, for the five selected passenger vehicles are depicted in Figure 7.

The speed of lateral response of a vehicle can also be measured using frequency-domain performance metrics such as the vehicle's lateral acceleration -3 dB bandwidth (in Hz). This is the frequency at which the magnitude of the transfer function, from the steering wheel to the vehicle's lateral acceleration, has dropped below 70.7% of its steady-state value. Again, there is a strong correlation between the vehicle's 90% rise time and its lateral acceleration bandwidth. For the 1989 Escort, that correlation is depicted in Figure 8: the larger the bandwidth, the smaller the 90% rise time. A linear approximation between

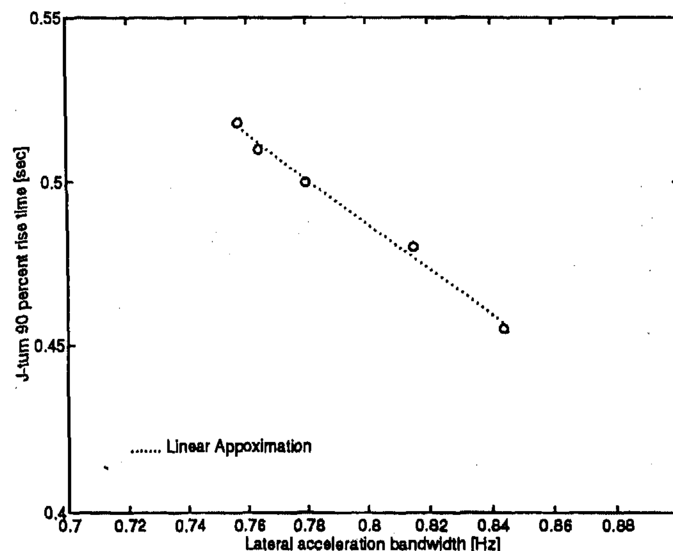


Figure 8: Correlation between vehicle bandwidth and 90% rise time

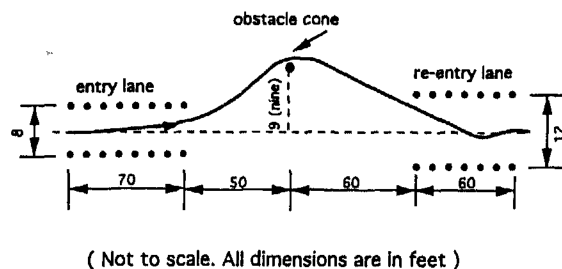


Figure 9: Consumers Union obstacle avoidance course

the bandwidth and rise time is given by:

$$t_{rise} \approx 1.0381 - 0.6888 \times BW\,. \qquad (4)$$

Here, BW is the vehicle's lateral acceleration bandwidth in Hz, and t_{rise} is the vehicle's J-turn 90% rise time in seconds. Since the lateral acceleration bandwidth is closely correlated with the 90% rise time, only the rise time is used in our study.

The Consumers Union obstacle avoidance course, depicted in Figure 9, was used to objectively evaluate the handling quality of passenger vehicles during emergency double lane change maneuvers. Details of this obstacle avoidance course are defined in Reference 13. The maximum speed that a passenger vehicle can successfully complete the double lane change maneuver is denoted by U_{max}. The magnitude of U_{max}, in km/hr, is another lateral response performance metric used in our study.

Selected Baseline Vehicle

For the purpose of dynamics analyses, the 1989 Ford Escort was selected as the baseline variable dynamic testbed vehicle. The following modifications are made to the production model to account for the added weights of the data acquisition system, four wheel steering actuator mechanism, and others:

- Sprung mass and moment of inertia of the production Escort were each increased by 28%. However, these inertia properties were increased without altering the longitudinal, lateral, and vertical positions of the vehicle's overall c.g. location.
- Spring rates and damper rates of the production vehicle's suspension were each increased by 28%. Since the vehicle sprung mass was increased by 28%, there was no change in either the frequency or damping ratio of the vehicle's heave mode.
- Torsional stiffness of the front anti-roll bar of the production Escort was increased by 28% (in magnitude). The production Escort has no rear anti-roll bar.
- "Larger" tires were selected for the modified Escort. The tire model used is P195/75R14.

With these modifications, the weight and moment of inertia of the "compact" Escort were now midway between those of "compact" and "mid-sized" vehicles. It was judged that such a modified Escort could better emulate the lateral response characteristics of both the "small" and "compact" production vehicles, that have higher accident statistics. If a mid-sized vehicle had been selected instead as the baseline vehicle, the added weights would have caused its weight to approach that of a "full-size" vehicle. This heavier baseline vehicle might not have been able to emulate the lateral response characteristics of both the "small" and "compact" vehicles as well.

Sensitivity Analyses

Two types of sensitivity analyses were performed to assess the degree that selected vehicle performance metrics could be influenced by two vehicle parameters: effects of (1) the torsional stiffnessess of the front and rear anti-roll bars on the vehicle's roll gradient, and (2) the damping rates of the suspension shock absorbers on the 90% rise time of the vehicle in J-turn maneuvers.

Five anti-roll bar configurations used in this study are depicted in Figure 10. In that figure,

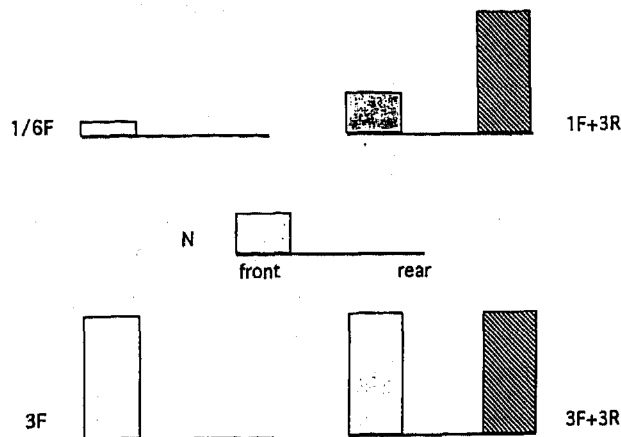


Figure 10: Anti-roll bar stiffness configurations used in a sensitivity analysis

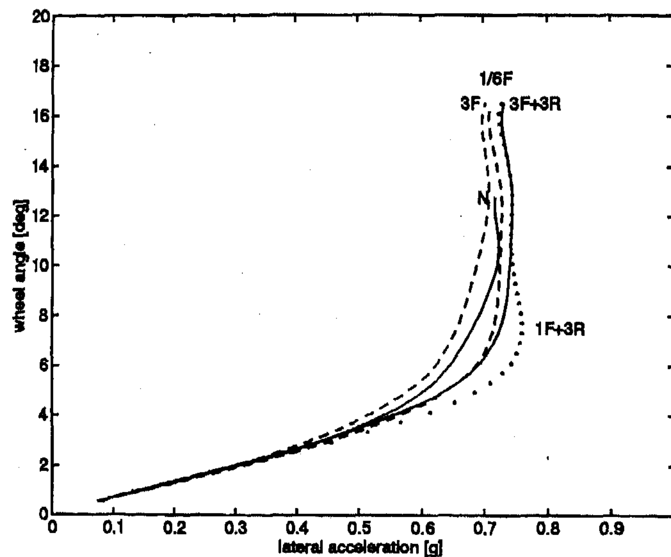


Figure 11: Turn circle maneuver results for five anti-roll bar configurations

"N" denotes for the "nominal" anti-roll bar configuration of the baseline VDTV: stiffnesses of the front and rear anti-roll bars are -288.6 and 0 Nm/deg, respectively (the baseline VDTV does not have a rear anti-roll bar). The stiffnesses of all the other anti-roll bar configuration variants are multiples of the nominal front anti-roll bar stiffness. For example, the front and rear anti-roll bar stiffnesses of the configuration labeled "1F+3R" are -288.6 and -865.8 Nm/deg, respectively.

Plots of steering wheel angle versus the vehicle's lateral acceleration obtained in turn circle maneuvers, for the five anti-roll bar configurations, are compared in Figure 11. Plots that compare the vehicle's understeer coefficient, roll gradient, and 90% rise time in a J-turn maneuver are given in Figures 12, 13, and 14, respectively. From Figure 12, we observe:

- For low-g (below 0.3 g) maneuvers, the anti-roll bars' stiffnesses have very little effect on the vehicle's understeer coefficient.
- At high-g conditions, the anti-roll bars' stiffnesses have the following influences on the vehicle's understeer coefficient:
  - understeer coefficient is increased by an increase in the front anti-roll bar stiffness. See results obtained for the "3F" configuration.

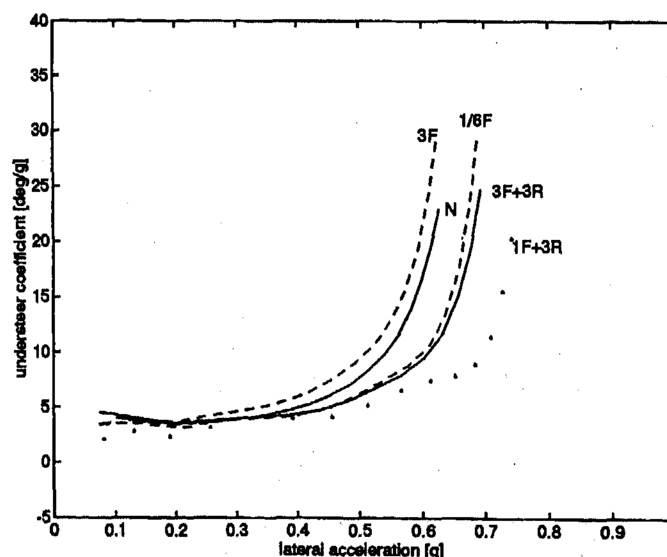


Figure 12: Understeer coefficient results for five anti-roll bar configurations

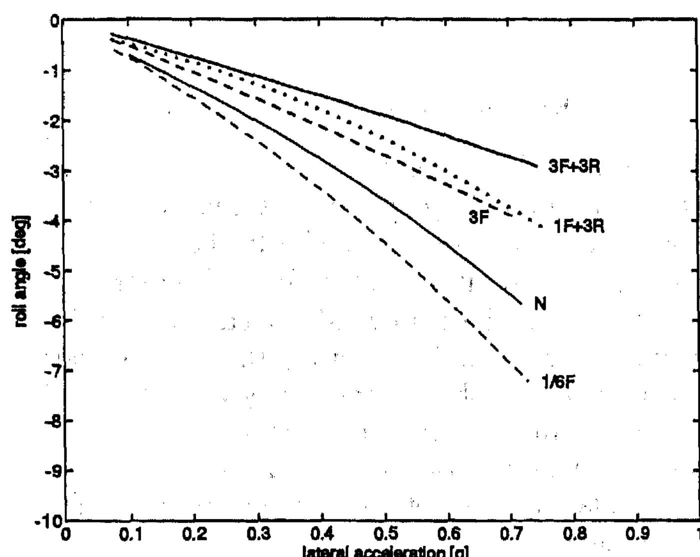


Figure 13: Vehicle roll gradient results for five anti-roll bar configurations

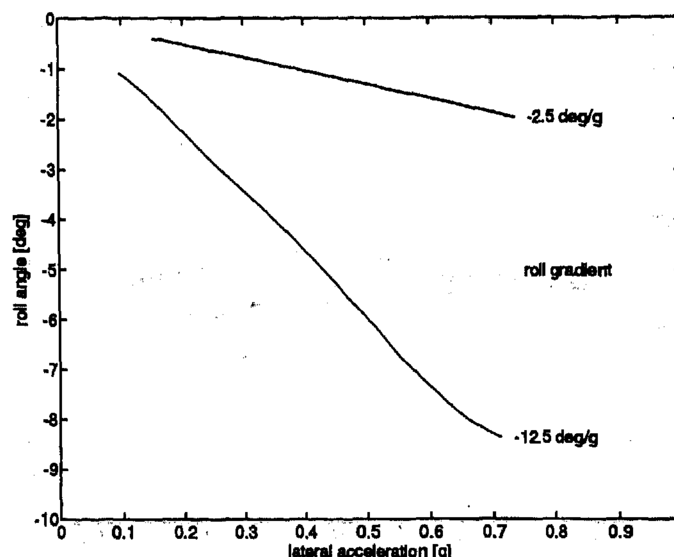


Figure 14: Emulation range of an active anti-roll bar controlled system

- understeer coefficient is decreased by an increase in the rear anti-roll bar stiffness. See results obtained for the "1F+3R" configuration. Alternatively, the vehicle's understeer coefficient can be decreased by decreasing the front anti-roll bar stiffness. See results obtained for the "$\frac{1}{6}$F" configuration.

From Figure 13, we conclude that the stiffnesses of the anti-roll bars have a dominant influence on the vehicle's roll gradient. The larger the magnitude of the total anti-roll bar stiffness (sum of both the front and rear anti-roll bar stiffnesses), the smaller is the vehicle's roll gradient. Hence, the roll gradient for the "3F+3R" configuration is the smallest among the five configurations studied. Hence, changing the stiffnesses of the vehicle's anti-roll bars (either mechanically or by using active anti-roll bar controlled systems) is an effective way to alter the vehicle's roll gradient. The alternative of using a fully active suspension system to control the vehicle roll gradient is likely to be more expensive.

The specified emulation range of the VDTV's roll gradient, depicted in Figure 14, is obtained using results depicted in Figure 13: (a) the lower limit is obtained by reducing 25% from the roll gradient of the "3F+3R" configuration, and (b) the upper limit is obtained by adding 25% to the roll gradient of the "$\frac{1}{6}$F" configuration. This specified emulation range of the vehicle's roll gradient also appeared in Section 7.1.2 of Reference 14. Note also that the upper and lower roll gradient limits depicted in Figure 13 completely envelop the a_y versus roll angle plots of the five passeneger vehicles given in Figure 5.

The effects that anti-roll bar stiffnesses have on the 90% rise time in J-turn maneuvers are less obvious. However, the following trends are observed in Figure 15:

- vehicle becomes more responsive (with a smaller J-turn 90% rise time) if the stiffness of the front anti-roll bar is increased. See results obtained for the "3F" configuration.

- vehicle becomes less responsive (with a larger J-turn 90% rise time) if the stiffness of the front anti-roll bar is decreased. See results obtained for the "$\frac{1}{6}$F" configuration.

- vehicle also becomes less responsive (with a larger J-turn 90% rise time) if the stiffness of the rear anti-roll bar is increased. See results obtained for the "1F+3R" configuration.

The objective of the second set of senstivity analyses was to assess to what degree the damping rates of the vehicle suspension influence the

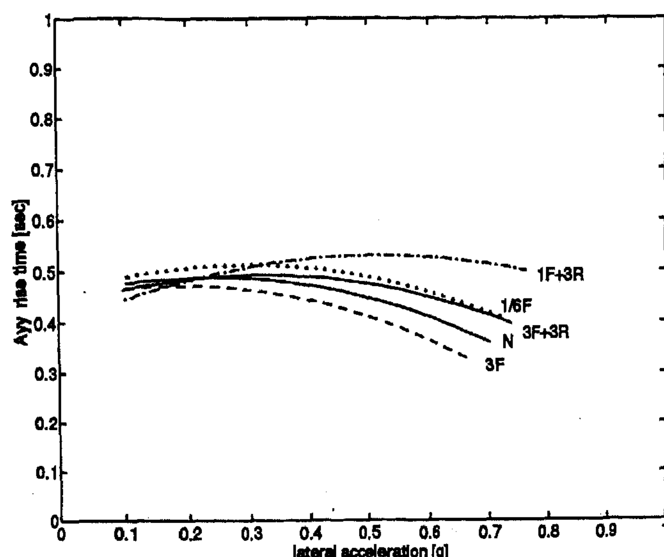


Figure 15: Rise time results for five anti-roll bar configurations

vehicle's J-turn 90% rise time. This analysis was performed for the Jet Propulsion Laboratory by Systems Technology Incorporated. Results obtained are summarized in the following two paragraphs while the details are given in Reference 15.

Three sets of damper rates were used in the sensitivity analyses. The "hardest" damping rates are: at a piston speed of 75 cm/sec, the damping forces are 3400 and 900 Newtons in extension and compression, respectively. The "softest" damping rates are: at a piston speed of 75 cm/sec, the damping forces are 1200 and 670 Newtons in extension and compression, respectively. The nominal damping forces are between those of the "hardest" and "softest" dampers. These damping rates are identical to those specified in sub-section 4.3.8 of Reference 14.

Using the above mentioned damping rates, the 90% rise times of a vehicle in a 80 km/hr 0.22 g J-turn maneuver were determined using VDANL. The 90% rise times obtained were 0.43, 0.51, and 0.49 seconds for the "softest," "nominal," and "hardest" damping rates, respectively. Differences among these rise times are quite small. Considering the accuracy of the VDANL simulation program, we concluded that suspension damping rates do not significantly influence the speed of response of a vehicle in cornering maneuvers.

Four-Wheel-Steering Control Algorithms

The lateral dynamics of a vehicle can be substantially altered by steering its rear wheels in conjunction with the front wheels. For example, the control sensitivity of a four-wheel-steering vehicle at a given forward speed can be increased/decreased by steering the rear wheels out-of-phase/in-phase respectively with the front wheels. Additionally, the transient lateral response characteristics of the vehicle can also be manipulated via carefully designed rear steering algorithms.

Both open-loop and closed-loop 4WS algorithms have been proposed and used in vehicle research.[16–19] Closed-loop 4WS algorithms use selected vehicle measurements (for example, vehicle's yaw rate) as well as the driver's steering command to control the rear steering actuator. On the other hand, open-loop algorithms do not feedback any vehicle measurements. The open-loop and closed-loop control algorithms described in the following sub-section are taken from Reference 19. They are given here to illustrate the flexibility available to alter the lateral response characteristics of a VDTV via four-wheel-steering algorithms.

- 4WSN Algorithm: This is an open-loop algorithm first suggested by Nissan Motor Company. Using a vehicle model, a speed-dependent ratio between the rear and front wheels is computed in order to achieve zero steady-state side velocity:

$$\delta_{rc}/\delta_{fc} = K_N(U)\,. \qquad (5)$$

  Here, δ_{fc} equals the driver steering wheel command divided by the steering ratio. The command to the rear steering actuator is given by δ_{rc}. The variable U is the forward speed of the vehicle. The function K_N is the "Nissan" ratio. At low speeds, the rear wheels are steered out-of-phase with the front wheels (i.e., K_N is negative) to enhance the vehicle maneuverability. At high speeds, the rear wheels are steered in-phase with the front

wheels (i.e., K_N is positive) to enhance the vehicle lateral stability. However, the lateral forces generated by both the front and rear wheels counteract one another, and the response time of the vehicle's yaw rate may deteriorate. To overcome this problem, we delay the steering of the rear wheel by τ_D seconds:[17]

$$\delta_{rc}(t) = K_N(U)\,\delta_{fc}(t - \tau_D)\,. \qquad (6)$$

The delay time τ_D is on the order of 0.1 seconds. The parameters K_N and τ_D in this "delayed" Nissan algorithm can be used to alter the steady-state and transient response characteristics of the vehicle, respectively.[17] Numerous other open-loop 4WS algorithms have also been suggested. See, for example, Reference 18. However, open-loop algorithms were not used in our study.

4WSY Algorithm: This is a simple closed-loop algorithm with feed-forward of the front steering command and feedback of the vehicle's yaw-rate:

$$\delta_{rc} \;=\; K_1(U)\,\delta_{fc} + K_2(U)\,r\,. \qquad (7)$$

Here, the variable δ_{fc} was defined in connection with the 4WSN algorithm. The variable r is the filtered yaw rate of the vehicle. This closed-loop 4WS algorithm, depicted in Figure 16, was the main focus of our study.

In equation (7), the feed-forward gain $K_1(U)$ chiefly alters the vehicle's steady-state responses. On the other hand, the feedback gain $K_2(U)$ affects both the steady-state and transient characteristics of the vehicle. Both parameters must be selected with care. The rear steering angle increases monotonically with the magnitude of K_1. Since the maximum rear steering actuator excursion is typically bounded (say, not more than 10 degrees), the magnitude of K_1 must be selected within this physical constraint. On the other hand, feeding back the vehicle's yaw rate can

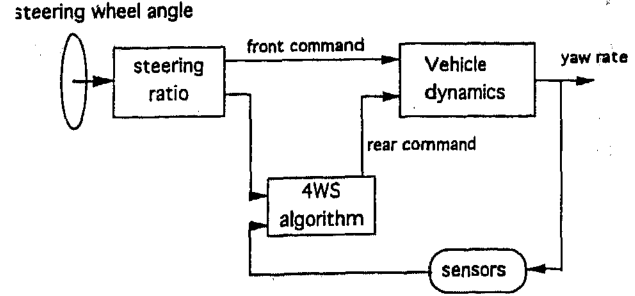


Figure 16: Four wheel steering controlled configuration

degrade the closed-loop stability of a four-wheel-steering vehicle. Care must be exercised to ensure that the damping ratio of the vehicle's yaw mode is always above a preselected minimum level. Appropriately selected, these control parameters allow us to vary the lateral dynamics of the VDTV so that it emulates the response characteristics of a broad range of vehicles.

For simplicity, the control parameters K_1 and K_2 are not made functions of the vehicle's forward speed in our study. In practice, they can be made functions of both the vehicle's forward velocity and other vehicle measurements such as its lateral acceleration. Other closed-loop 4WS algorithms have also been proposed in the literature.[18–19]

Plots of steering wheel angle versus the vehicle's lateral acceleration obtained from turn circle maneuvers, for four combinations of control parameters (K_1, K_2) are compared in Figure 17. In that figure, 4WS control algorithms that steered the front and rear wheels "in-phase" have positive K_1, and are denoted by "I1" and "I2". On the other hand, 4WS control algorithms that steered the front and rear wheels "out-of-phase" have negative K_1, and are denoted by "O1" and "O2". Plots that compare the understeer coefficients of the nominal two-wheel-steering vehicle, and the "I2" and "O2" four-wheel-steering vehicles are

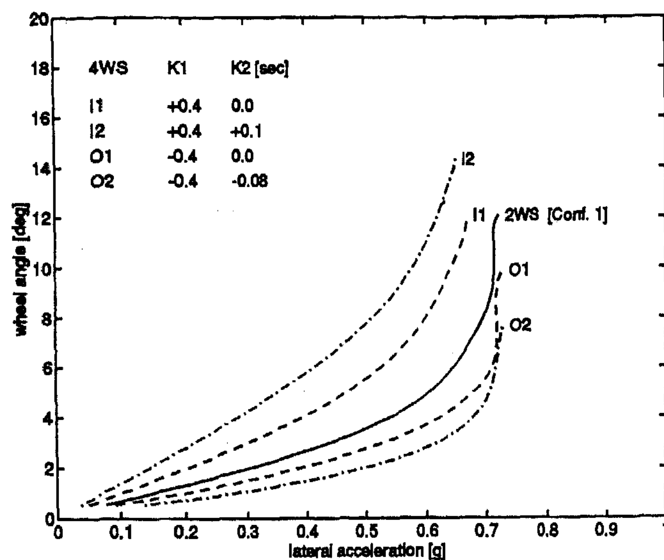


Figure 17: Turn cicrle maneuver results obtained with 4WS VDTV [Configuration I]

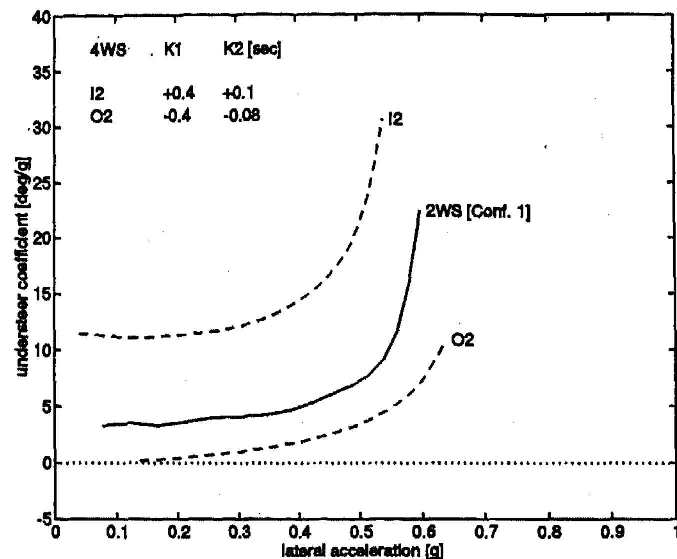


Figure 18: Understeer coefficient results obtained with 4WS VDTV [Configuration I]

shown in Figure 18. For clarity, results obtained for the "I1" and "O1" four-wheel-steering vehicles are omitted in that figure. The following observations can be made from Figure 18:

- the understeer coefficient of the four-wheel-steering vehicle is increased by steering the rear wheels in-phase with the front wheels, and can be further increased by having a negative yaw rate feedback (with positive K_2).
- the understeer coefficient of the four-wheel-steering vehicle is decreased by steering the rear wheels out-of-phase with the front wheels, and can be further decreased by having a positive yaw rate feedback (with negative K_2).

In Figures 17 and 18, the two-wheel-steering vehicle is denoted by "2WS [Configuration I]." Results obtained for a second two-wheel-steering vehicle, denoted by "2WS [Configuration II]" are given in Figures 19 and 20. The main difference between these two configurations is: all four tires used in the "Configuration I" are P195/75R14 tires (that are used by the Toyota Van in Reference 10). For "Configuration II", the two front tires of "Configuration I" are replaced by two P195/75R14 tires (that are used by the Chevrolet S10 in Reference 10).

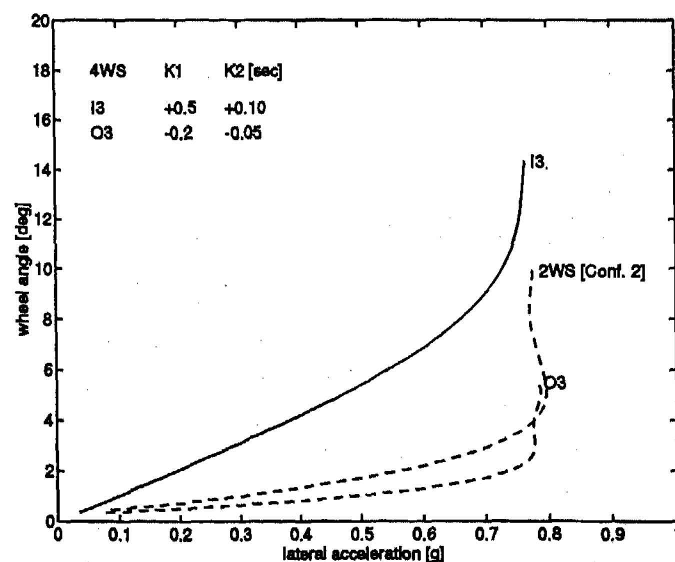


Figure 19: Turn cicrle maneuver results obtained with 4WS VDTV [Configuration II]

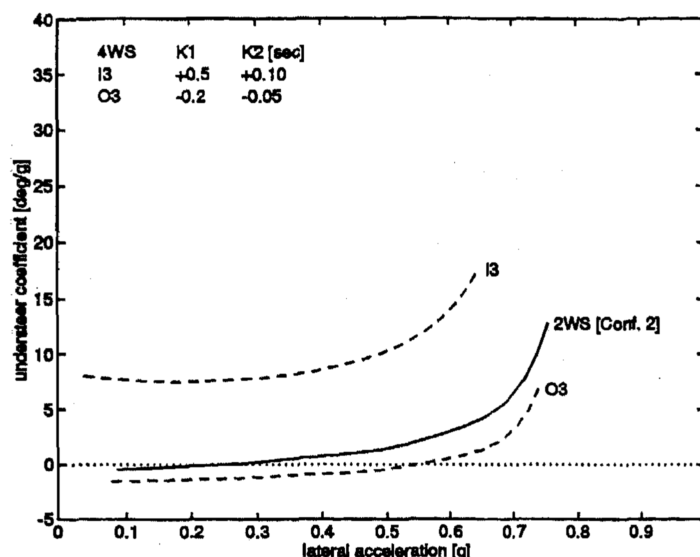


Figure 20: Understeer coefficient results obtained with 4WS VDTV [Configuration II]

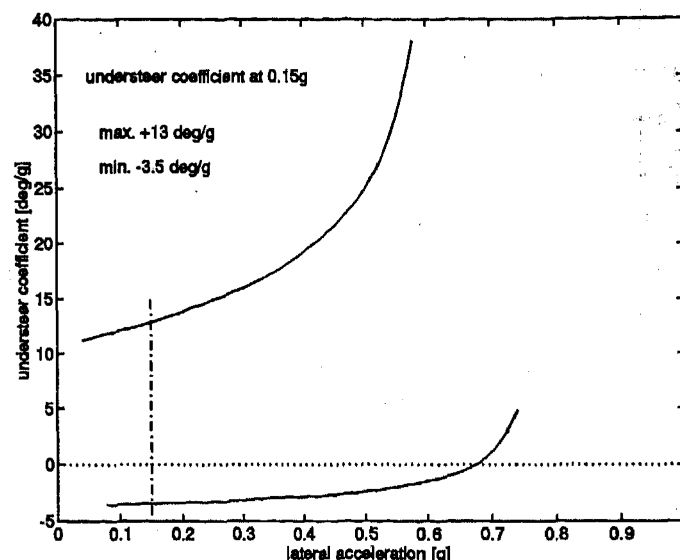


Figure 21: Emulation range of the VDTV understeer coefficient

Turn circle maneuver results obtained with the second two-wheel-steering vehicle and its four-wheel-steering derivatives are given in Figures 19 and 20. In Figure 20, the second two-wheel-steering vehicle controlled by the "O3" 4WS algorithm can generate negative understeer coefficient (i.e., oversteer). The emulation range of the VDTV's understeer coefficient, estimated using results given in Figures 18 and 20, is depicted in Figure 21. The upper bound in that figure was obtained by adding 25% to the result obtained for the "I2" four-wheel-steering vehicle (see Figure 18). The lower bound in that figure was obtained by subtracting 2.5 deg/g from result obtained for the "O3" four-wheel-steering vehicle (see Figure 20).

The effects that 4WS control algorithms have on the vehicle's 90% rise time in J-turn maneuvers are fairly significant. The following observations can be made from Figure 22:

- vehicle becomes more responsive (with a smaller J-turn 90% rise time) if the front and rear tires are steered in-phase and with a negative yaw rate feedback (positive K_2). See results obtained for the "I3" four-wheel-steering vehicle.
- vehicle becomes even more responsive if the stiffness of the front anti-roll bar is increased.

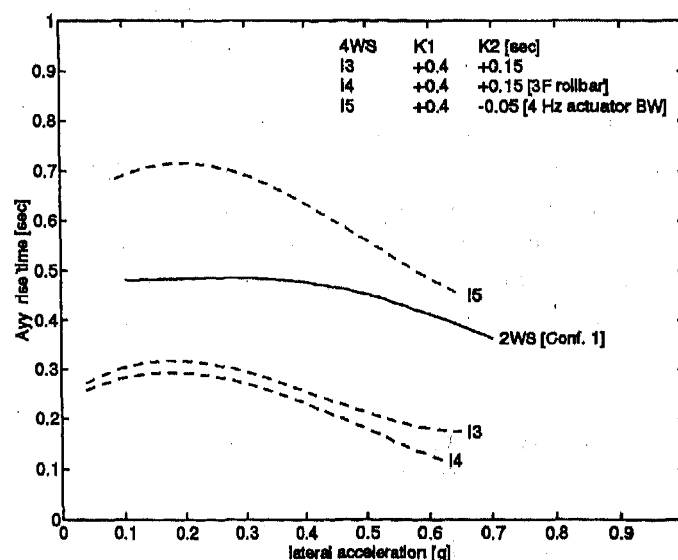


Figure 22: Rise time results obtained with a 4WS VDTV

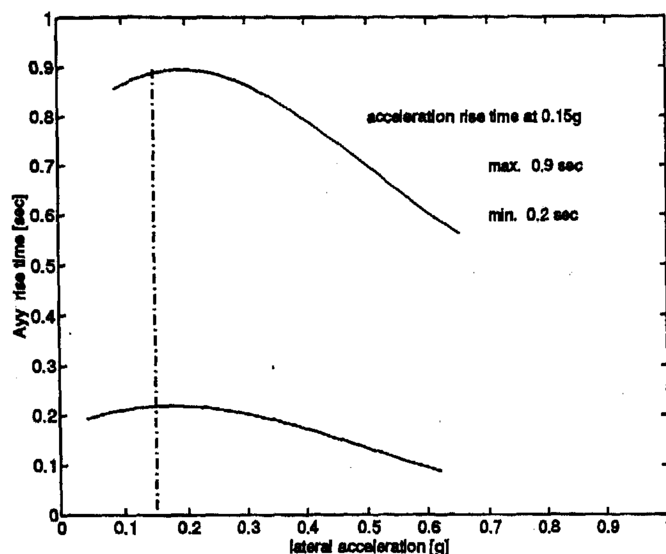


Figure 23: Emulation range of the VDTV J-turn 90% rise time

See results obtained for "I4" four-wheel-steering vehicle.

- vehicle becomes less responsive (with a larger J-turn 90% rise time) if the front and rear tires are steered in-phase and with a positive yaw rate feedback (negative K_2), as well as using a steering actuator with a lower bandwidth of 4 Hertz (the nominal bandwidth of the steering actuator is on the order of 10 Hz). See results obtained for the "I5" four-wheel-steering vehicle.

The emulation range for the VDTV's J-turn 90% rise time is depicted in Figure 23. The upper and lower bounds in that figure are obtained by increasing and decreasing, respectively, 25% of results obtained by the "I5" and "I4" four-wheel-steering vehicles that are depicted in Figure 22.

Results obtained from the Consumers Union Obstacle Course

Consumers Union obstacle avoidance course, depicted in Figure 9, was used to objectively evaluate the handling qualities of passenger vehicles during emergency double lane change maneuvers. Details of this obstacle avoidance course are defined in Reference 13. The VDANL closed-

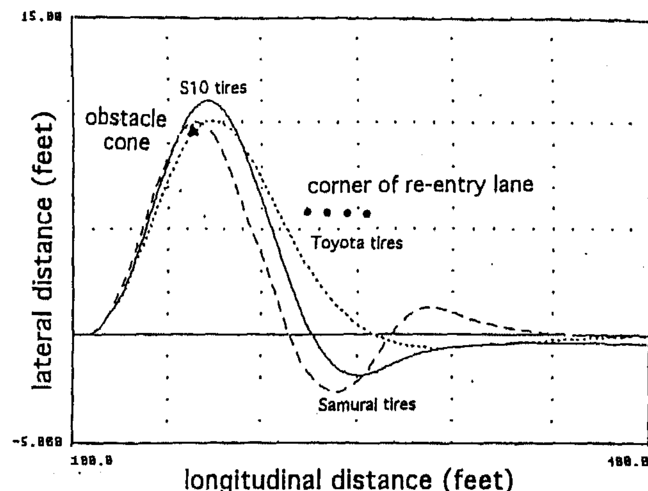


Figure 24: Obstacle avoidance lane change maneuver results

loop crash avoidance option was used to determine the maximum speed at which the baseline VDTV (with P195/75R14 Toyota Van tires) can successfully complete this double lane change maneuver. This analytical simulation was then repeated on two other VDTV variants: the first has P195/75R14 Chevrolet S10 tires and the second has P205/70R15 Samurai tires. The maximum speed was found using the Samurai tires, and was 52.7 km/hr. In Section 3.5.2.3 of Reference 14, we specified that the VDTV shall be able to successfully negotiate the Consumers Union obstacle course at all speeds below 55 km/hr.

Figure 24 displays the lateral versus longitudinal positions for the crash avoidance double lane change maneuver of the three VDTV variants with different tire models. As shown in that figure, all three VDTV variants can clear the corner of the re-entry lane, but only the variants with the S10 and Samurai tires can clear the obstacle cone. The VDTV with the Samurai tires is most "aggressive" because it can get back to the centerline of the re-entry lane in the shortest longitudinal distance. However, the vehicle trajectory beyond that "crossover" point is quite oscillatory. On the other hand, the vehicle trajectory obtained with the S10 tires, both before and after the crossover point is better damped. Time histories of the steering wheel angle, vehicle's lateral acceleration and sideslip angle, as well as tire

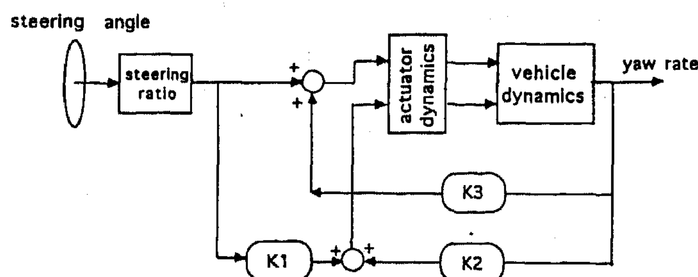


Figure 25: A combined four-wheel-steering and steer-by-wire control architecture

forces are given in Reference 15.

Concluding Remarks

A Variable Dynamic Testbed Vehicle concept has been proposed as a tool to perform driving-related human factors research. The goal of this study was to analytically investigate to what degree a VDTV with four-wheel-steering can emulate the lateral dynamics of a broad range of vehicle models. Using a selected compact-sized automobile as a baseline, our study indicated this vehicle can be controlled to emulate the lateral response characteristics (including the vehicle's understeer coefficient and the 90% lateral acceleration rise time in a J-turn maneuver) of a fleet of production vehicles, from low to high lateral acceleration conditions. Also, the roll gradient of the baselined vehicle can be altered via changes made to the torsional stiffnesses of the front and/or rear anti-roll bars to emulate the roll stiffnesses of a fleet of production vehicles.

The levels of emulation can potentially be improved if the VDTV has both the four-wheel-steering and steer-by-wire features depicted in Figure 25. With this steering configuration, the controller architecture given in (7) becomes:

$$\bar{\delta}_{rc} = K_1(U)\delta_{fc} + K_2(U)r\,, \quad (8)$$
$$\bar{\delta}_{fc} = \delta_{fc} + K_3(U)r. \quad (9)$$

Here, the variables δ_{fc} and r were defined in connection with the 4WSY control algorithm. Front and rear steering commands are given by $\bar{\delta}_{fc}$ and $\bar{\delta}_{rc}$, respectively. The added emulation benefits that one can derive with this controller architecture are not confirmed in this study, but experimental results obtained from a vehicle fitted with a similar steering system have shown promise.[3] The proposed VDTV, with both the steer-by-wire and four-wheel-steering features,[14] will be an ideal testbed to verify results obtained by the Simulator Vehicle of Reference 3.

Acknowledgments

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology. It was sponsored by the National Highway Traffic Safety Administration through an agreement with the National Aeronautics and Space Adminstration. The authors wish to thank their colleagues at Jet Propulsion Laboratory, including D. Griffin, M. Koffman, R. Norton, and R. Phen; W. Allen, D. Klyde, and T. Rosenthal at Systems Technology Incorporated, as well as J. Chrstos and R. Garrott at the Vehicle Research and Test Center for many helpful discussions. This research was conducted under the technical direction of L. Emery of the National Highway Traffic Safety Adminstration. All errors are our responsibility.

Disclaimer

The discussion in the text of this paper reflects the opinion and findings of the authors, and not necessarily those of the National Highway Traffic Safety Adminstration.

References

1. Leasure, W.A., Jr., "The NHTSA Collision Avoidance Program," IVHS America Workshop on Collision Avoidance, Reston, Virginia, April 21-22, 1994.

2. McKenna, K.J., "A Variable Response Vehicle - Description and Applications," Joint Automatic Control Conference, Austin, Texas, June 19-21, 1974.

3. Sugasawa, F., Irie, N., and Kuroki, J., "Development of A Simulator Vehicle for Conducting Vehicle Dynamics Research," International Journal of Vehicle Design, Vol. 13, No. 2, pp. 159-167, 1992.

4. Marriott, A., "Variable Dynamic Testbed Vehicle," SAE 950036, 1995.

5. Allen, W., Rosenthal, T., and Szostak, H., "Steady State and Transient Analysis of Ground Vehicle Handling," SAE 870495, 1987.

6. Allen, W., Szostak, H., Rosenthal, T., and Klyde, D., "Field Testing and Computer Simulation Analysis of Ground Vehicle Dynamic Stability," SAE 900127.

7. Bernard, J., et al, "Evaluation of Selected Vehicle Dynamics: Phase II Final Report," MVMA Project: IOWA 9114-C11302, Iowa State University, College of Engineering, June 1992.

8. Heydinger, G., "Vehicle Dynamic Simulation and Metric Computation for Comparison with Accident Data," NHTSA DOT HS 807 828, Final Report, March 1991.

9. Allen, W., Rosenthal, T., Klyde, D., Owens, K., and Szostak, H., "Validation of Ground Vehicle Computer Simulation Developed for Dynamic Stability Analysis," SAE 920054.

10. Allen, W., Szostak, H., Klyde, D., Rosenthal, T., and Owens, K., "Vehicle Dynamic Stability and Rollover," National Highway Traffic Safety Adminstration, DOT HS 807 956, Final Report, June 1992.

11. SAE Handbook (1992), Volume 4: "On-highway Vehicles and Off-highway Machinery," Vehicle Dynamics Terminology, SAE J670e. Published by SAE, Inc., 400 Commonwealth Drive, Warrendale, PA 15096-0001.

12. Lee, A.Y., "Emulating the Lateral Dynamic of A Range of Vehicles Using A Four-wheel-steering Vehicle," SAE 950304. See also SP-1074, "New Development in Vehicle Dynamics, Simulation, and Suspension Systems," pp. 11-22, 1995.

13. Riley, B. and Robinson, B., "Handling Test on Four-wheel-drive Multi-purpose Vehicles," Research Report 330, Transportation and Road Research Laboratory, U.K., 1991.

14. Griffin, D.C., "Variable Dynamic Testbed Vehicle (VDTV): Functional Requirements," JPL Publication D-13459, April 1996. Published by Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109-8099.

15. Klyde, D., Allen, W., and Rosenthal, T., "VDANL Sensitivity Analysis With A Modified Ford Escort," Working Paper 2544-1, December 18, 1995. Published by Systems Technology, Incorporated, 13766 South Hawthorne Blvd., Hawthorne, California 90250-7083.

16. Lee, A.Y., "Vehicle Stability Augmentation Systems Designs for Four Wheel Steering Vehicles," ASME Journal of Dynamical Systems, Measurements and Control, Vol. 112, No. 3, September 1990.

17. Fukunada, Y., Irie, N., Kuroki, J., and Sugasawa, F., "Improved Handling and Stability Using Four-Wheel-Steering," The 11^{th} International Conference on Experimental Technical Safety Vehicles, Washington, D.C., May 1987.

18. Takiguchi, T., Yasuda, N., Furutani, S., Kanazawa, H., and Inoue, H., "Improvement of Vehicle Dynamics by Vehicle-speed-sensing Four-wheel-steering System," SAE 860624, February 1986.

19. Lee, A.Y., "Performance of Four-wheel-steering Vehicles in Lane Change Maneuvers," SAE 950316. See also SP-1074, "New Development in Vehicle Dynamics, Simulation, and Suspension Systems," pp. 161-174, 1995.

Table 1

Estimated Values of Vehicle and Tire Parameters†

Parameter	LeCar	Sentra	Escort	T-bird	Mark VIII
Class	small	small	compact	mid-size	full-size
Year	1977	1983	1989	1987	1995
wheel base (m)	2.42	2.39	2.39	2.65	2.89
mean track width (m)	1.27	1.39	1.41	1.49	1.53
c.g. distance from front axle (m)	0.93	0.84	0.88	1.06	1.18
c.g. height (m)	0.54	0.52	0.56	0.56	0.55
track width ratio	1.18	1.34	1.26	1.33	1.39
wheelbase ratio	4.48	4.60	4.27	4.73	5.25
total weight (kg.wt.)	816	1068	1229	1649	1756
roll inertia (kg-m^2)	147.3	217.0	244.1	348.8	716.2
pitch inertia (kg-m^2)	797.7	1101.1	1342.4	2537.4	3500.2
yaw inertia (kg-m^2)	986.5	1315.3	1539.1	2976.1	3627.3
front/rear roll stiffnesses (Nm/deg)	165.1 182.8	393.7 295.3	379.8 379.8	305.4 356.2	329.1 421.8
front/rear roll damping (Nms/deg)	15.8 15.0	27.1 17.2	25.3 25.3	34.9 32.5	39.1 42.0
front/rear anti-roll bar stiffness (Nm/deg)	-145.8 -75.0	0 0	-224.8 0	-714.8 +87.7	-2118 -299.3
steering ratio (-)	20.8	16.9	18.2	14.3	13.4
tire model	P145/ SR13	P155/ SR13	P165/ 80R13	P215/ 70R14	P215/ 70R14
Calspan coeff's: A_0	1260	2380	0	733	733
A_1	13.20	9.21	15.66	19.50	19.50
A_2	1830	2280	2350	2900	2900
A_3	0.533	0.523	0.530	1.370	1.370
A_4	-31200	-7225	-24450	4420	4420

† Detailed information on these and other vehicle/tire parameters are available in Appendix E of Reference 10.

970562

Methodology for Validating the National Advanced Driving Simulator's Vehicle Dynamics (NADSdyna)

W. Riley Garrott and Paul A. Grygier
National Highway Traffic Safety Administration

Jeffrey P. Chrstos
JPC Engineering

Gary J. Heydinger
S.E.A., Inc.

Kamel Salaani and J. Gavin Howe
Transportation Research Center, Inc.

Dennis A. Guenther
The Ohio State Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper presents an overview of work performed by the National Highway Traffic Safety Administration's (NHTSA) Vehicle Research and Test Center (VRTC) to test, validate, and improve the planned National Advanced Driving Simulator's (NADS) vehicle dynamics simulation. This vehicle dynamics simulation, called NADSdyna, was developed by the University of Iowa's Center for Computer-Aided Design (CCAD). NADSdyna is based upon CCAD's general purpose, real-time, multi-body dynamics software, referred to as the Real-Time Recursive Dynamics (RTRD), supplemented by vehicle dynamics specific submodules. VRTC has "beta tested" NADSdyna, making certain that the software both works as computer code and that it correctly models vehicle dynamics. This paper gives an overview of VRTC's beta test work with NADSdyna.

The paper explains the methodology used by VRTC to validate NADSdyna. The techniques used to obtain vehicle parameters are discussed. Repeated experimental runs were performed at each test condition to generate sufficient data for statistical analyses. The processing paths for both experimental test data and simulation predictions are described. The techniques used to compare simulation predictions with experimentally measured data are presented.

The paper continues with an overview of the NADSdyna validation work performed for one vehicle, a 1994 Ford Taurus GL. Due to the complexity of this work, only an overview of this work is presented in this paper. Also included an overview of the simulation evaluation work for two other vehicle dynamics simulations, VDANL and VDM RoAD, that was performed using the Ford Taurus data. VRTC's efforts to improve NADSdyna's predicted steering system feel and to advance the state-of-the-art in steering system modeling are summarized.

INTRODUCTION

This paper presents an overview of the work that has been performed over the last three years by the National Highway Traffic Safety Administration's (NHTSA) Vehicle Research and Test Center (VRTC) to validate and improve the planned vehicle dynamics simulation for the National Advanced Driving Simulator (NADS).

The National Advanced Driving Simulator (NADS) is being developed by NHTSA and will be situated at the University of Iowa. The NADS will, when

completed in 1999, be the most capable driving simulator in the world. It will have advanced computer generated imagery with the capability of displaying 5.5 million pixels over a complete, 360 degree field of view. The use of higher pixel densities directly in front of the vehicle cab will provide close to real-life visual resolution in this critical area.

NADS will also have a motion base capable of generating large displacements, generating highly realistic kinesthetic cues. The motion base will consist of a yaw turntable, used to generate large yaw rotation angles, mounted on a six degree of freedom hexapod. The hexapod will be mounted on an X-Y Carriage that can generate up to 64 feet of movement in both the X and Y directions.

An advanced driving simulator such as the NADS must include a detailed, highly accurate vehicle dynamics simulation to predict the movements of the simulated vehicle in response to both control and disturbance inputs. The planned NADS vehicle dynamics simulation, called NADSdyna [6], was developed by the University of Iowa's Center for Computer-Aided Design (CCAD) with assistance from Iowa State University and the University of Wisconsin. NADSdyna is based upon CCAD's general purpose, real-time, multi-body dynamics software, referred to as the Real-Time Recursive Dynamics (RTRD) [7, 8, 35], supplemented by vehicle dynamics specific sub-modules. A control program (ICON_lite) is designed to support the non-simulator user in preparing a vehicle model at a workstation.

BRIEF DESCRIPTION OF NADSDYNA - The general purpose, multi-body, dynamics program, Real Time Recursive Dynamics (RTRD), was developed by the University of Iowa's Center for Computer-Aided Design and is the core dynamics of NADSdyna. The RTRD is used to precisely model chassis/suspension/steering linkages to capture geometric nonlinearities. It offers great capabilities in terms of rigid-body mechanism modeling, and accounts for all nonlinear kinematic effects, such as, roll steer, roll centers, and squat/lift forces. The RTRD can model either open or closed loop chains. The RTRD is based on a topological analysis of a rigid body mechanism formulated as modified recursive dynamics equations using minimum relative joint coordinates. The constraint library supports the following joints: revolute, spherical, translational, universal, spherical-translational, revolute-spherical, and translational-revolute. More joints can be added if needed. For closed chains, distance constraints, spherical, and revolute-spherical cut-joint types are supported. Springs, dampers, anti-roll bars, leaf springs are modeled as external forces, and nonlinear models can be used.

Five vehicle submodules are included in NADSdyna: power train, tires, brakes, steering and aerodynamics. These submodules provide NADSdyna with a broad range of functional capabilities and allow most production vehicles to be simulated. The following paragraphs contain a brief summary of the main capabilities of each of these submodules.

The power train submodule models the engine, transmission, and drive line. The engine is modeled using a map specifying engine torque as a function of throttle position and engine rpm. A simple, first order, time lag is also included. Both automatic and manual transmission are provided. For the drive line, transfer case, differential, and final drive gear box models are provided that permit most typical vehicles to be simulated. Vehicles may have front wheel drive, rear wheel drive, or four wheel drive.

The tire submodule models each of the vehicle's tires and calculates the spin angular velocity of each wheel. Four empirical tire models are provided for use in the simulation: Calspan's [36], System Technology Inc.'s [12], Gim/Nikravesh's [37], and a table look-up model. The wheel spin dynamics are solved using implicit integration of the first order differential equations for the longitudinal slips as suggested by Bernard [9]. This method allows simulation of the vehicle at low speed (including the singularity configuration, zero longitudinal velocity) and allows the simulated vehicle to stop smoothly on a hill.

The brake submodule includes models of the three main types of brakes: manual hydraulic brakes, power-assisted hydraulic brakes, and pneumatic brakes. Each type of brake can be modeled either with and without an antilock brake system.

The steering submodule can be modeled either kinematically or using a power assisted dynamic model. Vehicles with either rack and pinion or Pitman arm steering systems can be simulated.

The aerodynamic submodule calculates the aerodynamic forces acting on the vehicle. SAE J1594 is used to determine the quasi-steady state forces.

NADSdyna includes a number of pre-processors, which assist in the input of RTRD and submodule data, and initial vehicle parameter verification. The following pre-processors are provided: vehicle rigid-body, power train, tires, brakes, steering, and aerodynamics. Another pre-processor is used to run NADSdyna for non-real-time applications. The pre-processors are all OSF/Motif applications.

Two post-processors are provided with NADSdyna. One of these is used to perform "sanity" checks on NADSdyna data sets [38]. This post-processor has NADSdyna make a variety of standardized runs and checks that the resulting is sensible. The second post-processor assists users in examining and manipulating NADSdyna output. This tool is a Graphical User Interface built on top of Xmath, that allows users to easily examine NADSdyna output. It also includes an easy interface to Xmath and C that allows the user to manipulate NADSdyna output by creating custom functions and data reduction procedures.

OTHER COMMON U.S. DEPARTMENT OF TRANSPORTATION VEHICLE SIMULATIONS - While NADSdyna is an extremely capable vehicle dynamics simulation that is well suited for use with the NADS, it does have some drawbacks that make it less useful as a general purpose vehicle dynamics simulation. Specifically, NADSdyna requires an immense amount of vehicle parameter data, is relatively complicated to run, and only operates on UNIX workstations, not on typical personal computers. As a result, the U.S. Department of Transportation uses other, less complex, vehicle dynamics simulations for much of its vehicle simulation work. Two frequently used vehicle dynamics simulations are "Vehicle Dynamics Analysis, Non-Linear" (VDANL) and "Vehicle Dynamics Models for Roadway Analysis and Design" (VDM RoAD).

As is explained below, NHTSA decided to perform a complete validation of NADSdyna. Much of the difficulty and expense of performing vehicle dynamics simulation validation is in the vehicle parameter measurement and vehicle testing. Due to the high level of complexity of NADSdyna, researchers realized that no additional vehicle parameters measurements or vehicle testing would have to be performed to evaluate both VDANL and VDM RoAD at the same time. Therefore, the NADSdyna validation project also worked with VDANL and VDM RoAD.

BRIEF DESCRIPTION OF VDANL - The VDANL simulation used in this research is the latest version of the VDANL simulation being developed by Systems Technology, Inc. (STI) for the Federal Highway Administration (FHWA) as part of its "Interactive Highway System Design Model" (IHSDM). The original VDANL simulation was developed in the mid 1980's by STI for NHTSA as documented in [12, 14, 15].

STI has made numerous improvements to VDANL since the original version. Version 2 of VDANL, was developed by STI in 1989-1990 for NHTSA [10]. Versions 3 and 4 were developed during 1990 to 1995 when STI made VDANL a commercial product. Version 5.03 of VDANL, used for this research, contains basically the same vehicle and tire model as version 2. However, version 5.03 adds capabilities for trailer towing, simulation of tractor/semi-trailers, off-road tire operation, and complex terrain definitions (among other modifications/improvements). VDANL has been extensively evaluated by VRTC [17, 18, 32] and Iowa State University [21, 22]. The results of these evaluations have shown that VDANL is, in general, able to correctly predict vehicle directional responses, including transient behavior, up to moderately high g levels.

VDANL is a personal computer based program that runs under the DOS® operating system. Four degrees-of-freedom are used to model the vehicle sprung mass: pitch rotation, roll rotation, lateral translation, and vertical translation. Two additional degrees-of-freedom are used to model the total vehicle mass: yaw rotation and longitudinal translation. The front and rear unsprung masses are each modeled with three degrees-of-freedom: vertical translation, lateral translation, and roll rotation. Each wheel has a single spin degree-of-freedom, and the steering system has a single rotational degree-of-freedom. This gives a total of seventeen degrees-of-freedom for a passenger car (adding a two-axle trailer can add ten additional degrees-of-freedom).

The suspension model allows modeling both solid and independent suspension systems. Quasi-static compliance models of wheel camber and steer angle change due to tire lateral and longitudinal forces and aligning moments are included. VDANL contains a sophisticated tire model that predicts tire lateral and longitudinal forces and aligning moments over the complete range of slip angle, longitudinal slip, normal load, and camber angle. Open-loop control inputs of steering angle, throttle pedal position, and brake pedal force are allowed, along with a closed-loop speed controller that modulates the throttle to follow a user specified speed profile. The tire lateral forces are lagged using two first order lags to simulate tire dynamic characteristics.

BRIEF DESCRIPTION OF VDM RoAD - The VDM RoAD simulation from the University of Michigan Transportation Research Institute [16] is an eighteen degree-of-freedom model (UMTRI also developed a simpler fourteen degree-of-freedom simulation for the FHWA). There are no known previously published evaluations of VDM RoAD.

The VDM RoAD "package" consists of a "Simulation Graphical User Interface" (SGUI) [33], the simulation program executable (18DOF.EXE), and the UMTRI "Engineering Plotter" program. All user interaction is done through the SGUI, which allows vehicle parametric information to be entered, and simulation runs to be specified. The SGUI is a graphical front end to a database of vehicle parametric and simulation input/output files. Sample screens from the SGUI are shown in [25] and [33]. The VDM RoAD package evaluated in this report was run on a personal computer under the Microsoft Windows NT™ 3.51 operating system. Versions of VDM RoAD are also available to run on Macintosh and UNIX systems.

The vehicle simulation program has been developed by UMTRI using the AUTOSIM [34] multi-body code generator. A multi-body description of the vehicle model is input to AUTOSIM, and AUTOSIM symbolically derives the multi-body equations, and writes FORTRAN code. The compiled version of the FORTRAN code is called 18DOF.EXE. A description of the passenger car model can be found in the May/June 1995 progress report from UMTRI to FHWA. A more

complete description will be contained in UMTRI's final report.

18DOF.EXE models all six degrees-of-freedom for the vehicle chassis. Only independent suspension systems are modeled, and each wheel contains a vertical degree-of-freedom. The suspension models use tables to define spring and damper forces along with wheel toe and camber angles as a function of suspension motion. The four tire slip angles are modeled as first order differential equations producing tire lateral force and aligning moment lag. A fixed roll center model is used, and anti-pitch/dive forces are modeled. Wheel steer compliance to tire lateral and longitudinal force and aligning moment is modeled. The steering system model allows front Ackerman effects through table lookup as well as mechanical four wheel steering. Brake torque at each wheel is specified by a table as a function of brake pedal force. The six aerodynamic coefficients are specified as functions of aerodynamic side slip angle. Tire forces and moments can be specified using either tables or the STI tire model from version 1.00 of VDANL.

STEERING SYSTEM FEEL - One major area of concern for the NADS vehicle dynamics is the correct prediction of the "feel" of the steering system, i.e., the forces and moments that are fed back to the driver through the steering handwheel. However, steering system feel is also subjective as well as objective issue. When a subject is "driving" a simulated passenger car in the NADS, the subject should think that the steering system is that of an automobile as compared to that of, for example, a heavy truck. The subjectivity of driver impressions of steering system feel make correct prediction extremely difficult, because even small errors in the predicted forces and moments can spoil the subjective feel. As a result of this concern about steering system feel, a portion of this research has focused on this topic.

OBJECTIVES

NHTSA's Vehicle Research and Test Center is approximately halfway through a multi-year research project to test, validate, and improve the planned vehicle dynamics simulation for the NADS. The primary objectives of this project are:

1. To "beta test" the NADSdyna software originally developed by the University of Iowa's Center for Computer-Aided Design. NHTSA must be certain that this software works as computer code.
2. To demonstrate that NADSdyna correctly simulates the dynamics of a broad range of vehicles. Detailed validation work is being performed for two vehicles, a passenger automobile (a 1994 Ford Taurus) and a tractor-semi-trailer (a 1992 Volvo tractor with a Fruehauf van semi-trailer).
3. To develop four data sets that contain the parameters used to describe a vehicle to NADSdyna. The data sets to be developed will be for two passenger automobiles (a 1994 Ford Taurus; the second make/model is to be decided later), a sport utility vehicle (make/model to be decided later), and a tractor-semi-trailer (a 1992 Volvo tractor with a Fruehauf van semi-trailer).
4. To improve the functionality of NADSdyna beyond that originally included by CCAD. Examples of planned improvements include improving the prediction of steering system feel, adding failed vehicle subsystem models, adding interactions with roadway features (bumps and slippery patches), and adding the ability to model wind gusts.

The secondary objective of this project is evaluate two other vehicle dynamics simulations that are used by the U.S. Department of Transportation, VDANL and VDM RoAD. This work was performed because much of the difficulty and expense of performing vehicle dynamics simulation validation is in the vehicle parameter measurement and vehicle testing. Due to the high level of complexity of NADSdyna, no additional vehicle parameters measurements or vehicle testing had to be performed to evaluate both VDANL and VDM RoAD at the same time.

The primary objective of this paper is to present an overview of the research that has been performed by the Vehicle Research and Test Center on this project to date. The companion papers [1, 2, 3, 4, and 5] present additional details of this work. Several reports describing this research [20, 23, 24, 25, 26, 27, and 28] are either being written or have been completed. The current paper attempts tie these papers and reports together.

The major focus of this project's work, to date, has been validating NADSdyna for one vehicle, a 1994 Ford Taurus. The methodology used to perform this validation work is explained below. As this methodology explains, there are three major phases in simulation validation:

1. Experimental data collection. The work performed during this phase for the Ford Taurus is described in greater detail in the companion paper [1] and in the report [20].
2. Vehicle parameter measurement. The work performed during this phase for the Ford Taurus is described in greater detail in the companion paper [2] and in the report [24].
3. Comparison of simulation predictions with experimental data. The work performed during this phase for the Ford Taurus is described in greater detail in the companion paper [3] and in the report [23].

The Ford Taurus data was also used to perform simulation evaluation work for two other vehicle dynamics simulations, VDANL and VDM RoAD. This work is discussed in greater detail in the companion paper [4] and in the report [25].

VRTC's efforts to increase NADSdyna's functionality have, to date, focused on improving NADSdyna's predicted steering system feel. While VRTC plans to improve a number of other areas of NADSdyna, work has not yet begun in these other areas. More details about VRTC steering system research are presented in the companion paper [5] and in the reports [26, 27, and 28].

VALIDATION METHODOLOGY

The simulation validation methodology used in this project is an improved version of the methodology that was described in [18]. This methodology consists of three main phases, experimental data collection, vehicle parameter measurement, and comparison of simulation predictions with experimental data. This section of the paper begins with a definition of simulation validation followed by a detailed description of each of the main validation phases.

DEFINITION OF SIMULATION VALIDATION - A computerized, mathematical model of a physical system, such as a handling and control vehicle dynamics simulation, will be considered to be valid if, within some specified operating range of the physical system, a simulation's predictions of the system's responses of interest to specified input(s) agree with the actual physical system's responses to the same input(s) to within some specified level of accuracy. This definition is compelling and sound, but measuring the actual physical responses is not an error-free exercise due to the inability to control the systematic and random errors within the system. Therefore, the validation confidence will be improved if more than one vehicle are used. Analytical diagnostics, or sanity checks, should be applied along with this definition so that the simulation predictions match the desired model behavior.

The above definition contains several points that need to be explained in greater detail.

First, a simulation's predictions will, in general, only be correct within some portion of the physical system's operating range. An obvious example of this is that a vehicle dynamics simulation's predictions may be correct for low lateral acceleration maneuvers but become progressively worse as lateral acceleration increases and non-linear effects become more important. A second example is that, due to an incorrect modeling of the effects of steering system compliance, a simulation's predictions might get worse as steer angle increases.

Similarly, a simulation's predictions may only be correct for control inputs that predominantly contain frequencies within some specified range. For example, some vehicle dynamics simulations are valid for steady state and relatively slowly varying input conditions but have problems with fast transients that contain high frequencies.

Therefore, simulation validation should be performed in both the time and frequency domains. Validation in the time domain is good for demonstrating that the simulation can correctly predict steady state conditions and that non-linear effects are properly modeled. High frequency transient phenomena, however, are very difficult to study in the time domain. The effects of increasing input frequency on the correctness of a simulation's predictions are best determined through frequency domain studies.

A second significant point contained in the definition of simulation validation is that it is valid only for specified inputs/outputs groups. For example, for a handling and control simulation, simply because the simulation has been shown to be valid for braking and steering control inputs does not imply that the response to a road disturbance (such as a bump in the road) will be correctly predicted. Similarly, a simulation that successfully predicts lateral sprung mass acceleration might fail to predict vertical sprung mass acceleration.

The third significant point in the simulation validation definition is that whether or not a simulation can be considered to be valid depends upon how much the simulation's predictions can acceptably vary from the actual test results at a given operating point. The level of accuracy required to classify a simulation as valid depends on the intended uses of the simulation and the accuracy believed to be attainable. If only the trends of the response of a physical system are to be simulated, with little interest in predicting values, much less accuracy is required than when trying to predict exact values.

EXPERIMENTAL DATA COLLECTION PHASE - Vehicle dynamics simulation validation involves comparing a simulation's predictions of a vehicle's responses to control inputs (steering, braking, throttle) and disturbance inputs (wind, road features) to the actual vehicle's responses to the same inputs. While information on the actual vehicle's responses to inputs (the validation data) may be determined in any of several ways, in general, the best way is through experimental testing of the actual vehicle. This allows direct comparison between a simulation's predictions and the actual vehicle's responses.

The maneuvers performed during experimental testing should cover a broad range of vehicle operating conditions. The maneuvers performed for this research can be divided into six major categories. These are:

1. Steady state lateral performance — These primary validation maneuvers are low frequency turns. They are used to determine the vehicle's steady state lateral gains and understeer.
2. Transient lateral performance — These primary validation maneuvers input a broad range of frequencies at the steering handwheel. They are used to measure the vehicle's frequency response functions and thereby characterize the vehicle's

lateral response to transient inputs. Experience has shown that comparisons in the frequency domain between simulation predictions and actual vehicle responses are usually the best way to validate the transient portions of a simulation.

3. Longitudinal acceleration maneuvers — These primary validation maneuvers are performed either in a fixed gear or with the transmission allowed to change gears. They are used to measure the vehicle's response to throttle inputs.
4. Longitudinal deceleration maneuvers — These primary validation maneuvers are performed to measure the vehicle's response to braking inputs.
5. Road disturbance input maneuvers — These primary validation maneuvers are performed to determine the vehicle's to selected road disturbance inputs.
6. Other driving maneuvers — These maneuvers are not the primary maneuvers used for simulation validation. Instead, they are used to build confidence that validation results from the primary maneuvers can be generalized to other maneuvers including maneuvers in with multiple inputs.

The first five types of maneuvers are characterized as primary validation maneuvers. These are the minimum set of maneuvers needed for validation.

The first and second types of maneuvers characterize both the steady state and transient lateral performance of the vehicle, respectively. Unfortunately, braking and throttle maneuvers always contain both steady state and transient portions. Therefore, for the vehicle's braking and throttle responses, both steady state and transient characteristics must be determined from the same maneuvers.

Repetitive experimental testing of the actual vehicle allows statistical methods to be used to determine the random uncertainty present in the experimental testing and measurement process. Every experimental measurement contains random error superimposed onto the signal. Random errors are defined here as transducer measurement noise, errors due to unaccounted for variations disturbing the vehicle, and random minor changes in the vehicle. For vehicle testing, random error includes, for example, the effects of wind gusts, road roughness, tire non-uniformity, and changes in the brakes from test to test. Note that there are other sources of experimental non-repeatability including, for example, variability in the inputs.

Some of the disagreement between a simulation's predictions and experimental measurements will be due to this random error and errors in parameter measurements for the simulation. Since simulations cannot predict random error, if a simulation's predictions agree with experimental measurements to within the experimental random error level, the simulation should be considered to be valid.

Conversely, while simulations whose predictions are less accurate than the experimental random error level are useable for many tasks, the "best" simulation (for a specified operating regime and input/output set) of a vehicle is one whose predictions agree with experimental measurements to within the random error level.

The easiest way to determine the experimental random error level present in data is to repeat all experiments that are performed for the purpose of simulation validation multiple times. Given data from several tests, statistical procedures can be used to calculate the random error levels.

VEHICLE PARAMETER MEASUREMENT PHASE - A vehicle dynamics simulation is based on a set of equations derived from a model of the vehicle being simulated. Adjustable values, known as vehicle parameters, in these equations describe the particular vehicle or vehicle configuration being studied to the simulation.

During development of a simulation, each vehicle parameter should be unambiguously defined and methods developed to measure or obtain each value that is input to the simulation. Parameters that cannot be clearly defined or for which there is no means of measuring/obtaining are not acceptable in a simulation. Parameter definitions and techniques for measuring/obtaining parameter values should be included in a simulation's documentation.

During validation, vehicle parameters should be measured/obtained independently from the simulation validation data. If the vehicle parameters are not obtained independently, or if their values are adjusted during the validation process not due to improvements in the parameter determination process, but solely to improve the match between a simulation's predictions and the validation data, then the validation process becomes, at least in part, an exercise in curve fitting. While the ability to adjust the vehicle parameters to make a simulation's predictions match validation data is a strong argument for the correctness of a model, it is preferable to match simulation predictions made with independently measured, non-adjusted, vehicle parameters to validation data.

Simulation validation should be performed using several different vehicles. Using only one set of parameters (vehicle) could allow a problem to escape notice. The different vehicle parameter sets used should be for vehicles that differ substantially from each other. For the current work, validation is being performed for two very different vehicle, a medium-sized passenger automobile and a large tractor-semi-trailer.

COMPARISON OF SIMULATION PREDICTIONS WITH VALIDATION DATA - After the vehicle parameters have been measured, a simulation data set prepared, and the validation data collected and reduced, comparisons can be made between the simulation's predictions and the validation data. The simulation's predictions are calculated by feeding the

control inputs that were measured during the collection of the validation data into the simulation. The resulting simulation output is graphically compared to the experimental means and 95 percent confidence limits of the validation data.

FORD TAURUS EXPERIMENTAL DATA COLLECTION

This section briefly describes the data collection system, and data analysis used for the test program. Table 1 lists the data channels, the types of sensors used, and the units in which the data is stored. A more detailed description of the experimental setup and data storage system can be found in Reference [20].

DATA ORGANIZATION - Because of the size and scope of this vehicle testing project (more than 40 transducers, 75 maneuver types and conditions, and nearly 600 individual test runs), data organization is very critical. Therefore, guidelines were written for data file naming conventions, data file storage locations, transducer mnemonic naming conventions, and data units.

Because of the size of the testing program, record keeping is a major concern. For example, when simulating a particular test, the measured vehicle weight from the day of the test is necessary (during the course of this testing program, the total vehicle mass varied more than 70 kilograms, which is approximately 4 percent of the total mass). Daily test information is stored directly with the data. Some of this information is used by the data analysis programs, while some is simply for archival purposes.

DATA ACQUISITION EQUIPMENT - All on-board measurements are made using an Optim Megadac 2210C data recorder, that can digitize up to 128 analog channels with a maximum overall sample rate of 20,000 samples per second. All analog channels are sampled using a 16-bit analog-to-digital converter. The Megadac also has four 16-bit digital counters that may be cascaded.

Data Channel	**Sensor (units)**
Vehicle accelerations in X, Y, and Z-axes	Accelerometers (m/s^2)
Vehicle angular rates about X, Y, and Z-axes	Rate sensors (rad/s)
Power steering rack pressures	Pressure transducers (kPa)
Optical event	Optical trigger (V)
Chassis height	Ultrasonic distance sensors (m)
Brake line pressures	Pressure transducers (kPa)
Tie rod forces	Strain gages (N)
Vehicle speed	Tachometer generator (m/s)
Distance traveled	Fifth wheel (m)
Brake pedal force	Force transducer (N)
Transmission gearshift selector	Engine control module (V)
Engine speed	Engine control module (rad/s)
Gear transmission is in	Engine control module (V)
Wheel rotational velocities	ABS wheel sensors (rad/s)
Handwheel angle	Potentiometer (rad)
Handwheel rotational velocity	Tachometer generator (rad/s)
Handwheel torque	Strain gages (N-m)
Throttle position	Engine control module (%)
Strut accelerations along strut axis	Accelerometers (m/s^2)
Strut lengths	Potentiometers (m)

Table 1. Data Channels

At VRTC, the Megadac is controlled using a Grid model Gridcase 1500-series laptop computer that runs custom software to allow the system to be set up for each specific test. The test driver's primary interaction is through the laptop interface.

For the majority of the maneuvers run for this program, open-loop control inputs are desired. To make the control inputs as "open-loop as possible", three devices are used. Steering wheel angle inputs are constrained by a mechanical steering stop, a pneumatic ram is used to actuate the brake pedal, and a mechanical throttle pedal actuator is used during drive train testing.

TEST MANEUVERS - For many of the maneuvers, approximately 2 seconds of pre-trigger data are collected. This data is used to characterize the pre-maneuver state of the test vehicle. The pre-trigger data collection is a feature of the Megadac. After the Megadac is armed, data is collected and stored in a user definable size circular memory buffer. When the Megadac is triggered, the content of the buffer is written to the tape followed by the post-event data.

For maneuvers where repeatable open-loop inputs can be achieved, each test is repeated up to 10 times. The data from these repeat tests is then ensemble averaged in the time domain to obtain statistical measures of test repeatability. The maneuvers are summarized in [20].

DATA REDUCTION - Raw vehicle data must be converted into a form suitable for simulation evaluation. The process may involve (depending on the test) removing offsets, digital filtering, computing derived channels from measured channels, ensemble averaging in the time domain, and/or conversion of the time-domain data into the frequency domain. All data processing is done using the software program MATLAB, version 4.2c [3]. Some of the analysis also uses MATLAB Signal Processing Toolbox version 3.0b.

The first step in the processing of all raw data files is to remove channel offsets, digitally filter the data, and compute the derived channels. The MATLAB function BUTTER is used to design a two-pass phaseless Butterworth low-pass digital filter for all channels except longitudinal velocity, with a 3-dB down cutoff frequency of 10 hertz and a total of 12 poles. Because of significant cyclic noise in the longitudinal velocity channel (at the frequency of the fifth wheel rotational speed), a second filter is designed for the longitudinal velocity channel with a 3-dB down cutoff frequency of 2 hertz and a total of 6 poles. This additional filtering of longitudinal velocity was necessary to get "reasonable" wheel slip estimates.

A number of "derived" channels are also computed and included in the output file. Derived channels are channels computed by combining other channels, for example chassis roll angle, which is computed from the difference between the two ultrasonic ride height measurements. The derived channels are front and rear strut lengths, average steering wheel torque, steering rack position, power steering assist force, transmission gear number, chassis roll and pitch angles, and longitudinal wheel slip ratios.

TEST REPEATABILITY - When running repeat testing, one source of test-to-test variation that can be controlled is the repeatability of the vehicle control inputs. For this testing program, the control inputs are steering wheel angle, brake pedal force, throttle position, and for the constant speed runs, cruise control speed setting. Using the 33 meters/second, 6 meters/second^2 J-Turn tests, the repeatability of the steering wheel inputs and the vehicle longitudinal velocity are demonstrated. All of the filtered steering wheel angle data falls within approximately a 1-degree window, demonstrating that very good steering angle repeatability is achieved using the VRTC steering stop. Also, the average vehicle velocity from all 10 tests falls within a window of approximately 0.5 meters/second.

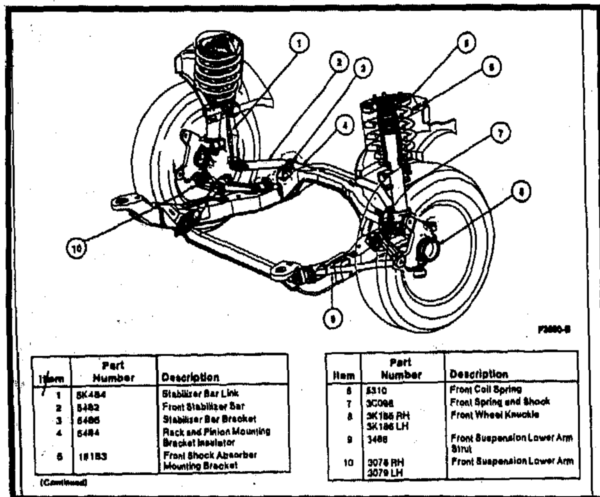
The throttle position repeatability is demonstrated by the 50-percent throttle, straight-line acceleration runs. Excellent repeatability is achieved using VRTC's mechanical throttle pedal actuator, with all of the data falling within a window of approximately 1 percent of throttle position. The brake pedal force repeatability is from the 5 meters/second^2 straight-line braking tests. The brake pedal force from all nine runs falls within a 6-Newton window, demonstrating good repeatability using the pneumatic brake ram.

FORD TAURUS VEHICLE PARAMETER MEASUREMENT

A passenger vehicle is a mechanical system consisting of a finite number of constrained rigid bodies. It depends on the vehicle type and the dynamic problem under consideration what type of constraints, how many bodies, and which degrees of freedom have to be assigned. The vehicle is assumed to be symmetric, and all the joints forming the suspensions are assumed to be frictionless. The front suspension is a MacPherson strut type as shown in Figure 1. The multi-body dynamics model shown in Figure 2 provides the front suspension with two degrees of freedom with respect to the chassis; vertical movements, and rotations about the king pin axis. The rear suspension is a MacPherson strut type with two parallel control arms as shown in Figure 3. The multi-body model shown in Figure 4 gives the rear wheel one degree of freedom about the chassis, a circular vertical movement around the tension strut link. In the actual vehicle, due to bushing compliance, the wheel moves in an "elliptical" circular direction as forced by the control arms and the tension strut. The RTRD's recursive formulation requires rigid bodies to be connected in an open loop structure, or a spanning tree.

The Taurus model contains ten closed loops, a joint on each loop is "cut" to get a spanning tree and the constraint equations associated with Lagrange multipliers are imposed on these cut joints to represent kinematically admissible motion. Graph theory is used to identify the topological structure of multi-body dynamic systems, and weighting factors of some general kinematic joints are used for the purpose of minimizing the chain length and maximizing efficiency in parallel computation [6].

The Inertia and center of gravity height were measured using the IPMD (Inertial Parameter Measurement Device) [39], located at VRTC. The location of all the suspension attachment points on the chassis, center of gravity, mass, and inertia of each component were measured. Bi-filar and tri-filar torsional pendulum's are used to measure the mass moments of inertia of small components. Nonlinear models of the spring, bump stop, and shock absorber are used, and the measurements were performed by Heitz Automotive Testing, Inc. The tire used during all field testing were General Ameritech P205/65R15 belted radials. The tests were performed at Smithers Scientific Services, Inc. on their MTS Flat-Track II flat-belt tire force and moment measuring machine. A complete description of the experiments can be found in [29]. The STI tire model is used , and the parameters were obtained using a preprocessor [40]. A quasi-static model of the compliances is used to avoid the inclusion of stiff elements, and the measurements were performed at Heitz Automotive Testing, Inc. [41]. The anti-roll bar stiffness was measured at VRTC and a quasi-static model is used. The aerodynamic coefficients were measured at the Lockheed Aeronautical Systems Company's Low Speed Wind Tunnel. The brake model with a dynamic lag is used (pneumatic model of NADSdyna), and the parameters were measured at VRTC. A detailed explanation of the measurements and NADSdyna data set can be found in [2, 23, 24, 30]. The power train and a speed dependent power steering system will be added in a later version of the simulation.



Item	Part Number	Description
1	5K484	Stabilizer Bar Link
2	5482	Front Stabilizer Bar
3	5486	Stabilizer Bar Bracket
4	5484	Rack and Pinion Mounting Bracket Insulator
5	18183	Front Shock Absorber Mounting Bracket

(Continued)

Item	Part Number	Description
6	5310	Front Coil Spring
7	3C098	Front Spring and Shock
8	3K185 RH 3K186 LH	Front Wheel Knuckle
9	3468	Front Suspension Lower Arm Strut
10	3078 RH 3079 LH	Front Suspension Lower Arm

Figure 1. 1994 Ford Taurus GL Front Suspension

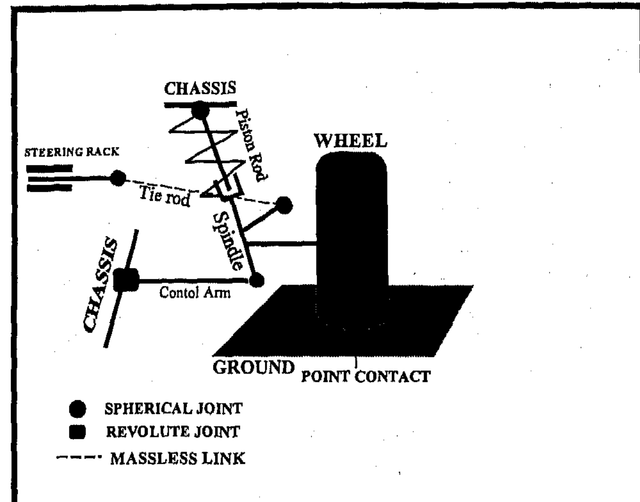
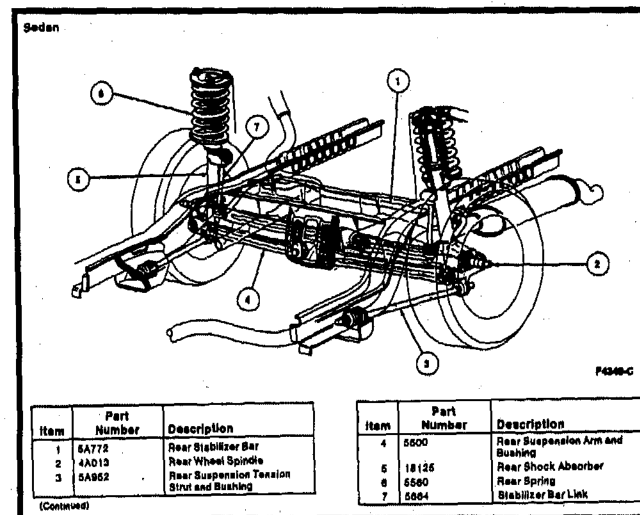


Figure 2. 1994 Ford Taurus GL Front Suspension Model



Item	Part Number	Description
1	5A772	Rear Stabilizer Bar
2	4A013	Rear Wheel Spindle
3	5A952	Rear Suspension Tension Strut and Bushing

(Continued)

Item	Part Number	Description
4	5500	Rear Suspension Arm and Bushing
5	18125	Rear Shock Absorber
6	5560	Rear Spring
7	5664	Stabilizer Bar Link

Figure 3. 1994 Ford Taurus GL Rear Suspension

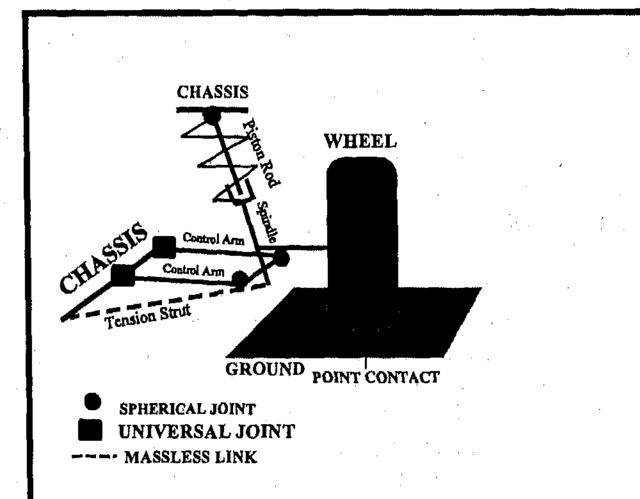


Figure 4. 1994 Ford Taurus GL Rear Suspension Model

NADSdyna VALIDATION FOR THE FORD TAURUS

The utility of the analysis of the aforementioned NADSdyna model depends on the extent to which it has

been evaluated. NADSdyna simulation results compared with experiments are shown in [3, 9, 31]. The following maneuvers are studied:

SLOWLY INCREASING STEER - The slowly increasing steer maneuver involved supplying the simulation with a slowly increasing handwheel angle, while maintaining a constant vehicle speed, until maximum vehicle lateral acceleration, rollover, or a spinout is achieved. A handwheel angle rate about *10 deg/sec* is used to eliminate dynamic effect and maintain steady state properties. The results of this maneuver indicated that the vehicle model has the appropriate non-linear suspension properties, as well as the tire model is appropriate to model severe cornering.

STEP STEER INPUT - This maneuver (J-turn) is a pseudo-step input of handwheel angle. The J-turn maneuvers are simulated to determine steady-state as well as transient responses. A physically realistic input rate on the order of *500 deg/sec* is used. The transient response comparisons between field test and computer model show good agreement in terms of timing of the transient events and peak levels of the variables that indicate limit performance conditions.

DIRECTIONAL DYNAMICS FREQUENCY RESPONSE (PULSE STEER) - Linear frequency responses can be developed for linear, time invariant systems, or approximated for a nonlinear system about some linearized operating condition. For vehicle handling dynamics, frequency responses are derived only for constant speed maneuvers, since handling responses are speed dependent. Additionally linearity is assumed applicable below lateral acceleration levels of *0.3-0.4* g's. The directional dynamic properties of the simulation are reasonable as indicated by the yaw rate and lateral acceleration frequency responses, which give some credibility to the composite lateral/directional dynamics. The roll angle measurements have a high noise to signal ratio, and the dynamics trends are reasonably predicted.

LANE CHANGE MANEUVERS - Once the steady state, transient, and frequency response predictions of the simulation have been evaluated, the simulation is checked against more complex maneuvers designed to simulate real world driving conditions, like lane change maneuvers. In this maneuver the steering input represents the driver/vehicle closed loop following performance. Figure 5 shows that the simulation predicted well the lateral acceleration, and the yaw rate.

BRAKING - The braking maneuvers are performed at different levels of severity from low g up to the limit. Despite uncertainty in road friction and other parameters, the simulation predictions are reasonable. A combined braking and steering maneuver is not necessarily valid since the tire forces are extremely non-linear. Predictions for this maneuver require accurate tire force model for the particular tire/surface used. The tire model does not fit well the experimental data at composite slip conditions [23, 28], which will be improved in a later version of the simulation software.

RIDE ANALYSIS - A simplistic approach is applied where road roughness is considered to be the input of vehicle vibration response. The transfer function of vehicle vertical acceleration at the chassis center of gravity to road acceleration input shows that the natural frequency of the chassis, and wheels, as well as the amplitude attenuation at higher frequencies are reasonably predicted by the simulation.

The evaluation of the 1994 Ford Taurus NADSdyna model have shown that the simulation predicted the fundamental mechanics of vehicle handling responses in general. The results show promise that NADSdyna will provide the fidelity required for the NADS vehicle dynamics simulation. This evaluation effort is being extended to include acceleration maneuvers, a complete dynamic steering system, as well as other vehicles including a tractor semi-trailer.

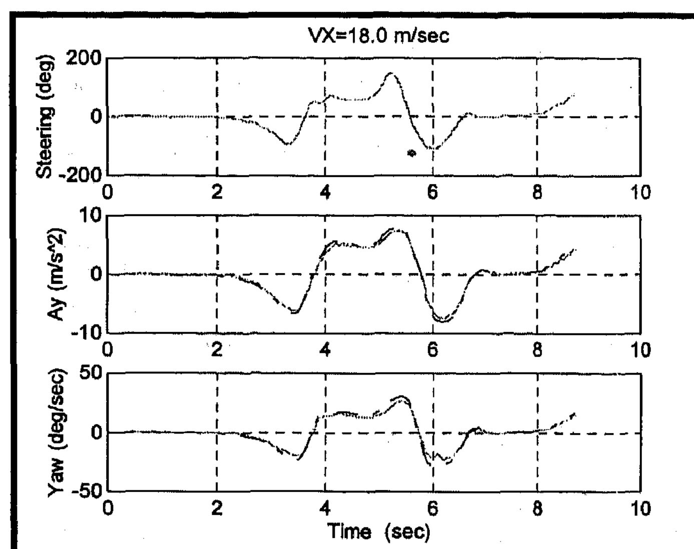


Figure 5. Lane Change Maneuver

VDANL AND VDM RoAD EVALUATION FOR THE FORD TAURUS

REASONS FOR INTEREST IN VDANL AND VDM RoAD - As part of the IHSDM research program, the Federal Highway Administration is using two vehicle dynamics simulations. These simulations are VDANL and VDM RoAD. The Federal Highway Administration's IHSDM program is a comprehensive project that is being developed to aid highway designers and traffic engineers in designing and implementing safe highway arteries and peripheries including, on-off ramps, interchanges, roadsides, medians, etc. The two simulations are an

integral part of the overall IHSDM program. Working in cooperation with the Federal Highway Administration, VRTC has evaluated these vehicle dynamics simulations.

The evaluations of VDANL and VDM RoAD are presented in two documents, the companion SAE paper [4] and a NHTSA report [25]. The SAE paper [4] evaluates the simulations' predictions in the linear range via comparison with analytical models. Consistency checks are made and lateral load transfer, roll gradient, and understeer predictions are evaluated. Linear range contributions of roll steer, roll camber, lateral force compliance, and aligning moment compliance to simulated overall vehicle understeer are computed and compared to the those computed using the analytical model. Both simulations do a good job of predicting expected vehicle responses in the linear range. Also in the linear range, the VDANL and VDM RoAD predictions of yaw rate frequency response are computed and shown to compare well with measured results from the 1994 Ford Taurus. Finally, to evaluate the nonlinear range of vehicle response, understeer gradient plots are generated up to the limit of lateral acceleration, and the results compared with experimental results. Both simulations are shown to predict the trends in the nonlinear understeer plots.

The NHTSA report [25] provides a thorough comparison of the VDANL and VDM RoAD predictions with experimental results from a 1994 Ford Taurus. Many types and severities of steering maneuvers and braking maneuvers are evaluated; and both qualitative and quantitative measures of the simulations' performance are presented. The report concludes that both simulations contain an appropriate order of model dynamics in the linear range of vehicle operation, that they both correctly predict gains and trends into the nonlinear range, and that they correctly predict longitudinal dynamics.

IMPROVING STEERING SYSTEM MODELING

VRTC's efforts to evaluate simulator performance and efforts to further develop steering system models are closely related. Several projects have been completed in support of the development of the steering system model and efforts are currently underway to develop a final steering system model and to properly evaluate its performance.

One project deals with the analytical modeling of the steering system. This project involves modeling the necessary steering system components (inertias, stiffnesses, damping, lash, power assistance, etc.) to fully capture the essence of steering feel via torque feedback through the steering wheel. The steering system model evaluations were done by implementing them in a VDANL-like vehicle dynamics computer simulation and studying the results for soundness using criteria such as: are the torques fed forward and backward through the steering system reasonable, are the time lags and hysteresis correct as modeled, etc. Two models have been developed and are currently being evaluated and refined.

Another project was the implementation of a steering system control hardware in VRTC's fixed-based driving simulator. This task included installing the steering torque feedback servo-motor, torque sensor, and steering angle and rate sensors in the vehicle used in VRTC's simulator. Once installed, these components were evaluated to characterize the performance measures (gain, bandwidth, and phase) of the torque feedback system.

The next step in VRTC's effort to develop tools for evaluating simulator performance is to implement and coordinate the analytical steering system models with the steering system control hardware on the simulator. With knowledge of the performance of the steering system hardware, suitable control logic can be implemented into the steering system model to compensate for the response of the hardware and generate the desired steering feedback torques to the driver.

Once implemented in VRTC's driving simulator, the entire system will be evaluated both objectively and subjectively. Objective measures, such as feedback torque gains and phase shifts, will be measured on the simulator and compared with similar measurements taken from an actual vehicle. Similarly, subjective performance measures of the simulator and an actual vehicle will be made to appraise the fidelity of the simulator's steering system.

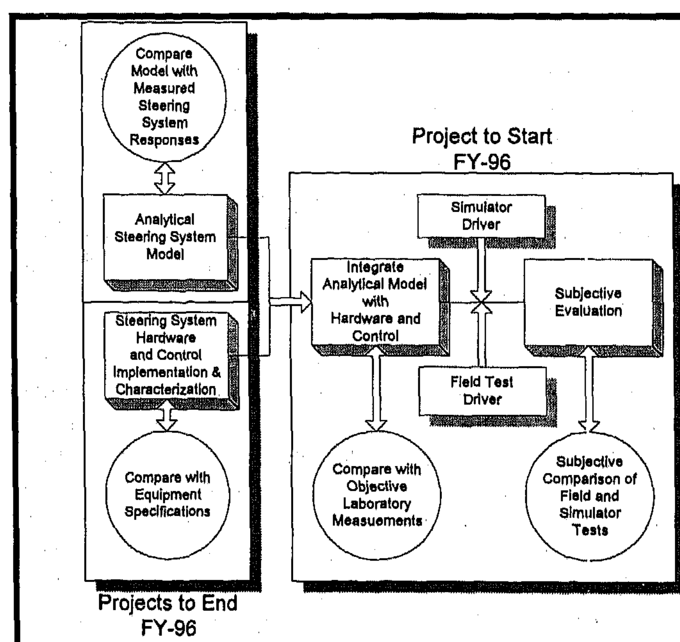


Figure 6. Schematic of Methodology to Develop Simulator Steering System Models/Hardware and Simulator Evaluation Tools

CONCLUSIONS

This project has shown that the current version of NADSdyna does a good job of predicting the performance of the 1994 Ford Taurus over a broad range of maneuvers. The results show promise that NADSdyna will provide the fidelity required for the NADS vehicle dynamics simulation. This validation effort is currently being extended to include acceleration maneuvers. Additionally, NADSdyna validation work is currently being performed for a Volvo-Fruehauf tractor-semitrailer. In work to date, a complete set of vehicle parameters has been measured for this vehicle and a large amount of experimental testing has been performed. The process of comparing NADSdyna predictions to this validation data for the Volvo-Fruehauf has just begun. This work will be reported upon in greater detail in future papers.

This project has also shown that the current versions of VDANL and VDM RoAD do good jobs of predicting the performance of the Taurus.

VRTC will continue to work to achieve the remaining goals of this project. The principle future activities will be:

1. Comprehensive testing of NADSdyna. NHTSA must be certain that all of the parts of this work before the code is placed in service with the NADS.
2. Development of two additional NADSdyna data sets. These will be for a sport utility vehicle and a passenger automobile. The precise make/models for which these data sets will be developed has not yet been decided.
3. Improvement of the functionality of NADSdyna beyond that originally developed by CCAD. Examples of planned improvements including the prediction of steering feel, adding failed vehicle subsystem models, adding interactions with roadway features (bumps and slippery patches), and adding the ability to model wind gusts.

REFERENCES

1.Chrstos, J. P. and Grygier, P. A., "Experimental Testing of a 1994 Ford Taurus for NADSdyna Validation," SAE Paper 970563, 1997.

2. Salaani, M. K, Chrstos, J. P. and Guenther, D. A., "Parameter Measurement and Development of a NADSdyna Validation Data Set for a 1994 Ford Taurus," SAE Paper 970564, 1997.

3. Salaani, M. K., Heydinger, G. J., and Guenther, D. A., "Validation Results from Using NADSdyna Vehicle Dynamics Simulation," SAE Paper 970565, 1997.

4. Chrstos, J. P. and Heydinger, G. J., "Evaluation of VDANL and VDM RoAD to Predict the Vehicle Dynamics for a 1994 Ford Taurus," SAE Paper 970566, 1997.

5. Howe, J. G., et al., "Improving Steering Feel for the National Advanced Driving Simulator," SAE Paper 970567, 1997.

6. "NADS Vehicle Dynamics Software - Release 4," Center for Computer Aided Design, The University of Iowa, August, 1995.

7. Haug, E. J., *Computer Aided Kinematics and Dynamics of Mechanical Systems - Volume I: Basic Methods*, Allyn and Bacon, Boston, 1989.

8. Haug, E. J., *Intermediate Dynamics*, Allyn and Bacon, Boston, 1989.

9. Bernard, J. E. and Clover, C. L., "Tire Modeling for Low-Speed and High-Speed Calculations," SAE Paper 950311, 1995.

10. Allen, R. W., et al., "Vehicle Dynamic Stability and Rollover," NHTSA Final Report DOT-HS-807-956, 1992.

11. Haug, E. J., et al., "Feasibility Study and Conceptual Design of a National Advanced Driving Simulator," NHTSA Final Report DOT-HS-807-597, 1990.

12. Allen, R. W., Rosenthal, T. J., and Szostak, H. T., "Analytical Modeling of Driver Response in Crash Avoidance Maneuvering - Volume II: An Interactive Tire Model," NHTSA Final Report DOT-HS-807-271, 1988.

13. Bernard, J. E. and Clover, C. L., "Validation of Computer Simulations of Vehicle Dynamics," SAE Paper 940231, 1994.

14. Allen, R. W., Rosenthal, T. J., and Szostak, H. T., "Analytical Modeling of Driver Response in Crash Avoidance Maneuvering - Volume III: A Trim Model and Computer Program for Determining Ground Vehicle Steady State Operating Conditions and Quasilinear Stability Coefficients," NHTSA Final Report DOT-HS-807-272, 1988.

15. Allen, R. W., Rosenthal, T. J., and Szostak, H. T., "Analytical Modeling of Driver Response in Crash Avoidance Maneuvering - Volume IV: User's Guide for Linear Analysis, Nonlinear Simulation, Part Task Simulation," NHTSA Final Report DOT-HS-807-273, 1988.

16. FHWA/UMTRI Contract DRFH61-93-R-00142.

17. Heydinger, G. J., "Improved Simulation and Validation of Road Vehicle Handling Dynamics," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1990.

18. Heydinger, G. J., et al., "A Methodology for Validating Vehicle Dynamics Simulations," SAE Paper 900128, 1990.

19. Bundorf, R. T., "The Influence of Vehicle Design Parameters on Characteristic Speed and Understeer," SAE Paper 670078, 1967.

20. Chrstos, J. P. and Grygier, P. A., "Instrumentation and Field Testing of 1994 Ford Taurus GL for NADSdyna Evaluation," NHTSA Final Report (draft), 1996.

21. Bernard, J., Bhatnagar, A., and Vanderploeg, M., "Evaluation of Select Vehicle Dynamics Models - Phase I Report," Motor Vehicle Manufacturers Association, Contract 9017-C0167B, 1990.

22. Bernard, J. E., Bhatnagar, A., and Clover, C. L., "Evaluation of Select Vehicle Dynamics Models - Phase II Final Report," Motor Vehicle Manufacturers Association, Contract 9114-C11302, 1992.

23. Salaani, M. K., "Development and Validation of a Vehicle Model for the National Advanced Driving Simulator," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1996.

24. Chrstos, J. P., "Parameter Measurement and Computation for 1994 Ford Taurus GL," NHTSA Final Report (draft), 1996.

25. Chrstos, J. P., "Simulation Evaluation of VDANL and VDM RoAD," NHTSA Final Report (draft), 1996.

26. Rupp, M. Y., "Passive Dynamic Steering System Model for Use in Vehicle Dynamics Simulation," M.S. Thesis, The Ohio State University, Columbus, Ohio, 1994.

27. Jang, B. C., "A Mathematical Model of a Power Steering System for Implementation in a Driving Simulator," M.S. Thesis, The Ohio State University, Columbus, Ohio, 1996.

28. Woodburn, C. M., "Modeling the Steering Torque Feedback Controller of a Fixed-Based Driving Simulator," M.S. Thesis, The Ohio State University, Columbus, Ohio, 1995.

29. Lee, S. W., Development of New Dynamic Tire Model For Improved Vehicle Dynamics Simulation, Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1994.

30. Salaani, M. K., "Vehicle Dynamics Modeling for the National Advanced Driving Simulator of a 1994 Ford Taurus Gl," DOT-NHTSA Report in Progress

31. Salaani, M. K., Heydinger, G. J. "Vehicle Dynamics Evaluation for the National Advanced Driving Simulator of a 1994 Ford Taurus Gl," DOT-NHTSA Report in Progress

32. Chrstos, J. P. and Heydinger, G. J., "Evaluation of NHTSA Light Vehicle Handling Simulations", NHTSA Final Report No. DOT-HS-807-868, 1992.

33. Sayers, M. W. and Mink, C., "A Simulation Graphical User Interface for Vehicle Dynamics Models", SAE Paper No. 950169, 1995.

34. AUTOSIM - Reference Manual, Mitchell and Gauthier Associates, Version 2.0, 1993.

35. Tsai, F. and Haug, E. J, "Automated Methods For High Speed Simulation of Multi-body Dynamic Systems," Technical Report R-47, Center for Computer-Aided Design, The University of Iowa, June 1989.

36. Segal, D. J., "Highway-Vehicle-Object Simulation Model - 1976, Engineering Manual - Analysis," Calspan Corporation, Zr-5461-V-8, FHWA Contract No. DOT-FH-11-8265, 1976.

37. Gim, G. and Nikravesh, P. E., " An Analytical Model of Pneumatic Tyres for Vehicle Dynamic Simulations, Part I, Pure Slips," International Journal of Vehicle Design, Vol. II, No. 6, 1990.

38. Gruening, J. W. and Bernard, J. E., "Verification of Vehicle Parameters for Use in Computer Simulation," SAE Paper 960176.

39. Garrott, W. R., Monk, M. W., and Chrstos, J. P., "Vertical Inertial parameters- Measured Values and Approximations," SAE 881767, February 1988.

40. Allen, R. W., Rosenthal, T. J and Chrstos, J. P., "A vehicle Dynamics Tire Model for Both Pavement and Off-Road Conditions," SAE Paper 970559.

41. Heitzman, E., "Topsy-A Modular Chassis Parameter Measurement System," SAE Paper 971056.

970563

Experimental Testing of a 1994 Ford Taurus for NADSdyna Validation

Jeffrey P. Chrstos
JPC Engineering

Paul A. Grygier
National Highway Traffic Safety Administration

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

As part of the National Advanced Driving Simulator (NADS) program, the Vehicle Research and Test Center (VRTC) in East Liberty, Ohio is evaluating the NADS vehicle dynamics software. As part of VRTC's effort, an extensive vehicle testing program to provide data for the simulation evaluation was performed. This paper describes VRTC's testing of a 1994 Ford Taurus GL passenger car. Each of the test maneuvers run by the Taurus are described, along with instrumentation setup, control actuation, test conditions, and driver procedures. The test data reduction and processing are detailed. Sample results of the testing and an analysis of test repeatablility and measurement noise are also presented.

INTRODUCTION

Starting in 1989, the National Highway Traffic Safety Administration (NHTSA) began researching the use and construction of a new state-of-the-art driving simulator, the National Advanced Driving Simulator (NADS) [1]. A site selection competition among United States universities was conducted, and in 1992, the University of Iowa was selected. As part of the University of Iowa's cost sharing, they contributed the core vehicle dynamics software for the NADS. This software program, called the Real Time Recursive Dynamics (RTRD) [2], is a minimum coordinate set, multi-body dynamics program that breaks the vehicle chassis and suspension system into a tree topology by cutting closed kinematic loops. This tree topology allows the multi-body system to be solved on parallel computers in real time.

After winning the NADS site selection competition, the University of Iowa supplemented the core vehicle dynamics (RTRD) with a collection of sub-modules that model portions of the vehicle such as the powertrain, tires, brakes, steering system, and aerodynamics that cannot be modeled using multi-body dynamics. The RTRD program containing these sub-modules is called NADSdyna [3].

NHTSA's Vehicle Research and Test Center (VRTC) in East Liberty, Ohio, is evaluating the NADSdyna vehicle dynamics software, developing vehicle parameter sets for NADSdyna, and enhancing various aspects of NADSdyna. As part of VRTC's effort, an extensive vehicle testing program to provide "benchmark" data for the simulation evaluation was performed. This paper describes VRTC's testing of a 1994 Ford Taurus GL passenger car. A similar testing program for an articulated tractor-semitrailer will be described in a later report from VRTC.

NADS is intended to be the most advanced driving simulator in the world when it is operational in the late 1990's. It is designed to be a research facility that will allow crash avoidance maneuvers and other high severity or dangerous driving situations to be safely studied. The broad scope of the NADS intended uses requires that the vehicle simulation be able to simulate vehicle ride, cornering, driving acceleration, and braking with a high degree of realism. VRTC's task of evaluating this simulation, therefore, requires a broad testing program covering both mild and severe vehicle ride, cornering, driving acceleration, and braking. The simulation evaluation methodology used is based on past simulation programs at VRTC [4, 5].

VRTC performed this testing program to provide the data needed to evaluate the NADSdyna software. However, there is very little publicly available instrumented vehicle handling and braking test data. One of the purposes of this paper therefore is to make this test data available to other researchers. Electronic versions of the test data described in this paper can be obtained be contacting W. Riley Garrott at VRTC, P. O. Box 37, East Liberty, Ohio, 43319, or via telephone at (937) 666-4511.

The test vehicle used for this project is a 1994 Ford Taurus GL. The Taurus VIN number is 1FALP5249RA164396. The Taurus has a 3.8-liter V6 engine, four-speed automatic transmission, speed sensitive power rack and pinion steering,

and four wheel anti-lock brakes. General Ameri-tech 205/65R15 tires with a cold inflation pressure of 35 psi are used for all testing. The tires were purchased together, and are from the same production batch. Force and moment test data for these tires (from the same batch) can be found in reference [6].

INSTRUMENTATION AND DATA RECORDING

This section briefly describes the transducers and data system used for the test program. Table 1 lists the data channels, the types of sensors used, and the units in which the data is stored. A more detailed description of the experimental setup and data storage system can be found in Reference [7].

DATA ORGANIZATION - Because of the size and scope of this vehicle testing project (more than 40 transducers, 75 maneuver types and conditions, and nearly 600 individual test runs), data organization is very critical. Therefore, guidelines were written for data file naming conventions, data file storage locations, transducer mnemonic naming conventions, and data units.

All data acquired for this project is stored in SI units. To avoid confusion, the decision was made to "build" all units out of the basic SI units. For example, the rotational speed of the engine is given in rad/s rather than in RPM. The axis system used is the SAE standard (from J670e) with X forward, Y to the driver's right, and Z down. The data is stored in engineering units in VRTC's DX file format. This file format allows test information (in ASCII) to be stored with the data (in floating-point numbers). For the purposes of data analysis, the test files are converted from the DX format into MATLAB binary format.

Because of the size of the testing program, record keeping is a major concern. For example, when simulating a particular test, the measured vehicle weight from the day of the test is necessary (during the course of this testing program, the total vehicle mass varied more than 70 kilograms, which is approximately 4 percent of the total mass). Keywords are a feature of VRTC's DX file format that allow information to be stored in the file header, and retrieved by using an associated keyword. This feature is used to store daily test information directly with the data. Some of this information is used by the data analysis programs, while some is simply for archival purposes. While Keywords are a feature of the DX file format, Keywords containing numerical information are also included in the raw data MATLAB files.

Data Acquisition Equipment - All on-board measurements are made using an Optim Megadac 2210C (Megadac) data recorder [8]. The Megadac can digitize up to 128 analog channels with a maximum overall sample rate of 20,000 samples per second. All analog channels are sampled using a 16-bit analog-to-digital converter. The Megadac also has four 16-bit digital counters that may be cascaded. All recorded data is stored on a 60-megabyte, quarter-inch tape cartridge.

At VRTC, the Megadac is controlled using a Grid model Gridcase 1500-series laptop computer. The laptop is connected to the Megadac using a IEEE-488 port. The laptop runs custom software written by VRTC that allows the system to be set up for each specific test (i.e. what channels to record, trigger mode, number and duration of test runs, etc.). The laptop software also keeps data specific to each test that is used when data is transferred from the data cartridge and converted to DX

Table 1. Data Channels

Data Channel	Sensor (units)
Vehicle accelerations in X, Y, and Z-axes	Accelerometers (m/s^2)
Vehicle angular rates about X, Y, and Z-axes	Rate sensors (rad/s)
Power steering rack pressures	Pressure transducers (kPa)
Optical event	Optical trigger (V)
Chassis height	Ultrasonic distance sensors (m)
Brake line pressures	Pressure transducers (kPa)
Tie rod forces	Strain gages (N)
Vehicle speed	Tachometer generator (m/s)
Distance traveled	Fifth wheel (m)
Brake pedal force	Force transducer (N)
Transmission gearshift selector	Engine control module (V)
Engine speed	Engine control module (rad/s)
Gear transmission is in	Engine control module (V)
Wheel rotational velocities	ABS wheel sensors (rad/s)
Handwheel angle	Potentiometer (rad)
Handwheel rotational velocity	Tachometer generator (rad/s)
Handwheel torque	Strain gages (N-m)
Throttle position	Engine control module (%)
Strut accelerations along strut axis	Accelerometers (m/s^2)
Strut lengths	Potentiometers (m)

files. The test driver's primary interaction with the data system is through the laptop interface.

STEER, BRAKE, AND THROTTLE CONTROLS - For the majority of the maneuvers run for this program, open-loop control inputs are desired. To make the control inputs as "open-loop as possible", three devices are used and are described below. Steering wheel angle inputs are constrained by a mechanical steering stop, a pneumatic ram is used to actuate the brake pedal, and a mechanical throttle pedal actuator is used during drivetrain testing.

The steering stop is based on a design from General Motors, and can be used to constrain the motion of the steering wheel and measure steering wheel angle and torque. In this program, it is only used to constrain the steering wheel motion. The steering stop hardware, shown in Figure 1, replaces the existing steering wheel. A torsional load cell is installed between the existing steering wheel mount and the steering stop's steering wheel. To the steering stop's steering wheel is attached a 20-centimeter diameter "Degree" wheel with tapped holes every 10 degrees that allow two adjustable steering stops to be installed. An adjustable length bar is mounted between the two A-pillars, just above the dashboard to provide a rigid base. To this bar is mounted an aluminum plate that has an engagement lever that the adjustable steering stops contact to stop steering wheel rotation. During this testing program, the repeatability of the steering wheel angle using the steering stop is approximately ±1 degree.

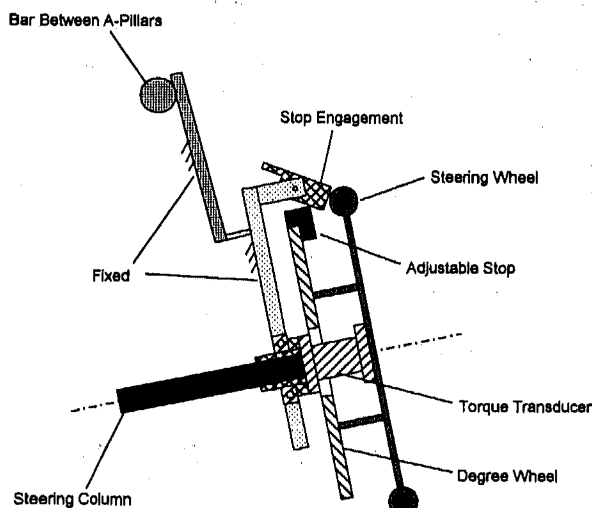


Figure 1. Diagram of steer stop

Repeatable brake pedal force actuation is achieved using a pneumatic ram. The ram, shown in Figure 2, is bolted to the vehicle floor, and oriented to act approximately perpendicular to the brake pedal force transducer. Two compressed air tanks are installed in the trunk for use with the ram. A pressure regulator is used to adjust brake ram force. The test driver uses a U-tube manometer to measure vehicle deceleration during a few test runs, and adjusts the regulated air pressure to the ram to achieve the desired deceleration. An electrically actuated solenoid valve is used to release the air to the ram from either a driver or Megadac command. During this testing program, the repeatability of the brake pedal force using the ram is approximately ±6 newtons.

The third control actuator that is used in this program is a throttle pedal actuator, shown in Figure 3. This actuator is simply an adjustable length strut that is hinged in the middle. The hinge is constrained to move in only one direction, and when the driver pushes down on the handle, the strut locks in its straight position (angle between the legs is 180 degrees). One end of the strut is hinged to the vehicle floor pan, while the other end of the strut engages the throttle pedal. During this testing program, the repeatability of the throttle position using the actuator is approximately ±1 percent of full throttle.

PRE-TEST CALIBRATION RUNS - Each time the Taurus is taken out to the test track, a pre-test calibration procedure is run. The pre-test calibration procedure begins with the car being driven onto the VRTC vehicle weigh station (four low profile Winslow force plates, mounted on the shop floor, that are used by VRTC to measure the four wheel loads simultaneously). The vehicle weight and tire tread depth is noted at each wheel. The Megadac is autobalanced and these values are saved by the Megadac, to be reloaded after the vehicle reaches the test area. Once the vehicle reaches the skid pad, a standing calibration is performed. This is a record of what the transducers are reading with the engine running, but no other inputs. After the standing calibration, the autobalanced values are restored. The next test is the 1000-ft (304.8-m) calibration run. It is designed to check the accuracy of the fifth wheel. The vehicle is driven the length of a 1000-ft course with the data collection system active. Data collected during this run is post-

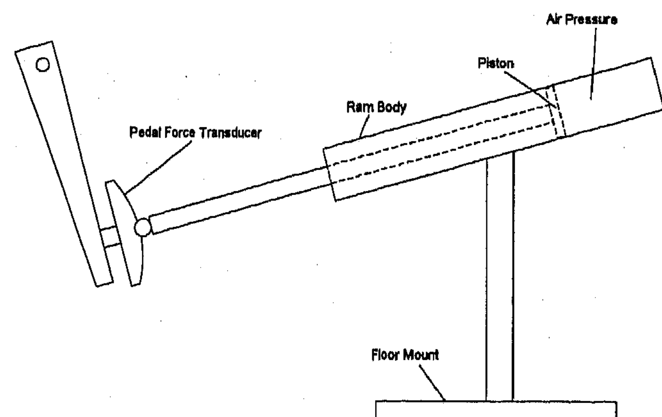


Figure 2. Pneumatic brake ram

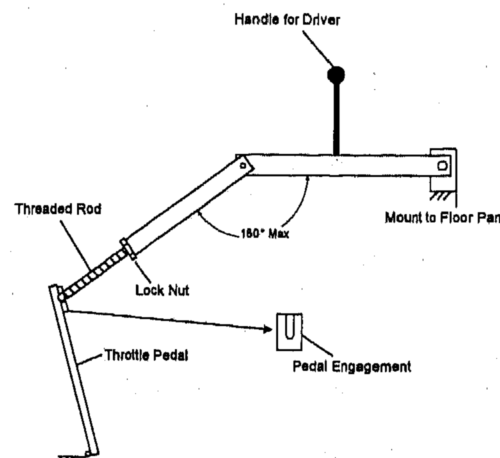


Figure 3. Throttle position actuator

processed to determine the average front ride height and the handwheel position offset can be found during post-test processing. The last calibration procedure is the 18 to 9 meters/second (40 to 20 miles per hour) run (the vehicle is driven in a straight line, and the speed is held constant at 18 and 9 meters/second for approximately 5 seconds each). This test is designed to check the accuracy of the vehicle speed sensor at two test speeds. Data from this run are used during post-test processing to determine the four tire rolling radii.

TEST MANEUVERS

This section describes the test maneuvers run for this project. Each basic maneuver type will be described in detail, and the actual test cases presented in tabular form. All testing was performed at the Transportation Research Center, in East Liberty, Ohio. Detailed descriptions of the TRC facilities used can be found in reference [9]. Basic descriptions of the facilities used are:

•**Vehicle Dynamics Area (VDA):** a 550 by 450-meter flat asphalt test surface with loops on two sides. The VDA is used for all handling type testing.

•**Skid Pad:** an 8-lane, 1200-meter long concrete test area with loops at each end. The skid pad is used for straight-line maneuvers such as driveline and brake testing.

•**Brake Slope:** 12-percent and 15-percent slope concrete test surfaces built to test vehicle parking brakes.

For many of the maneuvers, approximately 2 seconds of pre-trigger data are collected. This data is used to characterize the pre-maneuver state of the test vehicle. The pre-trigger data collection is a feature of the Megadac. After the Megadac is armed, data is collected and stored in a user definable size circular memory buffer. When the Megadac is triggered, the content of the buffer is written to the tape followed by the post event data.

For maneuvers where repeatable open-loop inputs can be achieved, each test is repeated up to 10 times. The data from these repeat tests is then ensemble averaged in the time domain to obtain statistical measures of test repeatability. The time-domain ensemble averaging is described later in this paper.

Each maneuver is designated by a two-character maneuver code. The first letter of the code shows the major maneuver type (e.g. "A" indicates a straight-line acceleration test). The second character designates a unique set of driver inputs (i.e. steering, throttle, and braking) for the test. This second character is referred to as the minor maneuver code and the major maneuver determines its meaning.

STRAIGHT-LINE ACCELERATION MANEUVERS - Straight-line acceleration maneuvers (maneuver code A) are run to provide data for characterizing the test vehicle's drivetrain performance. Table 2 shows details of the 10 test conditions run. All of the tests are run on TRC's skid pad. The throttle position is controlled using the mechanical throttle actuator described earlier. The test driver controls the steering wheel and attempts to keep the vehicle traveling in a straight line using the lane markings on the skid pad as an aid.

Minor maneuver codes A, B, and C are designed to test the overall drivetrain characteristics. For these runs, the gear selector is put in the drive ("D") position, and the transmission allowed to shift in its normal mode. Each run is started with the running vehicle stopped with no throttle input, in its test lane, with the driver's foot on the brake pedal. The data acquisition system is set to trigger on approximately 5-percent throttle input. When the data acquisition system is armed, the driver removes his foot from the brake pedal, and simultaneously pushes the throttle actuator into position. Minor maneuver codes A, B, and C correspond to fixed throttle inputs of approximately 10, 30, and 75 percent. Data recording for each test is continued for either 50 seconds or until the vehicle speed reaches 33 meters/second.0

Minor maneuver codes D, E, F, G, H, and I are designed to test the engine characteristics. For these runs, the gear selector is put in the first gear ("1") position so the transmission is not allowed to shift. Each run is started with the running vehicle stopped with no throttle input, in its test lane, with the driver's foot on the brake pedal. The data acquisition system is set to trigger on approximately 5-percent throttle input. When the data acquisition system is armed, the driver removes his foot from the brake pedal, and simultaneously pushes the throttle actuator into position. Run codes D, E, F, G, H, and I correspond to fixed throttle inputs of approximately 10, 20, 30, 50, 75, and 100 percent. Data recording for each test is continued until the vehicle speed reaches 15 meters/second, which corresponds to the maximum safe engine speed in first gear.

The next straight-line acceleration run is minor maneuver code J, which is designed to test the transmission down-shifting characteristics. For these runs, the gear selector is put in the drive ("D") position, and the transmission is allowed to shift in its normal mode. For this test, the vehicle is driven in a straight line at a constant 22 meters/second, utilizing the cruise control. The data system is set to trigger at 60-percent throttle, and the throttle actuator is set for 85-percent throttle. With the vehicle traveling at a constant 22 meters/second and the data system armed, the driver pushes the throttle actuator into position. This causes the transmission to downshift from fourth to third gear. The data system is set to keep approximately 2 seconds of pre-trigger data, and to record for approximately 8 seconds after the throttle application.

Table 2. Straight-Line Acceleration Maneuvers

Code	# Runs	Vx (m/s)	% Throttle	Gear Pos.
AA	9	0 -> 33	10	D
AB	9	0 -> 33	30	D
AC	11	0 -> 33	75	D
AD	5	0 -> 15	10	1
AE	5	0 -> 15	20	1
AF	5	0 -> 15	30	1
AG	5	0 -> 15	50	1
AH	5	0 -> 15	75	1
AI	5	0 -> 15	100	1
AJ	4	22 -> 33	85	D
AK	11	0 -> 33	10 -> 100	D

The final set of straight-line acceleration runs are found in maneuver code K, which are used to determine the shifting characteristics at various throttle positions. For these runs, the gear selector is put in the drive ("D") position, and the transmission is allowed to shift in its normal mode. Each run is started with the running vehicle stopped with no throttle input, in its test lane, with the driver's foot on the brake pedal. The data acquisition system is set to trigger on approximately 5-percent throttle input. When the data acquisition system is armed, the driver removes his foot from the brake pedal, and simultaneously pushes the throttle actuator into position. One run was performed at each throttle position, starting at 10 percent and ending at 100 percent with a 10 percent increase in throttle between each run. Data recording for each test continued until the vehicle reached either 33 meters/second or a constant speed below 33 meters/second.

BRAKE-IN-A-TURN MANEUVERS - Brake-in-a-turn maneuvers (maneuver code B) are run to provide data for characterizing the interaction of the lateral and longitudinal handling modes of the test vehicle. Table 3 shows the test specifications for the six brake-in-a-turn maneuvers, showing two test speeds, 22 and 33 meters/second, each with a low severity and a high severity brake-in-a-turn maneuver. The first four maneuvers were performed with the anti-lock brake system active; the last two were performed with it turned off. All of these maneuvers are run on the VDA.

Table 3. Brake-in-a-Turn Maneuvers

Code	# Runs	Vx (m/s)	Ax (m/s^2)	Ay (m/s^2)
BA	7	22.4	-2.5	-2.2
BB	10	22.4	-4.8	-6.5
BC	8	33.4	-3.7	2.1
BD	11	33.4	-4.7	5.7
BE	11	22.1	-4.0	-4.9
BF	12	33.3	-4.3	6.1

The steering wheel stop is installed, and the stops set at straight ahead and at the desired steering wheel angle for the test. The data system is set to trigger off the steering wheel angle, and is set up to save approximately 2 seconds of pre-trigger data. The data system is used to fire the pneumatic brake pedal ram when the steering wheel trigger level is reached. This means that except for propagation delays (approximately 0.1 seconds), the steering and braking inputs are applied at the same time. Data is recorded until the vehicle comes to a stop.

The brake-in-a-turn maneuver starts with the brake ram air pressure and steering stop position set to give the desired decelerations. The test driver uses the cruise control to attain the initial test speed and the data system is armed. The steering wheel is initially held against the zero (straight ahead) stop. The driver triggers the test by rapidly turning the steering wheel against the other stop and holding it there until the test is complete. As the test triggers, the brake ram fires and the vehicle slows to a stop. Each repeat test is run in approximately the same location on the VDA. Because of heat buildup in the brakes, especially with the 33 meters/second runs, the vehicle is driven for a few minutes between test runs to allow the brakes to cool.

CONSTANT SPEED J-TURN MANEUVERS - Constant speed J-turn maneuvers (maneuver code C) are run to provide data for characterizing the lateral handling mode of the test vehicle. Table 4 shows the test specifications for the twelve J-turn maneuvers, run at three test speeds of 11, 22 and 33 meters/second, each with four severity levels ranging from lateral accelerations of 2 meters/second^2 to 8 meters/second^2. All of these maneuvers are run on the VDA.

Table 4. Constant Speed J-Turn Maneuvers

Code	# Runs	Vx (m/s)	Ay (m/s^2)
CA	10	11.2	2.0
CB	9	11.3	-4.0
CC	10	11.5	6.0
CD	6	11.3	-8.0
CN	8	22.7	-2.0
CP	10	22.4	4.0
CQ	9	22.6	-6.0
CR	7	22.0	6.5
CV	8	33.0	2.0
CW	6	33.5	-4.0
CX	10	33.2	6.0
CY	5	33.2	-7.0

The steering wheel stop is installed, and the stops set at straight ahead and at the desired steering wheel angle for the test. The data system is set to trigger off the steering wheel angle. Approximately 2 seconds of pre-trigger data are saved, and data is recorded for approximately 5 seconds after the steering input to allow the vehicle to reach steady-state.

The J-turn maneuver starts with the steering stop position set to give the desired lateral acceleration. The test driver uses the cruise control to attain the initial test speed and arms the data system. The steering wheel is initially held against the zero (straight ahead) stop. The driver rapidly turns the steering wheel (which triggers data collection) until it hits the other stop and holds it there until the test is complete. Each repeat test is run in approximately the same location on the VDA.

SPECIAL MANEUVERS - Several special maneuvers were performed. Two of these maneuvers provide additional information about the steering system and the third about low-speed vehicle trajectory. The three maneuvers are the flick test, the J-turn and release test, and turning at an intersection.

The flick maneuvers (maneuver code F) are run to test the vehicle's steering stability. Table 5 shows the test specifications for the four maneuvers, run at test speeds of 11 and 22 meters/second. Lateral acceleration levels ranged from 3.6 meters/second^2 to 8.8 meters/second^2.

Table 5. Special Maneuvers

Code	# Runs	Vx (m/s)	Ay (m/s²)
FA	10	11	3.6
FB	7	11	-6.6
FC	7	22	-5.0
FD	8	22	8.8
GA	13	11	2.0
GB	13	11	-4.0
GC	12	11	6.0
GN	12	22	-2.0
GP	13	22	4.0
GQ	12	22	-6.0
TA	1	0 to 4	0 to -1.5
TB	1	0 to 4	0 to 1.4

The steering wheel stop is installed, and the stops set at straight ahead and at steering wheel angles of 90 degrees for maneuvers FA and FC and 180 degrees for maneuvers FB and FD. The data system is set to trigger off the steering wheel angle. Approximately 2 seconds of pre-trigger data are saved, and data is recorded for approximately 3 seconds after the steering input to allow the vehicle to reach steady-state.

The flick maneuver starts with the steering stop position set to give the desired steering input. The test driver uses the cruise control to attain the initial test speed and arms the data system. The steering wheel is initially held against the zero (straight ahead) stop. The driver rapidly turns the steering wheel (which triggers data collection) until it lightly contacts the other stop and then immediately releases the wheel. The vehicle is then permitted to resume its original path on its own. Each repeat test is run in approximately the same location on the VDA.

Constant speed J-turn and release maneuvers (maneuver code G) are run to provide data for characterizing the steering system of the test vehicle. Table 5 shows the test specifications for the six J-turn and release maneuvers, run at two test speeds of 11 and 22 meters/second, each with three severity levels ranging from lateral accelerations of 2 meters/second^2 to 6 meters/second^2. All of these maneuvers are run on the VDA.

The steering wheel stop is installed, and the stops set at straight ahead and at the desired steering wheel angle for the test. The data system is set to trigger off the steering wheel angle. Approximately 2 seconds of pre-trigger data are saved, and data is recorded for approximately 5 seconds after the steering input to allow the vehicle to reach steady-state.

The J-turn and release maneuver starts with the steering stop position set to give the desired lateral acceleration. The test driver uses the cruise control to attain the initial test speed and arms the data system. The steering wheel is initially held against the zero (straight ahead) stop. The driver rapidly turns the steering wheel (which triggers data collection) against the other stop and holds it there for 3 seconds. He then takes his hands off the steering wheel and allows the vehicle to return to a straight course. Each repeat test is run in approximately the same location on the VDA. These repeat tests are not used by the ensemble averaging routines (described later), but are run to better characterize the steering system.

Turning-at-an-intersectionmaneuvers(maneuver code T) are run to provide data for validating vehicle trajectory calculations. Table 5 shows the test specifications for the two turns, one left and the other right, run at low speed. These maneuvers are run on an area of asphalt delineated by two perpendicular lines and each turn is performed once.

The vehicle is positioned at the intersection of two perpendicular lines. Data collection is started and stopped by the driver. No pre-trigger data is collected. In addition to the on-board instrumentation, tire path data is collected. The tires are chalked and as the turn is performed, the chalk from the tires is transferred to the asphalt surface. The x and y-coordinates for multiple points on the path are determined by measuring the distances from the two perpendicular reference lines to the center of the chalk path. The vehicle paths for the left and right turns are shown in Figures 4 and 5, respectively. The lateral acceleration was approximately 1.5 meters/second^2.

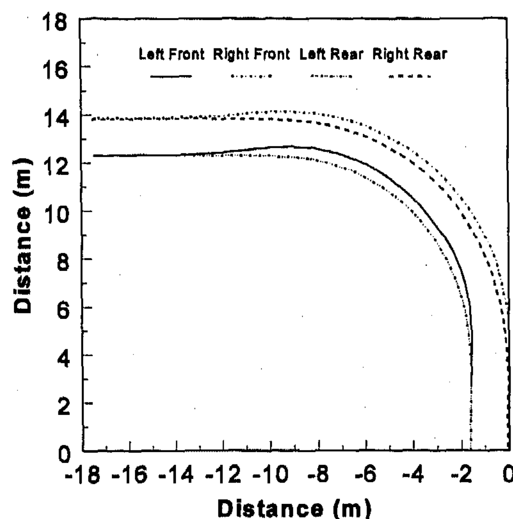


Figure 4. Left turn at an intersection (maneuver code TA)

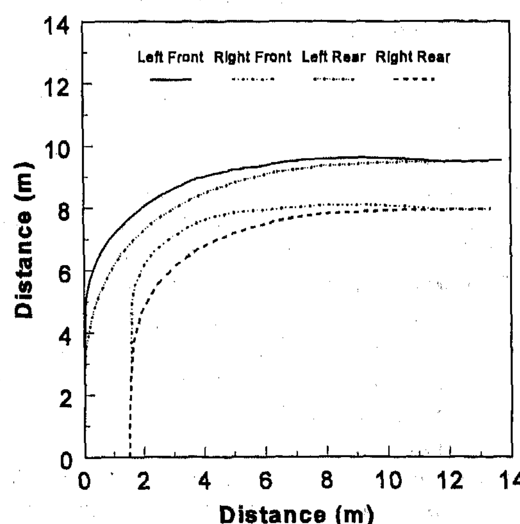


Figure 5. Right turn at an intersection (maneuver code TB)

In Figures 4 and 5, the front axle track appears to be narrower than the back axle track. This is because the front wheels are turned and the front axle is not perpendicular to the vehicle path but is skewed. This effect is more noticeable in the right turn (Figure 5) because that turn is sharper than the left.

Pulse Steer Maneuvers - Constant speed pulse steer maneuvers (maneuver code P) are run to provide data for characterizing the lateral handling mode of the test vehicle in the frequency domain. Table 6 shows the test specifications for the six pulse steer maneuvers (minor maneuver codes B through F) run at three test speeds of 11, 22 and 33 meters/sec, each with positive and negative steering pulses having magnitudes sufficient to achieve lateral accelerations of approximately 3 meters/second^2 (minor maneuver code A was run to debug the test procedure, and is not at one of the standard test speeds). All of these maneuvers are run on the VDA.

Table 6. Pulse Steer Maneuvers

Code	# Runs	Vx (m/s)	Steer Amplitude (rad)
PA	9	20.2	0.75
PB	6	11.4	1.2
PC	7	11.5	-1.1
PD	8	22.8	0.57
PE	10	22.9	-0.55
PF	8	33.7	0.35
PG	9	33.5	-0.38

The steering wheel stop is installed, and the stops set at straight ahead and at the desired maximum steering wheel angle for the test. The data system is set to trigger off the steering wheel angle. Approximately 2 seconds of pre-trigger data are saved, and data is recorded for approximately 5 seconds after the steering input to allow the vehicle to reach steady-state.

The pulse steer maneuver starts with the steering stop position set to achieve a maximum lateral acceleration of approximately 3 meters/second^2. The test driver uses the cruise control to attain the initial test speed, and arms the data system. The steering wheel is initially held against the zero (straight ahead) stop. The driver triggers the test by rapidly turning the steering wheel to the other stop, then immediately returning it to the zero stop and holding it there until the test is complete. Typical pulse duration is less than 0.2 seconds, providing good steering input power out to 3.5 to 4 hertz. Each repeat test is run in approximately the same location on the VDA.

DOUBLE LANE-CHANGE MANEUVERS - Double lane-change maneuvers (maneuver code LA, only one minor maneuver codes is used for this type of test) are run to provide data for characterizing the lateral handling mode of the test vehicle. Table 7 shows the test specifications for the nine double lane-change maneuvers run at test speeds ranging from 11 to 20 meters/second, with peak lateral accelerations ranging from 4 meters/second^2 to 8.5 meters/second^2. All of these maneuvers are run on the VDA through the same lane change course. Only the vehicle speed is varied. The course geometry is laid out using cones, with lane widths are 3.65 meters. The length of the gates are 18.3 meters, and the distance between the two gates is 22.9 meters.

Table 7. Double Lane-Change Maneuvers

Code	Vx (m/s)	Max. Ay (m/s^2)
LA00	11.0	-4.6
LA01	12.5	4.8
LA02	15	-6.0
LA03	11	-4.2
LA04	13	5.0
LA05	15.5	-5.0
LA06	16	-7.0
LA07	18	7.5
LA08	20	8.5

The data system is set to trigger off the steering wheel angle. Approximately 2 seconds of pre-trigger data are saved, and data is recorded until the driver hits the "Esc" key on the laptop. The driver is instructed to use the cruise control to set the desired test speed, arm the data system, and get the vehicle aligned with the entrance lane. The steering stop is not used for this test, so the steering control input represents the driver/vehicle closed-loop course-following performance. The driver's task is to keep the vehicle within the course boundaries marked by the cones. The driver is also instructed to steer as "smoothly" as possible. Unlike most of the maneuver types, repeat test are not run for the double lane-change maneuver. This is because the driver steering input is closed-loop, and is inherently non-repeatable.

Straight-Line Braking Maneuvers - Straight-line braking maneuvers (maneuver code S) are run to provide data for characterizing the braking characteristics of the test vehicle. Table 8 shows the test specifications for five straight-line braking maneuvers (minor maneuver codes B through F) with an initial test speed of 22 meters/second, with longitudinal accelerations ranging from -2 to -9 meters/second^2. All of these maneuvers are run on the skid pad.

Table 8. Straight-line Braking Maneuvers

Code	# Runs	Vx (m/s)	Ax (m/s^2)
SB	10	22	-2
SC	8	22	-3
SD	9	22	-5
SE	9	22	-7.5
SF	12	22	Max
SG	10		Brake on 12% slope
SH	10		Brake on 15% slope

The steering wheel stop is not installed, and the driver is instructed to keep the vehicle traveling in a straight line. The data system is set to trigger off the brake pedal force, and is setup to save approximately 2 seconds of pre-trigger data. The driver uses a switch to fire the pneumatic brake pedal ram. Data is recorded until the vehicle comes to a stop.

The maneuver starts with the brake ram air pressure set to give the desired deceleration. The test driver uses the cruise control to attain the initial test speed. With the data system armed, the driver pushes the switch to fire the ram. Each repeat test is run in approximately the same location on the skid pad. Because of heat build up in the brakes, the vehicle is driven for a few minutes between test runs to allow the brakes to cool.

Minor maneuver codes G and H are "stop on a hill" tests, and are used to provide data to evaluate the simulation's tire model at low speeds and when the vehicle is stopped (standard tire models used in handling simulations can not simulate tire force at low or zero speed, and special models must be developed for use in driving simulators). For these tests, the vehicle is driven onto the brake slope (either 12-percent or 15-percent slope) with an initial speed of approximately 11 meters/second (attained by the driver controlling the throttle, not using the cruise control). Using his foot to control the brake pedal, the driver brings the vehicle to a stop. The data system triggers off the brake pedal force, and saves approximately 2 seconds of pre-trigger data. The driver lets the data system record for approximately 5 seconds after the vehicle has stopped.

SLOWLY INCREASING STEER MANEUVERS - Constant speed, slowly increasing steer maneuvers (maneuver code XC, only one minor maneuver code is used for this type of test) are run to provide data for characterizing the lateral handling mode of the test vehicle. Table 9 shows the test specifications for the twenty slowly increasing steer maneuvers run at three test speeds of 11, 22 and 33 meters/second. Runs 0 to 5 have maximum lateral accelerations of approximately 4 meters/second^2, while runs 6 to 19 are run to the vehicle's maximum lateral acceleration (approximately 8 meters/second^2). All these maneuvers are run on the VDA.

The steering wheel stop is not installed for these tests, with the driver attempting to input a slow ramp input to the steering wheel. The data system is set to trigger off the steering wheel angle. No pre-trigger data is saved, and data is recorded until the driver manually stops the recording.

The maneuver starts with the test driver using the cruise control to attain the desired test speed. With the data system armed, and the vehicle traveling in a straight line, the driver starts the data recording. After a few seconds, the driver slowly starts to steer in the specified direction, trying to input a constant rate steering wheel angle ramp input. When the desired maximum lateral acceleration is reached, the driver stops the data recording, and brings the vehicle to a stop. At all test speeds, the maneuvers are run with both positive and negative steering inputs to characterize any vehicle asymmetries.

While this is a closed-loop steering maneuver, each vehicle speed/steering direction combination is repeated two or three times. These repeat tests are not used by the ensemble averaging routines, but are run to better characterize the vehicle steady-state limit performance.

COAST-DOWN MANEUVERS - Coast-down maneuvers (maneuver code XF, only one minor maneuver code is used for this type of test) are run to provide data for characterizing the aerodynamic and rolling drag of the test vehicle. Table 9 shows the test specifications for the 10 coast-down runs, all of which start at 33 meters/second. All of these runs are done on the skid pad.

The steering wheel stop is not installed for these tests, with the driver attempting to keep the vehicle traveling in a straight line with as little steering as possible. The data system is triggered manually by the test driver. No pre-trigger data is saved, and data is recorded until the vehicle speed drops below approximately 6.5 meters/second.

The maneuver starts with the test driver using the cruise control to attain the desired test speed. With the data system armed and the vehicle traveling in a straight line, the driver shifts the transmission into neutral, then starts the data system. The driver's task now is to keep the vehicle in its lane using as few steering corrections as possible.

ON-CENTER WEAVE MANEUVERS - On-center weave maneuvers (maneuver code XH, only one minor maneuver code is used for this type of test) are run to provide data for characterizing the on-center handling and steering system of the test vehicle. A detailed description of this test can be found in reference [10]. Table 9 shows the test specifications for the 25 on-center weave runs, which are run at speeds of 13, 22, and 33 meters/second. All these runs are done on the VDA.

The steering wheel stop is not installed for these tests, with the driver attempting to input a 0.2 hertz sinusoidal steering input with an amplitude to produce a maximum lateral acceleration of 2 meters/second^2. The data system is triggered manually by the test driver. No pre-trigger data is saved, and data is recorded until the driver manually stops the data system.

The maneuver starts with the test driver using the cruise control to attain the desired test speed. With the data system armed, and the vehicle traveling in a straight line, the driver starts the data system. The drivers task now is to input a 0.2 hertz sinusoidal steering input with an amplitude to produce a maximum lateral acceleration of 2 meters/second^2. The driver inputs one-and-a-half to two full cycles, then returns the steering wheel to zero (straight a head), and turns off the data system.

While this is a closed-loop steering maneuver, each test is repeated approximately six times. These repeat tests are not used by the ensemble averaging routines, but are run to better characterize the vehicle steering system.

On-Center Transition Maneuvers - On-center transition maneuvers (maneuver code XI, only one minor maneuver code is used for this type of test) are run to provide data for characterizing the on-center handling and steering system of the test vehicle. A detailed description of this test can be found in reference [10]. Table 9 shows the test specifications for the 49 on-center transition runs, which are run at speeds of 13, 22, 33 meters/second. At each test speed, runs with positive and negative steering inputs are run. All these runs are done on the VDA.

The steering wheel stop is not installed for these tests, with the driver attempting to gradually move the steering wheel from a "hands off" straight ahead condition to a low-g cornering condition. The data system is triggered manually by the test driver. No pre-trigger data is saved, and data is recorded until the driver manually stops the data system.

The maneuver starts with the test driver using the cruise control to attain the desired test speed. With the data system armed and the vehicle traveling in a straight line with his hands off the steering wheel, the driver starts the data system. The driver's task now is to lightly place his palms on the steering wheel, and slowly turn it either clockwise or counterclockwise.

The driver turns the wheel until the vehicle is cornering at 1 to 2 meters/second^2, then turns off the data system.

While this is a closed-loop steering maneuver, each test is repeated approximately six times in each steering direction. These repeat tests are not used by the ensemble averaging routines, but are run to better characterize the steering system.

COAST-TO-STOP ON SLOPE MANEUVERS - Coast-to-stop on slope maneuvers (maneuver code XP and XQ) are run to provide data for characterizing the low speed tire models used with driving simulators. Table 9 shows the test specifications for tests done on 12 and 15-percent slopes. All of these runs are done on the brake slope facility.

These tests are very similar to tests SG and SH, (described earlier) except rather than braking to a stop, the slope is used to stop the vehicle, before the brakes are applied. The data system is triggered manually by the test driver. No pre-trigger data is saved, and data is recorded until the driver manually stops the data system.

The maneuver starts with the test driver approaching the brake slope at approximately 10 meters/second. Before the beginning of the brake slope, the data system is manually triggered, and the throttle released. When the vehicle comes to a stop on the slope, the driver pushes the brake pedal to keep the vehicle from rolling back down the slope. After waiting a few seconds, the driver manually turns off the data system.

Table 9. Miscellaneous Maneuvers

Code	Description	# Runs	Vx (m/s)
XC00	Slowly Increasing Steer - negative	1	12
XC01	Slowly Increasing Steer - negative	1	22
XC02	Slowly Increasing Steer - negative	1	32
XC03	Slowly Increasing Steer - positive	1	11
XC04	Slowly Increasing Steer - positive	1	23
XC05	Slowly Increasing Steer - positive	1	33
XC06-07	Slowly Increasing Steer - negative	2	11
XC08-09	Slowly Increasing Steer - positive	2	11
XC10-11	Slowly Increasing Steer - negative	2	22
XC12-13	Slowly Increasing Steer - positive	2	22
XC14-15	Slowly Increasing Steer - negative	2	33
XC16-19	Slowly Increasing Steer - positive	4	33
XF	Coast-Down	10	33
XH01-7	On-Center: Weave	7	22
XH08-12	On-Center: Weave	5	13
XH13-18	On-Center: Weave	6	22
XH19-25	On-Center: Weave	7	33
XI00-11	On-Center: Transition	12	22
XI18-27	On-Center: Transition	10	13
XI28-39	On-Center: Transition	12	22
XI40-54	On-Center: Transition	15	33
XN	Brake Pressure/Pedal Hysteresis	2	0
XP	Coast-to-stop on 12% slope	10	
XQ	Coast-to-stop on 15% slope	9	

DATA REDUCTION

Up to now, this paper has described the acquisition and storage of the "raw" vehicle data. This section will describe the processing of the raw data into a form suitable for simulation evaluation. The process may involve (depending on the test) removing offsets, digital filtering, computing derived channels from measured channels, ensemble averaging in the time domain, and/or conversion of the time-domain data into the frequency domain. All data processing is done using the software program MATLAB, version 4.2c[MATLAB® - Reference Guide, The Mathworks, Inc., July 1993.1511]. Some of the analysis also uses the Signal Processing Toolbox version 3.0b (the Signal Processing Toolbox is an add-on to MATLAB). All VRTC written MATLAB .M files (MATLAB's scripting language) used to process the data are available electronically from VRTC. All of the processing software is "hard coded" to read and write files based on VRTC's directory structure and file naming conventions.

OFFSET REMOVAL, DIGITAL FILTERING, AND DERIVED CHANNEL COMPUTATION - The first step in the processing of all raw data files is to remove channel offsets, digitally filter the data, and compute the derived channels. The MATLAB function BUTTER is used to design a two-pass phaseless Butterworth low-pass digital filter for all channels except longitudinal velocity, with a 3-dB down cutoff frequency of 10 hertz and a total of 12 poles. Because of significant cyclic noise in the longitudinal velocity channel (at the frequency of the fifth wheel rotational speed), a second filter is designed for the longitudinal velocity channel with a 3-dB down cutoff frequency of 2 hertz and a total of 6 poles. This additional filtering of longitudinal velocity was necessary to get "reasonable" wheel slip estimates.

The only exception to the above is for the filtering of the ultrasonic displacement transducers. These transducers tend to have drop outs of a few samples rather than the random noise found on the other channels. For this type of signal, a running median filter is more appropriate. A running median filter takes a window of data around each point of the raw data, computes the median of the data in the window, and puts this median value in the output matrix. This procedure is repeated for each point in the input matrix. This type of filter has the characteristic of rejecting large amplitude, short duration spikes in the data, such as a dropout. The wider the window, the wider the dropout that will be rejected, but the lower the filter's cutoff frequency will be. The MATLAB command MEDFILT1 is used to filter the ultrasonic transducer output with a window size of 10 points.

A few of the channels also have offsets removed. The tape reading program uses the data from the pre-test calibration run to compute the steering wheel angle offset and the offset of the ultrasonic transducer readings. These values are stored in the data files as Keywords, and after these channels are filtered, the offset values are subtracted from each data point. The rate transducers used to measure roll, pitch and yaw rate are known to have a slow drift. Since the transducers are not zeroed by the Megadac before each test run, the measurements can get an offset. If there is more than 1 second of pre-trigger data in the test, and the mean value of the steering wheel angle is less than 10 degrees, then the mean value of the rate measurement, from the first data point until 25 points before time zero, is subtracted from the entire rate measurement channel. This can only be done for runs where there is some "straight ahead" running at the beginning of the data. For runs where there is no pre-trigger data, the rate measurements may be offset, but there is no way for the processing software to know their magnitudes.

A number of "derived" channels are also computed and included in the output file. Derived channels are channels computed by combining other channels, for example chassis roll angle, which is computed from the difference between the two ultrasonic ride height measurements. The derived channels are front and rear strut lengths, average steering wheel torque, steering rack position, power steering assist force, transmission gear number, chassis roll and pitch angles, and longitudinal wheel slip ratios.

Time-Domain Ensemble Averaging - In the late 1980's, VRTC conducted a large simulation evaluation project [4, 5] whose aim was to rank NHTSA's vehicle dynamics simulations. Much of the methodology developed during that project has been used in the current project. One development of the earlier project was to measure random experimental error by running repeat tests, and ensemble averaging each data channel in the time domain. There are two significant benefits to this procedure. First, by repeating each test (typically 10 times at VRTC), the influence of unmeasureable disturbances is greatly reduced (a wind gust, or a small area of the test surface with reduced friction for example). Secondly, it provides a measure of experimental uncertainty. It is VRTC's opinion that when comparing simulation predictions to experimental measurements, that the required accuracy of the simulation predictions can only be assessed within the experimental uncertainty (e.g. if experimentally, the uncertainty of the chassis roll angle measurement is ±1 degree, then if the simulation's roll angle prediction is within 1 degree of the experiment, it is accurate to within the uncertainty of the experiment).

This section describes the time-domain ensemble averaging of the test data. For each channel (measured and derived) in the filtered data files, the mean value and the 95-percent confidence interval of the mean are computed as a function of time. The interpretation of the confidence limits is that given the number of experiments, the mean value of the data is within the confidence interval to within 95-percent confidence (it does not say that the data will fall within the confidence interval to within 95-percent confidence, just that the actual mean of the data is within the interval). The 95-percent confidence interval around themean is computed by:

$$\varepsilon = t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \qquad (1)$$

where ε is the distance from the mean to the upper or lower boundary of the confidence interval, $t_{\alpha/2,n-1}$ defines the spread of the Gaussian distribution using the student t-distribution, s is the standard deviation of the data normalized by n-1 (computed using the MATLAB function std), and n is the number of runs used to compute the mean (see references 4 and 5 for a more complete explanation).

COMPUTATION OF VEHICLE FREQUENCY RESPONSES - The filtering and ensemble averaging procedures

described in the previous two sections prepare the experimental data for use with time-domain simulation evaluation methods. At VRTC, simulation evaluation is also performed in the frequency domain. Vehicle frequency responses to steering inputs are computed from the pulse steer (maneuver code P) tests for lateral acceleration, yaw rate, roll rate, and roll angle. The method used to compute the frequency responses is described in detail in [12], and the mathematics are not repeated here.

The frequency responses are computed based on repeat pulse steer runs. The test matrix defines both positive and negative steering angle pulse inputs at each of the three test speeds, and frequency responses are computed for each of the six conditions. For each run, the Power Spectral Density (PSD) of the steering wheel angle, the Cross Spectral Density (CSD) between lateral acceleration, yaw rate, roll rate, and roll angle and steering wheel angle, and the coherence between lateral acceleration, yaw rate, roll rate, and roll angle and steering wheel angle are computed. The MATLAB functions PSD, CSD, and COHERE are used to compute the PSD, CSD, and coherence. These functions are from the signal processing toolbox. All three functions use Welch's averaged periodogram method, which averages overlapping windowed (Hanning window) segments of the input vectors. The segments are 256 points long with a 128-point overlap.

After the PSD's, CSD's, and coherences have been computed for each run, for each frequency response to be computed, the mean value of the PSD, CSD, and coherence is computed (refer to reference [12] for details). The mean value of the frequency response is then computed by dividing the mean PSD by the mean CSD. The magnitude and phase angle of the complex result are found using MATLAB's abs and angle commands. The 95-percent confidence interval is computed using the equation

$$\varepsilon(f) = \cos^{-1}\sqrt{\frac{1-\left(\dfrac{1-\gamma_{io}^{2}(f)}{\sqrt[n]{1-P}}\right)}{\gamma_{io}^{2}(f)}} \qquad (2)$$

where $\varepsilon(f)$ is the confidence as a function of frequency f, $\Upsilon^2_{io}(f)$ is the computed coherence, n is the number of runs used, and P is the confidence level (in this case P=0.95).

SAMPLE RESULTS

This section presents a sample of the experimental measurements in both its raw and processed form. Issues of both experimental noise and repeatability are discussed. Computed vehicle frequency responses are also shown.

While the repeat testing and ensemble averaging performed at VRTC reduces the effects of random measurement noise, it is still important to keep the noise levels low. The term "measurement noise" is used here to represent the random variations in the data from transducer uncertainty, electrical and digitizing uncertainty, road roughness, variations in the tire/road friction, wind noise, other vibrations from the car itself, etc. This list is meant to represent everything in the physical measurement process and the vehicle that is not modeled by the vehicle dynamics simulations. The simulations represent idealized, deterministic vehicles operating on smooth, constant friction roads. The actual vehicle/road/measurement system is a stochastic process, operating in the presence of unknown disturbances.

To demonstrate the nominal levels of measurement noise, Figure 6 shows the lateral acceleration measurement from an 11 meters/second longitudinal velocity, 2 meters/second^2 lateral acceleration J-Turn run (V0RCA05). The dotted line is the raw data, and the solid line is the digitally filtered data. In the time range of 2 to 5 seconds, the vehicle is in steady-state (from a vehicle handling point of view). Using the raw data in this time period, the mean lateral acceleration is 1.69 meters/second^2, and the standard deviation is 0.067 meters/second^2. Using two standard deviations around the mean to contain 95 percent of the data, for this run, 95 percent of the measured data is within 8 percent of the mean. Figure 7 shows the lateral acceleration measurement from a 33 meters/second longitudinal velocity, 6 meters/second^2 lateral acceleration J-Turn run (V0RCX06). For this run, the mean lateral acceleration is 6.11 meters/second^2, and the standard deviation is 0.195 meters/second^2, so 95 percent of the measured data was within about 6 percent of the mean.

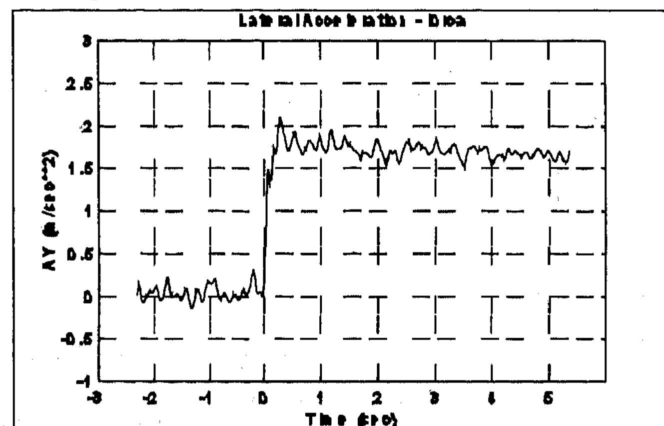


Figure 6. Lateral acceleration measurement from 2 m/s^2 J turn

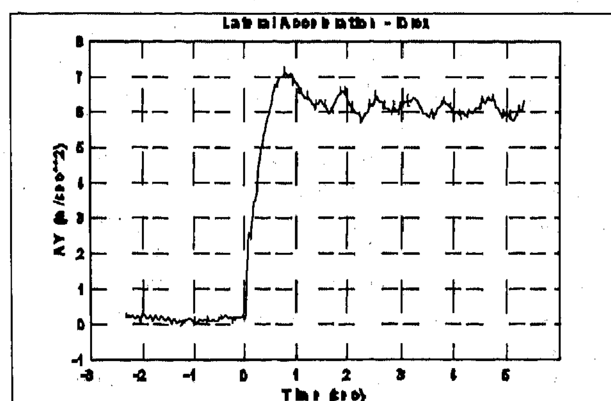


Figure 7. Lateral acceleration measurement from 6 m/s^2 J-turn

To show the test-to-test repeatability, the results of the ensemble averaging are used. Figure 8 shows the results of the time-domain ensemble averaging of the lateral acceleration of the 2 meters/second^2 J-turn used above. The dotted lines are the

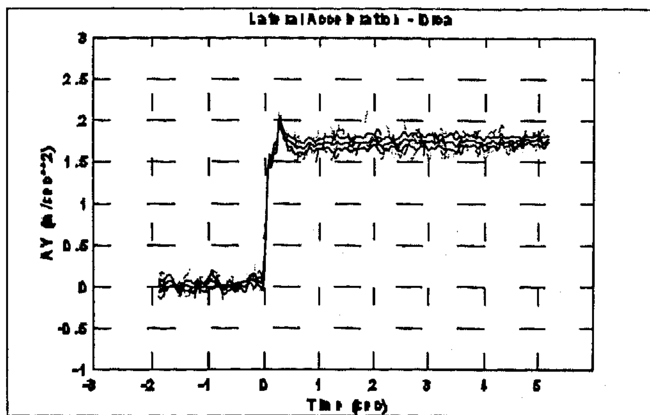


Figure 8. Ensemble averaging results from 2 m/s² J-turn filtered lateral acceleration measurements from the 10 repeat tests. The three solid lines represent the mean value, and the 95-percent confidence on the mean, based on the repeat tests. Figure 9 shows the same data with an expanded scale. As shown, the 95-percent confidence of the mean value is known to approximately ±0.05 meters/second².

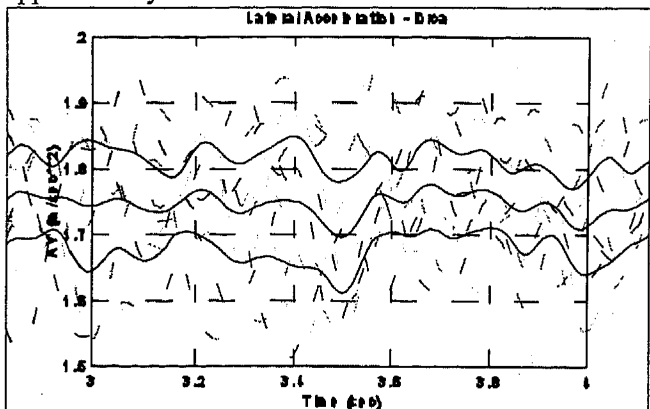


Figure 9. Ensemble Averaging Results - Expanded Scale

When running repeat testing, one source of test-to-test variation that can be controlled is the repeatability of the vehicle control inputs. For this testing program, the control inputs are steering wheel angle, brake pedal force, throttle position, and for the constant speed runs, cruise control speed setting. Using the 33 meters/second, 6 meters/second² J-Turn tests (V0RCX), the repeatability of the steering wheel inputs and the vehicle longitudinal velocity are demonstrated. Figure 10 shows the ensemble averaged results for the steering wheel angle. All of the filtered steering wheel angle data falls within approximately a 1-degree window, demonstrating that very good steering angle repeatability is achieved using the VRTC steering stop. Figure 11 shows the ensemble averaged results for the vehicle longitudinal velocity. As can be seen, the vehicle velocity from all 10 tests falls with in a window of approximately 0.5 meters/second.

Figure 12 shows the throttle position repeatability from the 50-percent throttle, straight-line acceleration runs (V0RAG). Excellent repeatability is achieved using VRTC's mechanical throttle pedal actuator, with all of the data falling within a window of approximately 1 percent of throttle position. Figure 13 shows the brake pedal force repeatability from the 5 meters/second² straight-line braking tests (V0RSD). The brake pedal force from all nine runs falls within a 6-newton window, demonstrating good repeatability using the pneumatic brake ram.

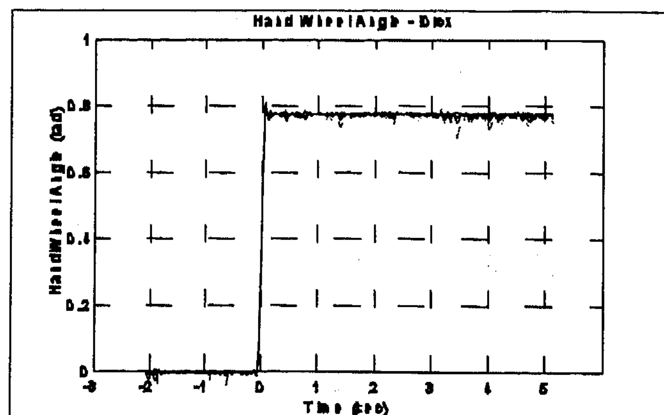


Figure 10. Steering Wheel Angle Repeatability

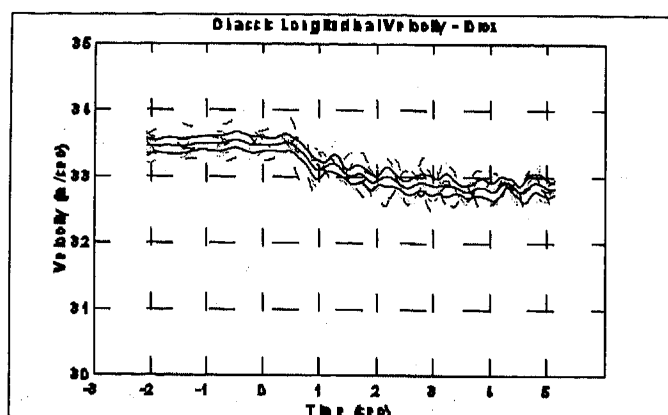


Figure 11. Longitudinal Velocity Repeatability

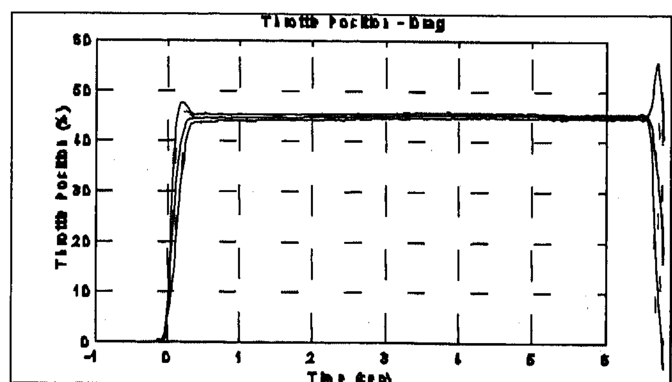


Figure 12. Throttle Position Repeatability

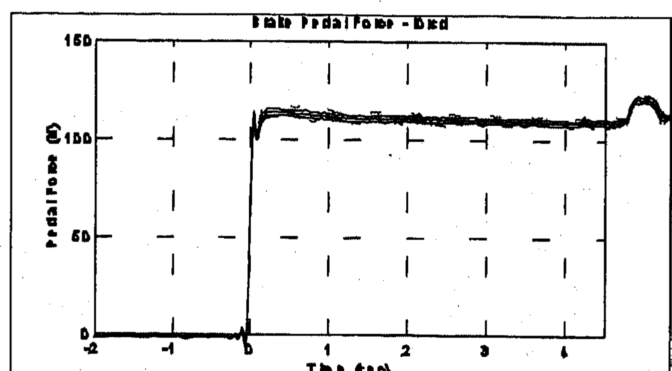


Figure 13. Brake pedal force repeatability

Figure 14 shows the vehicle yaw rate frequency response computed from the 22 meters/second runs (V0RPD). The frequency response is displayed from 2.5 to 20 radians per second (approximately 0.5 to 3 hertz). The top graph shows the magnitude mean and 95-percent confidence intervals of the mean. The center graph shows the phase angle mean and confidence intervals, while the bottom plot shows the mean coherence. As can be seen, for this set of runs, the mean coherence is near one across the entire frequency range. This indicates that there was sufficient input power at all frequencies, and that the input and output (steering wheel angle and vehicle yaw rate) are related in a "fairly" linear way under this maneuvering condition. The confidence intervals are relatively narrow, giving good confidence in all of this data.

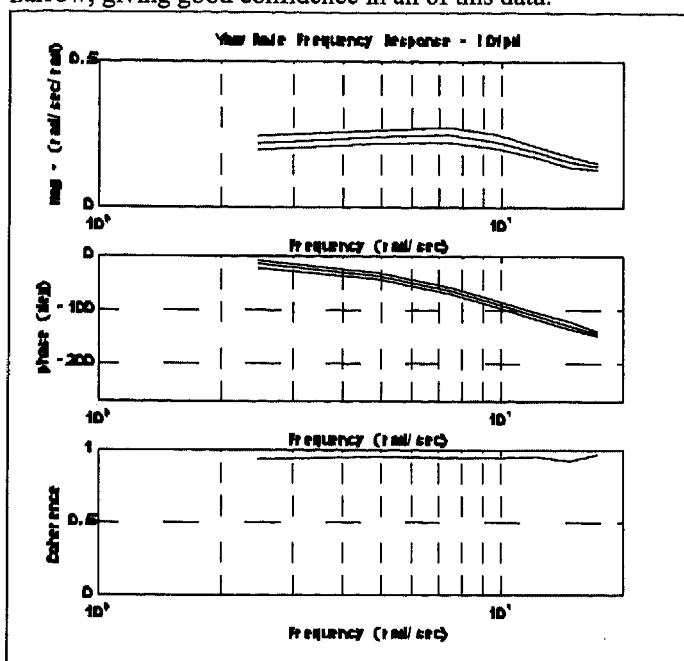


Figure 14. Computed yaw rate frequency response for 22 m/s runs

Figure 15 shows the lateral acceleration frequency response for the same set of test runs. Unlike the yaw rate frequency response, the coherence for the lateral acceleration starts to drop as the input frequency gets higher. Because of the lower coherence, the confidence intervals on the mean magnitude and phase angle get wider as the input frequency increases.

SUMMARY AND CONCLUSIONS

This paper describes the instrumentation and field testing of a 1994 Ford Taurus GL to generate benchmark data for VRTC's evaluation of the vehicle dynamics simulation to be used in NADS. The on-board instrumentation measured more than 40 transducers. The experimental program included 75 maneuver types/conditions and nearly 600 individual test runs.

The on-board instrumentation and data channels are listed. Each of the test maneuvers run by the Taurus is described, along with instrumentation setup, control actuation, test location, and driver procedures. The data reduction and data processing of the test data are detailed. Sample results of the testing, and an analysis of test repeatability and measurement noise, are also presented.

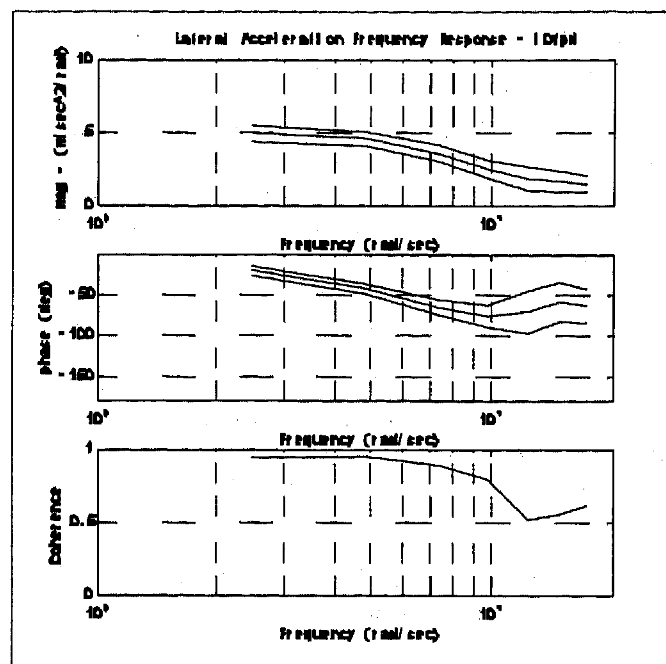


Figure 15. Computed Lateral Acceleration Frequency Response for 22 m/s Runs

VRTC performed this testing program to provide the data needed to evaluate the NADSdyna software. However, there is very little publicly available instrumented vehicle handling and braking test data, especially from a test program as comprehensive as the one described here. One of the purposes of this paper therefore is to make this test data available to other researchers. Electronic versions of the test data described in this paper can be obtained be contacting W. Riley Garrott at the Vehicle Research and Test Center at P. O. Box 37, East Liberty, Ohio, 43319, or via telephone at (937) 666-4511.

Future reports from the VRTC will describe the simulation parameter measurement and computation for the Taurus, and the simulation evaluation methodology and results. A similar program at VRTC will be performed using a Volvo-Fruehauf tractor-semitrailer, with reports and data to follow.

REFERENCES

1. Haug, E. J., et. al., "Feasibility Study and Conceptual Design of a National Advanced Driving Simulator", NHTSA Contract DTNH22-89-07352, Report No. DOT-HS-807-597, March 1990.
2. Tsai, F. F. and Haug, E. J., "Automated Methods for High Speed Simulation for Multibody Dynamic Systems", Technical Report R047, Center for Computer Aided Design, The University of Iowa, 1989.
3. NADS Vehicle Dynamics Software - Release 4", Center for Computer Aided Design, The University of Iowa, August 1995.
4. Heydinger, G. J., "Improved Simulation and Validation of Road Vehicle Handling Dynamics", Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1990.
5. Heydinger, G. J., Garrott, W. R., Chrstos, J. P., and Guenther, D. A., "A Methodology for Validating Vehicle Dynamics Simulations", SAE Paper 900128, 1990.

6. Lee, S., "Development of a New Dynamic Tire Model for Improved Vehicle Dynamics Simulation", Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1994.

7. Chrstos, J. P. and Grygier, P. A., "Instrumentation and field Testing of 1994 Ford Taurus GL for NADSdyna Evaluation", DOT report in progress.

8. Technical Manual for the Megadac Series 2000; Release 3.1, November 1986, Optim Corporation, Germantown, MD

9. Transportation Research Center Services & Facilities, Transportation Research Center Inc., East Liberty, Ohio.

10. Farrer, D. G., "An Objective Measurement Technique for the Quantification of On-Centre Handling Quality", SAE Paper No. 930827, 1993.

11. MATLAB® - Reference Guide, The Mathworks, Inc., July 1993.

12. Heydinger, G. J., Grygier, P. A., and Lee, S., "Pulse Testing Techniques Applied to Vehicle Handling Dynamics", SAE Paper 930828, 1993.

970564

Parameter Measurement and Development of a NADSdyna Validation Data Set for a 1994 Ford Taurus

Mohamed Kamel Salaani
Transportation Research Center, Inc.

Jeffrey P. Chrstos
JPC Engineering

Dennis A. Guenther
Ohio State Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper discusses the development of a 1994 Ford Taurus vehicle model for the National Advanced Driving Simulator's planned vehicle dynamics simulation, NADSdyna. The front and rear suspensions of the Taurus are modeled using recursive rigid body dynamics formulations. To complement vehicle dynamics, subsystems models that include steering, braking, and tire forces are included. These models provide state-of-the-art high fidelity vehicle handling dynamics for real-time simulation. The realism of a particular formulation depend heavily on how the parameters are obtained from the physical system. Therefore, the development of a data set for a particular model is as important as the model itself. The methodology for generating the Taurus data set is presented. The power train model is not yet included, so the simulation is run with the vehicle either at constant speed or decelerating.

INTRODUCTION

The U.S. Department of Transportation's (DOT) National Highway Traffic Safety Administration (NHTSA) is pursuing the development of the National Advanced Driving Simulator (NADS), an operator-in-the-loop high fidelity real-time vehicle driving simulator. NADS will materially assist NHTSA, other federal agencies, and commercial organizations in enhancing human factors and crash avoidance research. Research areas include: driver-vehicle-road interactions like driving under the influence of alcohol, human factors research concerning driver workload and adaptation to emerging Intelligent Transportation System (ITS) equipment and technology, and regulations regarding automotive safety. A realistic mathematical vehicle dynamics model is a prerequisite for performing this complicated task. Therefore, this research is principally concerned with developing a vehicle dynamics model of a typical passenger car with a high degree of realism for safety and stability analysis. A data set for the 1994 Ford Taurus is developed for simulation validation [27,28].

With the advent and proliferation of powerful computer hardware and with the availability of better DAE (Differential Algebraic Equations) numerical solutions, general purpose software based on multi-body formalisms have emerged and are playing the same role for mechanical systems as finite element methods are playing for complex structures. The multi-body program Real Time Recursive Dynamics (RTRD) was developed at the University of Iowa Center for Computer-Aided Design (CCAD) to solve general purpose multi-body dynamic systems, and will provide the core dynamics for the NADS. The RTRD is based on a topological analysis of a rigid mechanism formulated as modified recursive dynamics equations using minimum relative joint coordinates, and can be implemented with parallel computational algorithms [3, 18]. To simulate a vehicle, CCAD, aided by Iowa State University and the University of Wisconsin, has supplemented the RTRD(Figure 1). The RTRD with these subsystems are called NADSdyna. This program provides capabilities close to those offered by leading commercial multi-body dynamics software, such as ADAMS and DADS, in terms of rigid-body mechanism modeling. However, only a limited number of vehicle models have been developed using the RTRD [21], and a comprehensive validation of these models under severe crash avoidance scenarios has not yet been assessed.

Users of the multi-body approach must still understand and specify the vehicle response effects that they wish to model and work this into the multi-body approach allowed by a given formulation [15]. Furthermore, the multi-body dynamics do not cover wheel spin dynamics, tire forces, engines, aerodynamics, power steering systems, automatic controls, etc. These subsystems are among the primary elements that dominate vehicle behavior and control.

MULTI-BODY DYNAMICS MODELING

A passenger vehicle is a mechanical system consisting of a number of constrained rigid bodies. It depends on the vehicle type and the dynamic problem under consideration what type of constraints, how many bodies, and which degrees of freedom have to be assigned. In the proposed model, all the bodies are constrained

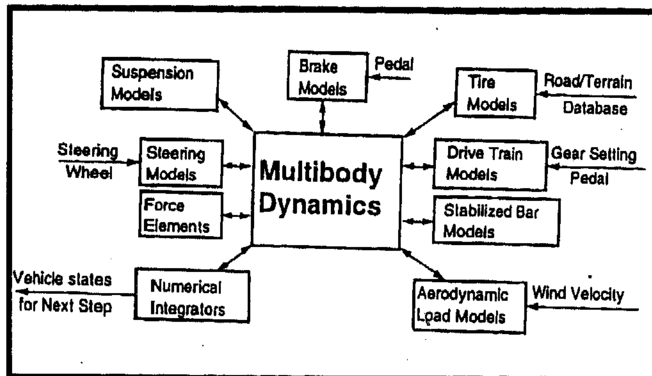
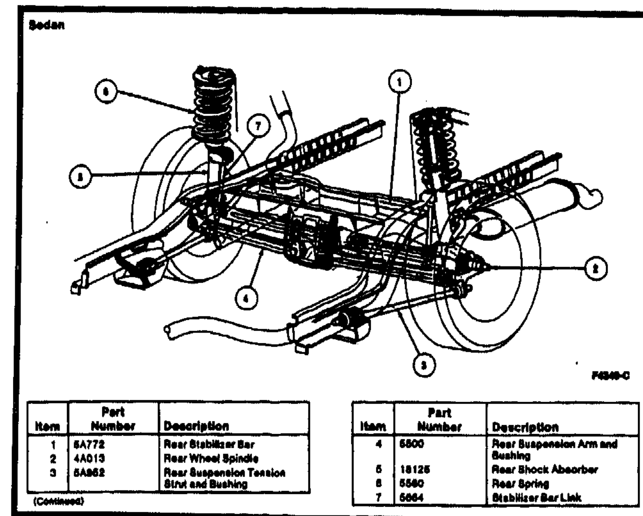


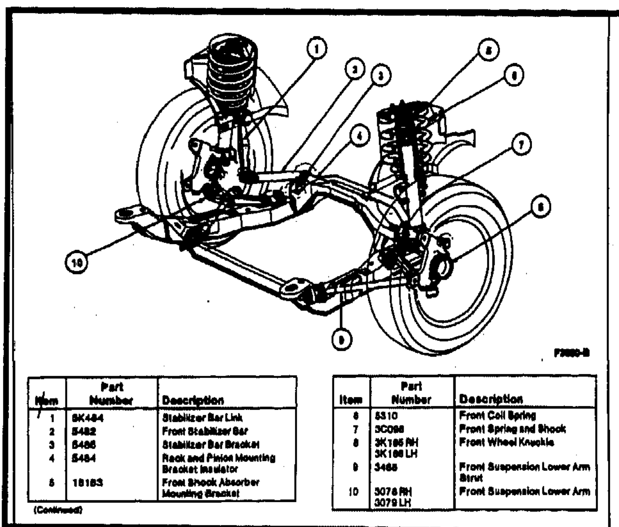
Figure 1. NADSdyna Systems Interactions



Item	Part Number	Description
1	5A772	Rear Stabilizer Bar
2	4A013	Rear Wheel Spindle
3	5A962	Rear Suspension Tension Strut and Bushing

Item	Part Number	Description
4	5500	Rear Suspension Arm and Bushing
5	18125	Rear Shock Absorber
6	5560	Rear Spring
7	5664	Stabilizer Bar Link

Figure 4. 1994 Ford Taurus Rear Suspension



Item	Part Number	Description
1	5K484	Stabilizer Bar Link
2	5482	Front Stabilizer Bar
3	5486	Stabilizer Bar Bracket
4	5484	Rack and Pinion Mounting Bracket Insulator
5	18183	Front Shock Absorber Mounting Bracket

Item	Part Number	Description
6	5310	Front Coil Spring
7	3C098	Front Spring and Shock
8	3K185 RH 3K186 LH	Front Wheel Knuckle
9	3468	Front Suspension Lower Arm Strut
10	3078 RH 3079 LH	Front Suspension Lower Arm

Figure 2. 1994 Ford Taurus Front Suspension

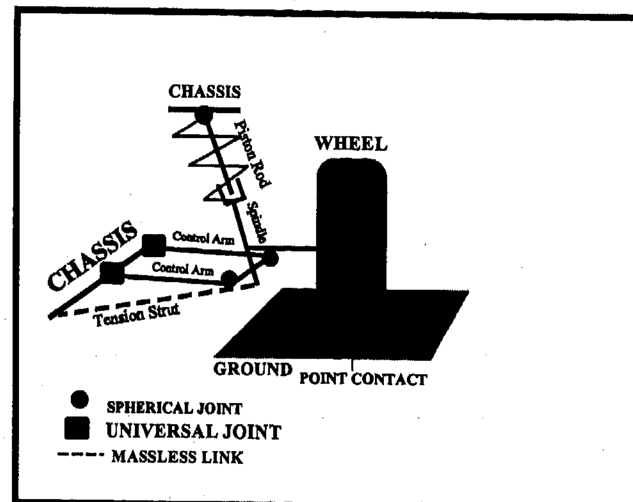


Figure 5. 1994 Ford Taurus Rear Suspension Model

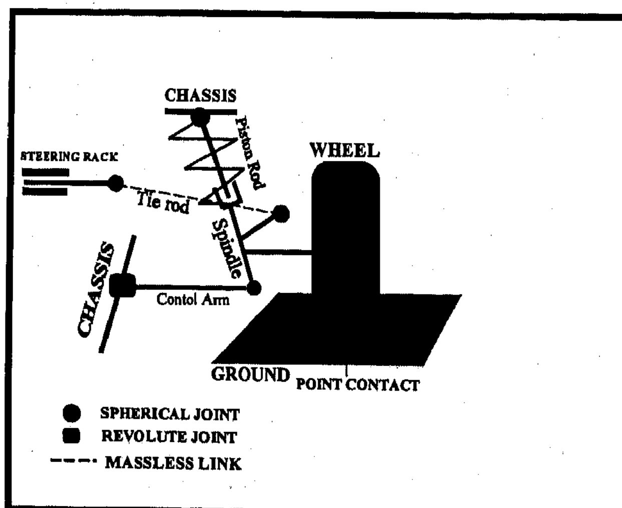


Figure 3. 1994 Ford Taurus Front Suspension Model

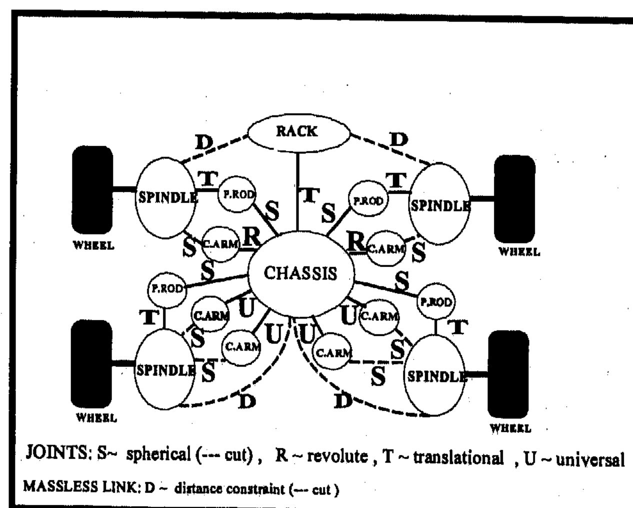


Figure 6. 1994 Ford Taurus RTRD Model

with rigid connections or joints. The left and right side of the vehicle are assumed to be symmetric. Friction forces in the suspension system exist at all metal-to-metal bearings. All the joints forming the suspensions are assumed to be frictionless.

FRONT SUSPENSION - The front suspension is a MacPherson strut type as shown in Figure 2. The kinematics of the model is similar to [12], and the spindle, lower control arm, and the piston rod are modeled as rigid bodies. The constraints are spherical joints between the spindle and the lower control arm, a spherical joint between the piston rod and the chassis, a translational joint between the piston rod and the spindle, and a revolute joint between the chassis and the piston rod. The tie rod connecting the spindle with the steering rack is not modeled as a rigid body; a kinematic distance constraint is used since its dynamics have a minor effect on the steering and suspension dynamics. If a rigid body had been used instead, the constraints would be a spherical joint between the tie rod and the spindle, and a universal joint between the steering rack and the tie rod. The steering rack and the chassis are connected with a translational joint that allows the steering rack to move parallel to the chassis frame to steer the front two wheels. This model provides the front suspension with two degrees of freedom with respect to the chassis; vertical movements, and rotations about the king pin axis. The model is shown in Figure 3.

REAR SUSPENSION - This suspension is a MacPherson strut type with two parallel control arms as shown in Figure 4. The multi-body model has the spindle, the piston rod, and the two control arms modeled as rigid bodies. The constraints are, a universal joint between the chassis and the control arms where rotations are around the vertical and horizontal axes, a spherical joint between the control arms and the spindle, a spherical joint between the piston rod and the chassis, and a translational joint between the spindle and the piston rod. The tension strut is not modeled as a rigid body, but with a kinematic distance constraint between the chassis and the spindle. This model gives the rear wheel one degree of freedom about the chassis, a circular vertical movement around the tension strut link. In the actual vehicle, due to bushing compliance, the wheel moves in an "elliptical" circular direction as forced by the control arms and the tension strut. The model is shown in Figure 5.

MODEL TOPOLOGY - The RTRD's recursive formulation requires rigid bodies to be connected in an open loop structure, or a spanning tree [1, 2]. The Taurus model contains ten closed loops, a joint on each loop is "cut" to get a spanning tree and the constraint equations associated with Lagrange multipliers are imposed on these cut joints to represent kinematically admissible motion. The procedure is based on cutting the joints that will result in a model with optimum computational efficiency characterized by the minimum number of coordinates and constraint equations [3]. Graph theory is used to identify the topological structure of multi-body dynamic systems, and weighting factors of some general kinematic joints are used for the purpose of minimizing the chain length and maximizing efficiency in parallel computation. For the dynamic steering system, the rack reaction forces can be derived from the Lagrange multipliers that correspond to the cut distance constraints between the wheels and the rack. The overall model is shown in Figure 6.

MULTI-BODY DYNAMICS PARAMETER MEASUREMENT - The vehicle center of gravity position, and mass moments of inertia were measured at VRTC on the Inertial Parameter Measurement Device (IPMD) [32]. The IPMD is a facility used to measure total vehicle center of gravity height and its mass moments of inertia about its center of gravity around its x, y, and z axis (as defined by SAE). The IPMD is a simple stable pendulum structure that the vehicle is driven onto and secured. Center of gravity height is estimated by applying a torque about the IPMD's pivot axis. Mass moments of inertia about the vehicle's X and Y axes (roll and pitch) are estimated by causing the vehicle/IPMD to oscillate about its pivot axis, and measuring the period of oscillation. Mass moment of inertia about the vehicle's Z axis (yaw) is measured by connecting springs to a table bearing that supports the vehicle, and causing the system to oscillate and measuring the period of oscillation. The location of the vehicle center of gravity in the horizontal plane and the total vehicle mass is estimated by measuring the four wheel weights. The location of all suspension attachment points on the chassis, the geometry of each suspension component, the mass, the center of gravity location, and mass moments of inertia were measured for VRTC by the Transportation Research Center Inc. Bi-filar and tri-filar torsional pendulum's are used to measure the mass moments of inertia of small components, and the charts and schematics are included in [25].

The geometric data consists all the three-dimensional coordinates of all connecting joints of the rigid bodies included in the dynamic formulations. These geometric data are transformed to generalized data format that consists of points and vectors forming the connectivity and constraints of the model with respect to a chassis reference point. The spindle assembly is made of the following components: wheel, brake rotor, strut without spring, knuckle, and brake caliper. The chassis inertia, mass and center of gravity are approximated by deducting the properties of the four spindles from the total vehicle. The chassis roll, pitch, and yaw inertia's are computed as follows :

$$I_{xc} = I_{xx} - m_c(z_t - z_c)^2 - 2m_f\left((z_f - z_t)^2 + y_f^2\right) - 2m_r\left((z_r - z_t)^2 + y_r^2\right) - 2I_{fx} - 2I_{rx} \quad (1)$$

$$I_{yc} = I_{yy} - m_c\left((z_t - z_c)^2 + (x_t - x_c)^2\right) - 2m_f\left(l_f^2 + (z_t - z_f)^2\right) \quad (2)$$
$$-2m_r\left(l_r^2 + (z_t - z_r)^2\right) - 2I_{fy} - 2I_{ry}$$

$$I_{zc} = I_{zz} - m_c(x_t - x_c)^2 - 2m_f\left(l_f^2 + y_f^2\right) - 2m_r\left(l_r^2 + y_r^2\right) - 2I_{fz} - 2I_{rz} \quad (3)$$

Where:

I_{xx}, I_{yy}, I_{zz} : Roll, pitch and yaw inertia of the total vehicle

I_{xc}, I_{yc}, I_{zc} : Roll, pitch and yaw inertia of the chassis

I_{fx}, I_{fy}, I_{fz} : Inertia of the front spindle

I_{rx}, I_{ry}, I_{rz} : Inertia of the rear spindle

l_f, l_r : Front and rear x-distance of the spindle c.g. to the vehicle c.g.

y_f, y_r : Y-distance from the vehicle x-axis to the front and rear spindle respectively.

m_t, m_c, m_f, m_r : The total vehicle, chassis, front spindle, and rear spindle masses respectively.

x_t, x_c, x_f, x_r: x-distance from the reference point to the c.g. of the total vehicle, chassis, front spindle, and rear spindle respectively

z_t, z_c, z_f, z_r : z-distance from the reference point to the c.g. of the total vehicle, chassis, front spindle, and rear spindle respectively

The multi-body model was first implemented in DADS [27]. After satisfactory qualitative results were achieved, the data were transformed to RTRD formulations. To transform the geometric data to a relative coordinate system, the direction cosine of each rigid body is computed with respect to the global coordinate system to obtain the orientation matrices. All the coordinates and

constraint vectors are transformed to global vectors with origins at the local frame center, then transformed to local frame orientations. The assembly is checked using the generalized inverse or pseudo inverse method to solve the nonlinear constraint equations [18].

The inertia is transformed to local frame, using the similarity transformation as follows,

$$J' = A^T J A \qquad (4)$$

where:

- J : Inertia in global coordinate system
- J' : Inertia tensor in local frame
- A : Transformation matrix of the rigid body

The RTRD rigid body preprocessor is used to analyze and generate topology and rigid body input files using the geometric and inertial data [18]. The initial conditions that consist of rigid body position and velocity are determined from a drop test simulation.

SPRING AND SHOCK ABSORBER - The two main dissipative elements of the suspension are dry friction and shock absorber viscous damping. For general handling maneuvers, cubic spline interpolation of the velocity-force shock absorber experimental curve is a good approximation (Figure 7). The springs and bump stops are modeled as nonlinear springs (Figure 8). In the multi-body formulations, these forces are formulated as external forces [18, 19].

The stiffness and the damping characteristics of the four strut units on the Taurus were measured for VRTC by Heitz Automotive Testing, Inc. A fixture was built to hold the strut top mount, with or without the spring installed, and drive the strut body using a hydraulic servo cylinder. The stiffness test was run by stroking the cylinder slowly (0.1 in/sec). A small amplitude, 100 Hz dither, was added to the input to simulate the vibration during actual driving. Force levels were between 0-1600 pounds, which provided data from the spring and the bump and rebound stops of the strut. The damping was measured by removing the springs and commanding ± 3 inch triangular inputs with varying ramp velocities between 1 and 70 inches per second. The spacing between velocities was closer for the lower velocities, and greater for the higher velocities.

VEHICLE SUBSYSTEMS MODELING

For a complete vehicle model, the multi-body formalism is augmented with other subsystems consisting of engine, power train, wheels, tires, steering, braking, aerodynamic forces, and terrain database. The power train and engine models are important to study vehicle traction characteristics, traction control systems, and the influence of drive-wheel traction on steering control [15]. A Ford Taurus model of the power train and engine is currently under development. At this phase of the project a simple cruise control is used to keep the vehicle at constant speed.

WHEEL KINEMATIC VARIABLES - The wheels are not modeled as rigid bodies, only the spin degree of freedom is formulated using this equation:

$$J\dot{\omega} = T - F_x r \qquad (5)$$

Where :

- J : Wheel spin inertia
- T : Applied torque
- $\dot{\omega}$: Wheel spin acceleration
- F_x : Tire longitudinal force
- r : Wheel rolling radius.

The longitudinal and lateral slips are solved using first-order differential equations as suggested by Bernard et al. [10]. The equations are the following;

$$\dot{\kappa} + \frac{|u|}{\lambda}\kappa = \frac{|u| - r\omega sign(u)}{\lambda} \qquad (6)$$

$$\frac{\partial\, tan(\alpha)}{\partial t} + \frac{|u|}{\beta} tan(\alpha) = \frac{v}{\beta} \qquad (7)$$

Where:

- κ : Longitudinal slip
- λ : Longitudinal relaxation length
- u : Longitudinal velocity of the wheel center
- ω : Wheel spin velocity
- α : Lateral slip
- β : Lateral relaxation length
- v : Lateral velocity of the wheel center.

The lateral relaxation length is obtained from experimental lateral dynamics testing. The phase angle of the lateral slip differential equation is the following,

$$\phi = tan^{-1}(f\tau) \qquad (8)$$

Where:

- ϕ : is the phase angle (rad)
- τ : time constant (sec)
- f : frequency (rad/sec)

The relaxation length is,

$$\beta = \tau \times u \qquad (9)$$

From experiment:

$\phi = 0.785$ rad

$f = 21.4$ rad/sec

$u = 13.5$ m/sec

then $\tau = 0.0468$ sec [4, 23] and $\beta = 0.628$ m (22). A procedure to measure longitudinal relaxation length has not yet developed. The longitudinal relaxation length is set to $\lambda = 0.25$ m, which is according to simulation frequency and how responsive the tire is to longitudinal force excitation.

TIRE FORCES - The vertical tire force is computed from a spring/damper/mass tire model using either the point contact follower or the envelope contact method. The contact follower model has been extensively used by others, and is valid for the frequency range of interest (~10 Hz), when the surface profile inputs do not contain short wavelengths compared to the tire contact patch. The tire vertical force is then formulated as follows:

$$F_n = k\delta + B\dot{\delta} \qquad (10)$$

Where:

- F_n : Tire normal force
- k : Tire vertical stiffness
- B : Tire vertical damping
- δ and $\dot{\delta}$: Tire vertical deflection and deflection rate respectively

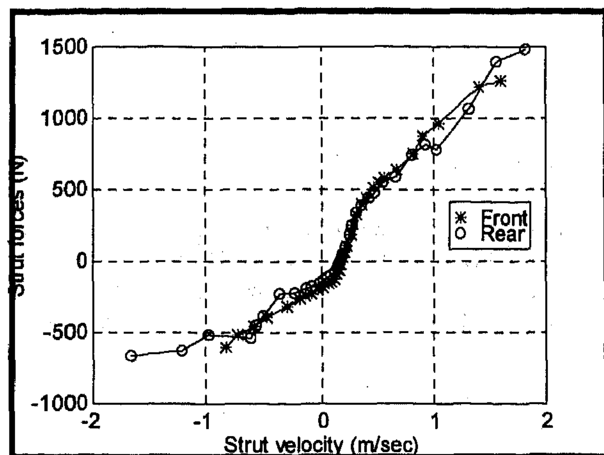


Figure 7. Shock Absorber

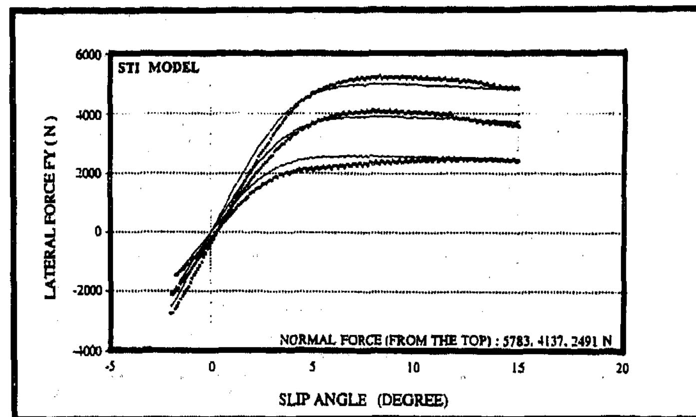


Figure 10. Tire Lateral Force
General Ameritech P205/65R15--Solid: model; dashed : exp.

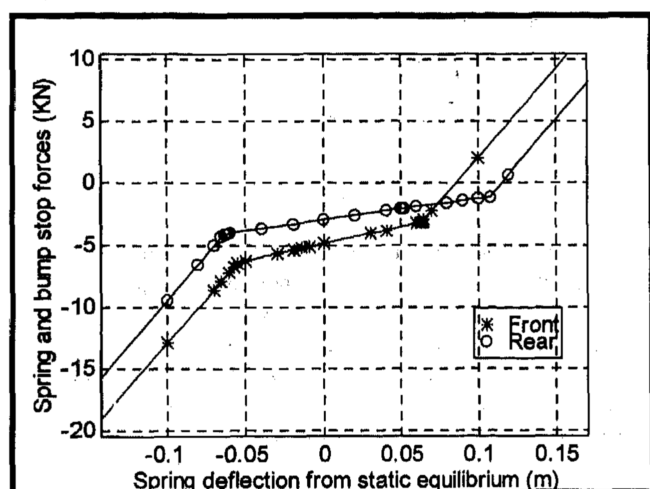


Figure 8. Suspension Spring and Bump Stops

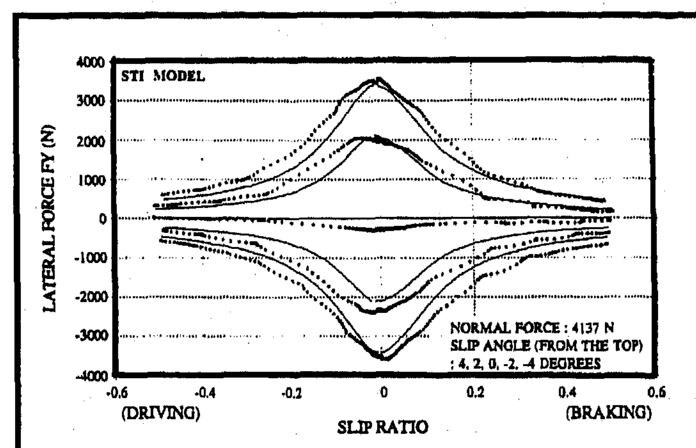


Figure 11. Tire Lateral Force Vs. Slip Ratio
General Ameritech P205/65R15--Solid: model; dashed : exp.

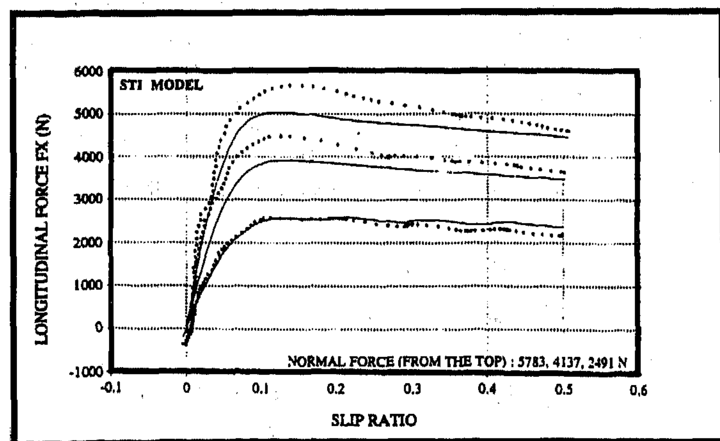


Figure 9. Tire Longitudinal Force
General Ameritech P205/65R15--Solid: model; dashed : exp.

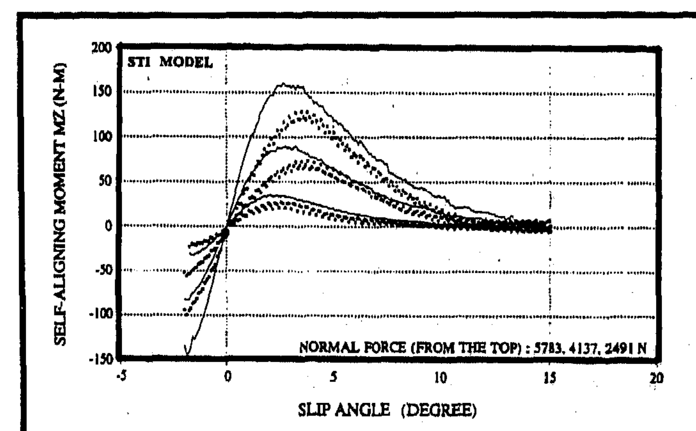


Figure 12. Tire Alignment Moment
General Ameritech P205/65R15--Solid: model; dashed : exp.

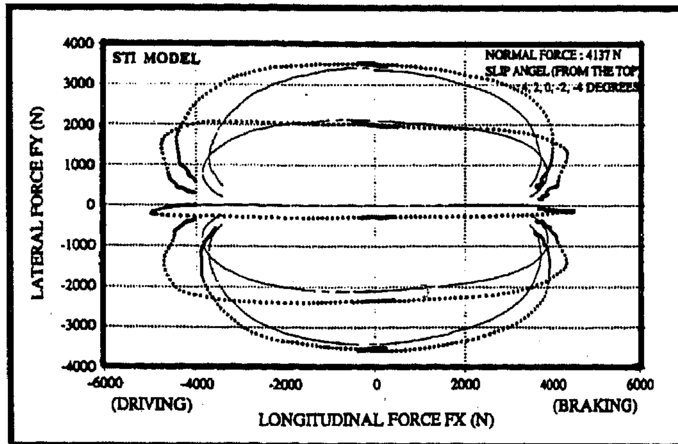


Figure 13. Tire Lateral Force Vs. Longitudinal Force
General Ameritech P205/65R15–Solid: model; dashed: exp.

The vertical tire stiffness, *220.3* KN/m, is determined using linear regression of sinusoidal vertical loads applied on the wheel, and the vertical damping is estimated to be 500 Ns/m [27].

A clear understanding of the lateral and longitudinal tire forces is the key for any successful vehicle simulation. These forces depend on the complex interaction of properties of these three subsystems: tire kinematics and elasticity, road profile, and the tribology of the contact. These forces are best formulated empirically using either Pacejka [9], or STI [5,6] models. These models are a means for presenting a complex mechanics problem as cost effectively as possible. The STI model provides good predictions in severe combined braking and cornering maneuvers, and better insight to the mechanics of the tire, since most of its parameters correspond to some physical properties. The lateral force dynamic effect is included as part of the lateral slip angle first order linear differential equation.

The tires used on the Taurus during all field testing were General Ameritech P205/65R15 steel belted radials. Twenty six tires were purchased from a single manufacturing batch in an attempt to minimize the tire to tire variability. Six of these were sent to Smithers Scientific Services, Inc., and were tested on their MTS Flat-Track II [31] flat-belt tire force and moment measuring machine. A complete description of these testing and data can be found in [4], and the complete test matrix is documented in its Appendix A. Seven types of tests were run at three normal loads (560, 930, and 1300 lb.), and at a variety of severity's. The tests are:

1. Quasi-static steering
2. Quasi-static braking/driving
3. Dual step steering
4. Pseudo pulse braking
5. Pseudo pulse steering
6. Discrete sinusoidal steering
7. Discrete sinusoidal loading

The parameters for the STI model are determined using a local VRTC STI preprocessor [23], and the model compared to experimental measurements are shown in Figures 9-13. The rolling resistance was determined from averaged coast down tests (starting at 33 m/sec) using the following equation,

$$R_r = \sum_{i=1}^{n} \frac{m a_{xi} - 0.5\rho C_F A_F V_i^2}{nW} \qquad (11)$$

where:

m : Vehicle mass (kg)

a_x : Vehicle deceleration (m/sec^2)

$0.5\rho C_F A_F V_i^2$: Aerodynamic drag force (N)

W : Vehicle weight (N)

R_r : Rolling resistance

n : Number of runs $(n = 10)$

This value is actually the consolidation of tire rolling resistance, bearing friction, and drive line drags. Since vehicle deceleration is small, and noise to magnitude ratio is significant, the deceleration is determined by numerically differentiating the longitudinal velocity as follows,

$$V_i = \sum_{j-5}^{j+5} \left(V_j / 10\right) \qquad (12)$$

$$c_2 t_i^2 + c_1 t_i + c_0 = V_i \qquad (13)$$

$$a_{xi} = 2c_2 t_i + c_1 \qquad (14)$$

Where:

V_i: Filtered velocity

V_j: Measured velocity

a_{xi}: Longitudinal acceleration

c_2, c_1, c_0 : Coefficients determined using least square of (i-10,i+10) data points.

The friction coefficient between the tire and the road is of a paramount importance. However, it is impossible to get the exact value, since the friction is dependent on many uncontrolled variables, such as road profile, road contaminants, temperature, and tire wear during testing. Based on prior testing experience, a nominal value of 0.85 is used. This might mask errors if the center of gravity height of the vehicle and other vehicle parameters are erroneous.

COMPLIANCES - Compliances result from flexibility and deflection of the suspension components. These deflections cause some additional steering angle to be produced by the wheels which affect understeer properties and consequently lateral/directional handling and stability [11]. The compliances can be modeled either by a quasi-static change of the steer and camber angles, or by an addition of stiff compliant elements in bushing localities. However, quasi-static implementation is sufficient since compliances have a much higher frequency than vehicle handling dynamics response and the inclusion of compliant elements result in a dramatic decrease of the integration time step that would impede real-time practicality. Stiff compliant elements will be important if wheel shimmy effect or any other high frequency vibration modes are to be analyzed. The deflection of the wheel due to compliance effects is formulated as follows,

$$\gamma_c = \lambda_{\gamma y} F_y + \lambda_{\gamma z} M_z \qquad (15)$$
$$\delta_c = \lambda_{\delta y} F_y + \lambda_{\delta z} M_z$$

Where:

$\lambda_{\gamma y}$, $\lambda_{\gamma z}$, $\lambda_{\delta y}$, and $\lambda_{\delta z}$: Compliance coefficients

γ_c, and δ_c : Camber and steer angles due to tire forces

F_y and M_z : Lateral force and alignment moment of the tire.

The camber, and lateral slip angles used in the tire model are adjusted as follows,

$$\gamma_t = \gamma + \gamma_c \qquad (16)$$
$$\alpha_t = \alpha + \delta_c$$

Where:

γ : Wheel camber angle

α : Lateral slip angle

γ_t : Camber angle used in tire forces computations

α_t : Lateral slip angle used in tire forces computations

The suspension compliances to tire forces for the Taurus were measured for VRTC by Heitz Automotive Testing, Inc. Heitz has developed the "Heitz Topsy chassis parameter measurement system" that secures a passenger vehicle to a base-plate frame, and applies forces at the tire contact patch. Wheel steer and camber angles are measured along with the transitional motions of the wheel center. The chassis can be adjusted vertically, allowing the compliance tests to be run at different tire normal loads. Forces at the tire contact patch are supplied by two hydraulic servo cylinders that can be configured to provide lateral and longitudinal forces and aligning moments individually or in combination [22, 33]. The data were measured for left and right sides individually. In the rear, they can be expected to be independent, since left and right sides are not connected. In the front, opposing and adding configurations are different when reacted differently through the compliance of the steering gear and column.

ROLL STIFFNESS - Maneuvering causes normal load to be transferred from the inside to the outside tires away from the direction of the turn due to vehicle lateral accelerations, which degrades lateral force sensitivity of the inside tires. The stability effects that are influenced by load transfer include directional and roll motions and rear axle wheel lockup. The effect of roll stiffness can be modeled either by simply applying quasi-static forces, or by including rigid bodies connected by stiff torsional springs. Since the contribution of the stabilizing torsional bar to the vehicle handling dynamics is a low frequency adjustment of the applied vertical tire force, it is not necessary to add more computational burden by including more bodies and stiff springs. This element is modeled by applying external forces on the spindle as follows,

$$F_t = \frac{K_t(\varphi_c - \varphi_w)}{t_w} \qquad (17)$$

Where:

φ_c and φ_w : Chassis roll angle and wheel axle roll angle respectively (rad)

k_t : Torsional stiffness of the bar(N.m/rad)

t_w : Track width (m)

F_t : Applied force on the spindle (N)

A restoring moment acting about the longitudinal x-axis of the chassis is applied as follows,

$$M_t = -K_t(\varphi_c - \varphi_w) \qquad (18)$$

Where:

M_t : Applied moment about the chassis x-axis.

The value of the stiffness can be approximated theoretically using strain energy formulations. However, experimental procedure is used because it is simple and the results are more comprehensive. The roll stiffness are measured at VRTC using the roll test described in [29]. The roll moment is applied by mounting a beam under the vehicle, parallel to the vehicle's Y-axis. A weight is placed on the beam at various locations to apply various roll moments.

AERODYNAMIC FORCES - Aerodynamic forces and moments acting on moving vehicles are of growing interest to the vehicle dynamicist. Through the motion relative to the air, a pressure field is generated on the vehicle surface whose integral effect is felt as a resultant aerodynamic force with time-varying points of contact. Positive lift on front wheels reduces the tire-to-road forces, and leads to decrease of effective wheel torque in tractive wheels. As lift forces increase, handling quality such as lateral guidance can rapidly deteriorate, reducing driving comfort and creating dangerous maneuvering at high speeds. A review of road accidents due to wind effects shows that these are caused during high speed travel and rapid temporal change of ambient wind [14]. The cause of these accidents is the slow response of the vehicle to corrective measures of the driver that leads to collision or displacement off the road.

SAE J1594 JUN87 aerodynamic recommended practice is used to model aerodynamic forces [17]. These forces are formulated by first computing the relative wind velocity in the chassis coordinate system as follows:

$$V_{rw} = A_c^T(V_w - V_c) \qquad (19)$$

Where:

V_w : Wind velocity in global coordinate system

V_c : Chassis velocity in global coordinate system,

V_{rw} : Velocity of the wind relative to the chassis in chassis local coordinate system.

Then the forces and moments are computed as follows,

$$F_a = AQ(C_x, C_y, C_z)^T$$
$$M_a = AQL(C_{mx}, C_{my}, C_{mz})^T \qquad (20)$$

$$Q = 0.5\rho V_{rw}^2 \qquad (21)$$

Where:

A : Vehicle frontal area

L : Vehicle wheel base

Q : Aerodynamic air pressure

ρ : is the air density

C_x, C_y, C_z : Aerodynamic force coefficients

C_{mx}, C_{my}, C_{mz} : Aerodynamic moments coefficients.

These coefficients are obtained from cubic spline interpolation of the aerodynamic curves (Figure 14) using the slip angle β.

$$\beta = tan^{-1}\left(\frac{V_{rw}(2)}{V_{rw}(1)}\right) \qquad (22)$$

Aerodynamic force and moment coefficients were measured at the Lockheed Aeronautical Systems Company's Low Speed Wind Tunnel (LSWT) in Marietta, Georgia. The LSWT is "...horizontal, atmospheric pressure, single return circuit, closed throat system having an overall circuit centerline length of 238 meters" [30]. The test section where the Taurus was tested is 23.5 feet wide by 16.25 feet high. The vehicle sits on a six-component pyramidal external balance system. The balance is a small-seeking balance with weights driven by lead-screws. The aerodynamic coefficients were measured at 25, 45, and 65 mph air speed. The yaw angle was varied from -90 to 90 degrees. The data are expanded to include full vehicle yaw rotation using vehicle geometric symmetry [27]. Other aerodynamic effects, such as wind gusts, will be added to a later version of the simulator software.

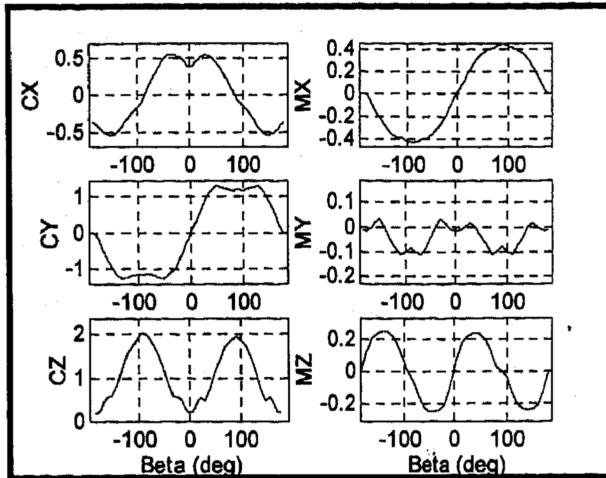


Figure 14. Aerodynamic Curves

BRAKE SYSTEM - The brake system is formulated as follows,

$$\tau \dot{P}_f + P_f = f_1(F_b) \qquad (23)$$

$$P_r = P_f \qquad P_f \le P_l$$
$$P_r = P_l + K_p(P_f - P_l) \qquad P_f > P_l \qquad (24)$$

$$T_{bf} = f_2(P_f) \qquad (25)$$

$$T_{br} = f_3(P_r) \qquad (26)$$

Where:

F_b : Applied brake pedal force

P_f : Front line brake pressure

P_r : Rear line brake pressure

T_{bf} : Front brake torque

T_{br} : Rear brake torque

τ : Brake system time lag

$f_1(F_b)$: Function relating brake force to front-line brake pressure

$f_2(P_f)$: Function relating front-line brake pressure to wheel brake torque

$f_3(P_r)$: Function relating rear-line brake pressure to wheel brake torque

K_p : Rear-to-front brake line proportion after knee point

The hysteresis effect is not modeled, but the parameters will be estimated within the hysteresis envelope. Suspension deflection effect is not included so far; a more comprehensive model will be developed later.

The measurements of brake pedal force, front-line brake pressure, and rear-line brake pressure were done at VRTC. The vehicle was stationary and the engine was running, then a cyclic increasing brake pedal force was applied slowly. The brake pedal force to front-line brake pressure relation (Figure 15) was obtained, and linear regression was used to determine the curve data points used in NADSdyna. The rear line brake pressure to front line brake pressure proportioning (Figure 16) was determined using the same experiment. The brake line pressure to brake torque relation is determined using VRTC's Road Transducer Plates (Figure 17).

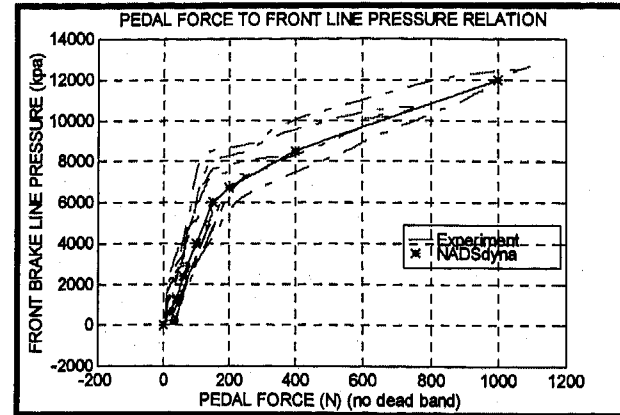


Figure 15. Front Line Brake Pressure to Brake pedal Force

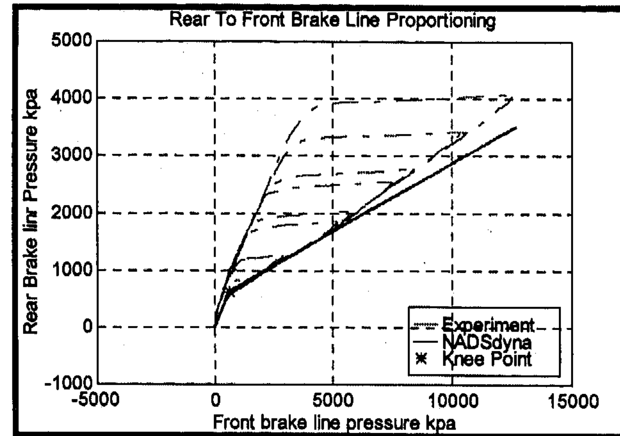


Figure 16. Rear To Front Brake Line Pressure Proportioning

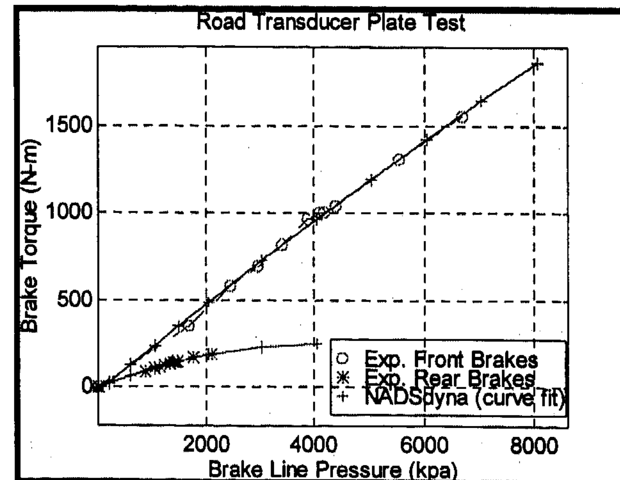


Figure 17. Brake Line Pressure to Braking Torque

STEERING SYSTEM - The rack is kinematically driven through a translation joint relative to the chassis. Its displacement is computed as follows,

$$d_{rack} = \frac{\omega_s}{r_p}$$
$$\dot{d}_{rack} = \frac{\dot{\omega}}{r_p} \quad (27)$$

Where:

d_{rack}, $\dot{d}_{rack}$: Rack displacement, and velocity respectively

ω_s, $\dot{\omega}_s$: Handwheel steering angle, and velocity respectively

r_p : Effective pinion pitch radius (134.8 mm/rad)

This model lacks the dynamics of the steering system from steering handwheel to the rack and the torque feedback to the driver. A variable assisted dynamic model is currently under development.

CRUISE CONTROL - The power train subsystem used in NADSdyna is based on a one-dimensional torque formulation that includes the torque and angular velocity transmitted through each model component [18]. The components include engine, automatic transmission gearbox, and open differential. Since at this phase of the project, measurement of power train data of the Ford Taurus has not yet been completed, the vehicle is simulated at either constant speed or in a deceleration state. Therefore, a cruise control model is applied to a quasi-steady-state engine model, and it is formulated as follows:

$$\delta_{th} = k_v\left(u_d - u_x\right) + k_a a_x \text{ if } u_x < u_d$$
$$\delta_{th} = 0 \qquad \text{if } u_x \geq u_d \quad (28)$$

Deactivate control if brake is applied

Where:

δ_{th} : Throttle input

k_v , k_a : Control constants

u_x , a_x : Longitudinal vehicle velocity and acceleration

u_d : Desired vehicle longitudinal velocity

The data of the power train are adopted from [18] of the HMMWV vehicle. The power train data do not affect simulation results, since only a small drive torque feedback to the wheel spin dynamics is applied to keep the vehicle at constant speed.

CONCLUSION

The model developed is for high fidelity vehicle dynamics simulation. Recursive multi-body dynamic formulations are used to model rigid-body nonlinear motions, compliances with high spring stiffness have been modeled as quasi-static force elements. The vehicle subsystems models constitute the fundamentals to complement vehicle dynamics simulation. These are as important as rigid body dynamics, especially tire forces that determine vehicle motion on the ground. Power train and variable assist dynamic power steering system models will be included in a later version of model development.

This model is evaluated in [28], and the results show promise that the Ford Taurus NADSdyna model is a high fidelity vehicle dynamics simulation.

Report [25] contains a complete description of the model and the data set, and report [26] discusses its validation. The modeling effort is being extended to include a complete dynamic steering system, as well as other vehicles including a tractor semi-trailer.

REFERENCES

1. Bae, D.S., and Haug, E. J., "A Recursive Formulation for Constrained Mechanical Systems, Part I Open Loop," Mechanics of Structures and Machines, Vol. 15, No.3, pp. 359-382, 1987.

2. Bae, D.S., and Haug, E. J., "A Recursive Formulation for Constrained Mechanical Systems, Part II Closed Loop," Mechanics of Structures and Machines, Vol. 15, No.4, 1987.

3. Tsai, F.F., and Haug, E. J., "Automated Methods for High Speed Simulation for Multi-body Dynamic Systems," Technical Report R047, Center for Computer Aided Design, The University Of Iowa, 1989.

4. Lee, S. W, "Development of New Dynamic Tire Model For Improved Vehicle Dynamics Simulation," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1994.

5. Allen, R. W., Rosenthal, T.J. and Szostak, H.T., "Analytical Modeling of Driver Response in Crash Avoidance Maneuvering," DOT-HS-807-270, Systems Technology, Inc., April 1988.

6. Allen, R. W. , "Vehicle Dynamic Stability and Rollover," NHTSA DTNH 22-88-C-07384, Systems Technology, Inc., January 1991.

7. Haug, E. J., et al., "Feasibility Study and Conceptual Design of a National Advanced Driving Simulator," Report No. DOT-HS-807-597, March 1990.

8. Pacejka, H. B., Bakker, E. and Nyborg, L., "Tyre Modeling for Use In Vehicle Dynamics Studies," SAE Paper 870421, February 1987.

9. Pacejka, H. B., Bakker, E. and Lidner, L., "A New Tire Model with an Application in Vehicle Dynamics," SAE Paper 890087, February 1989.

10. Bernard Jim, and Clover Chris, "Tire Modeling for Low-Speed and High-Speed Calculations," SAE Paper 950311, February 1995.

11. Bergman, W., "Effects of Compliances on Vehicle Handling Properties," SAE paper 700369.

12. Cronin, D. L., "MacPherson Strut Kinematics," Mechanism and Machine Theory, Vol. 16, No. 6, pp. 631-644, 1981.

13. Kortum, W., and Schiehlen, W., "General Purpose Vehicle System Dynamics Software Based on Multi-body Formalisms," Vehicle System Dynamics, Vol. 14, pp. 229-263, 1985.

14. Ahmed, S. R., Gawthorpe, R. G., and Mackrodt, P. A., "Aerodynamics of Road and Rail Vehicles," Vehicle System Dynamics, Vol. 14, pp. 319-392, 1985.

15. Allen, R. W., and Rosenthal, T. J., "Requirements for Vehicle dynamics Simulation Models," SAE paper 940175, February 1994.

16. Freeman, J. S., et al., "The Iowa Driving Simulator: An Implementation and Application Overview," SAE paper 950174, February 1995.

17. "Vehicle Aerodynamics Terminology," SAE Recommended Practice, SAE J1594 JUN87, SAE Handbook, 1991, V.4.

18. NADS Vehicle Dynamics Simulation, Release 4.0, Center for Computer-Aided Design, National Advanced Driving Simulator, The University of Iowa, Iowa City, IA., August 1995.

19. DADS Revision 7.5 Reference Manual Volume 1, CADSI, Computer Aided Design Software, Inc.

20. DADS Revision 7.5 Reference Manual Volume 2, CADSI, Computer Aided Design Software, Inc.

21. Polston, R. K., "The Influence of Vehicle Suspension Modeling on Virtual Prototyping," Master Thesis, The University Of Iowa, Iowa City, IA, 1994.

22. Chrstos, J. P., "Parameter Measurement and Computation for 1994 Ford Taurus GL," DOT-NHTSA Report in Progress.

23. Chrstos, J. P., Grygier, P.A., "Instrumentation and Field Testing of 1994 Ford Taurus GL for NADSdyna Evaluation," DOT-NHTSA Report in Progress.

24. Chrstos, J. P., "Simulation Evaluation of VDANL and VDM RoAD," DOT-NHTSA Report in Progress.

25. Salaani, M. K., "Vehicle Dynamics Modeling for the National Advanced Driving Simulator of a 1994 Ford Taurus GL," DOT-NHTSA Report in Progress.

26. Salaani, M. K, Heydinger, G. J., " Vehicle Dynamics Evaluation for the National Advanced Driving Simulator of a 1994 Ford Taurus GL," DOT-NHTSA Report in Progress.

27. Salaani, M. K., "Development and Validation of a Vehicle Model for the National Advanced Driving Simulator," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1996.

28. Salaani, M. K., Heydinger, G. J., Guenther, D. A., " Validation Results from Using NADSdyna Vehicle Dynamics Simulations," SAE 970565, February 1997.

29. Chrstos, J. P., "A Simplified Method for the Measurement of Composite Suspension Parameters," SAE 910232, February 1991.

30. Users Manual -Low Speed Wind Tunnel, Lockheed Aeronautical Systems Company, January, 1989.

31. Pottinger, M. G., "Flat-Track II Machine helps in Tire Force Moment Measuring," Rubber and Plastic News, March 19, 1990.

32. Garrott, W. R., Monk, M. W., and Chrstos, J. P., "Vertical Inertial parameters- Measured Values and Approximations," SAE 881767, February 1988.

33. Heitzman, E., "Topsy-A Modular Chassis Parameter Measurement System," SAE Paper 971056.

970565

Validation Results from Using NADSdyna Vehicle Dynamics Simulation

Mohamed Kamel Salaani
Transportation Research Center, Inc.

Gary J. Heydinger
SEA Inc.

Dennis A. Guenther
Ohio State Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper presents an evaluation of a vehicle dynamics model intended to be used for the National Advanced Driving Simulator (NADS). Dynamic validation for high performance simulation is not merely a comparison between experimental and simulation plots. It involves strong insight of vehicle's subsystems mechanics, limitations of the mathematical formulations, and experimental predictions. Lateral, longitudinal, and ride dynamics are evaluated using field test data, and analytical diagnostics. The evaluation includes linear and non-linear range of vehicle dynamics response.

INTRODUCTION

The utility of the analysis of the simulation depends on the extent to which it has been validated. The inconsistencies between computer models and actual vehicle response could be due to problems in several areas that include model formulation, simulation programming, vehicle parameter identification, numerical accuracy and stability, and low quality experimental results. The evaluation methodology is to compare simulation results with experimental measurements using the same driver control inputs [1]. This method is compelling and sound, but it should be necessarily applied with some analytical diagnostics and common sense, because the simulation might mask errors that could not be detected from comparisons with experiments.

The dynamic characteristics of road vehicles include complex interactions between many subsystems. Trying to find sources of discrepancy between simulation and experiment by looking at the vehicle as one unit is a hopeless practice. Thus, each subsystem must be evaluated and checked separately. The requirements for the appropriate vehicle dynamics response intended for handling and ride behavior are outlined in the following table:

MANEUVERS :	REQUIREMENTS :
Vehicle linear handling range	Proper non-linear geometry of roll-steer, camber angle, steering kinematics. Requires accurate tire stiffness, suspension compliances, lateral tire dynamics, and proper inertial and geometric data.
Trends in the nonlinear handling range	Accurate suspension kinematics, and advanced tire ground force model.
Accurate predictions in the nonlinear range, up to the limit maneuvers	Measure tire data specific for the particular tire/surface used for maneuvering experiments, accurate nonlinear suspension kinematics, tire data at extreme loads and combined orientations, and nonlinear characteristics of compliant elements.
Ride Analysis	Tire contact model should be appropriate for the particular surface in testing and desired frequency range. Nonlinear suspension spring stiffness, bushings, and damping are required for extreme road disturbances.
Straight Line Braking	Requires appropriate tire data for the particular tire/surface combination, proper inertial and geometric data for longitudinal load transfer and squat/ lift forces characteristics, accurate wheel spin dynamics algorithm, and accurate brake model with the appropriate dynamic lag.
Braking in a Turn	Accurate tire data for a combined lateral and longitudinal slip ratio for the particular tire/surface used, and an advanced tire model has to be used.

Table 1. Vehicle Simulation Guidelines

The vehicle dynamics simulation, NADSdyna, was developed primarily by the University of Iowa [7].

NADSdyna uses a multi-body recursive formulations to solve the dynamics of rigid bodies [2]. Subsystem models for the tires, aerodynamics, steering system, braking systems, and power train are coupled with the recursive dynamics core of NADSdyna. Iowa State University implemented four existing tire models which are available within NADSdyna. Iowa State University also developed models for the wheel spin dynamics for zero longitudinal speed.

The vehicle simulated is a 1994 Ford Taurus. The Taurus model, developed at the Vehicle Research and Test Center (VRTC), is detailed in [15]. Details of the model parameter determination procedures executed by VRTC are given in [13]. All experimental field testing of the Taurus was done by VRTC at the Transportation Research Center, Inc.

VEHICLE FIELD TESTING

The vehicle was instrumented at VRTC, and the schematics of the channels of data are shown in Figure 1. The vehicle field testing was designed to minimize the effects of driver variability and provide repeatable open loop experimentation. For each maneuver, the driver control inputs were repeated, then the mean values and 95% confidence intervals were computed. The mean experimental driver input used to drive the simulation (~250 Hz), is obtained using linear interpolation of experimental sampled values (100 Hz). The following maneuvers are compared with simulation: slowly increasing steer, step steer, pulse Steer, double lane-change, straight line braking, brake-in-a-turn, and ride. Maneuvers involving vehicle acceleration were also performed, but this paper does not include these results. However, the power train model evaluation will be presented at a future date.

Experimental determination of the model parameters result unavoidably in stochastic variability due to some uncontrolled conditions. These parameters are estimated and implemented in the mathematical models that approximate the physics of the desired vehicle dynamic response through numerical simulation. This simulation is compared to field test data which are characterized also by some variabilities due to the undetermined nature of the tribological contact between the tire and the ground, tread wear, tire temperature, changes in control inputs and others. Therefore, the experiments and the measurements have to be designed based on strong understanding of vehicle dynamics and insights of parameters sensitivities. Comparing analytical approximations with experimental measurements is not an easy task that guarantees acceptable results or explanation of discrepancy within a statistical bound, particularly when extreme maneuvering conditions are considered. For limit performance maneuvers, only the trends or relatively close results should be expected, since the mechanics at these extreme conditions (especially tires) is still an open area for the natural philosopher to device a theory.

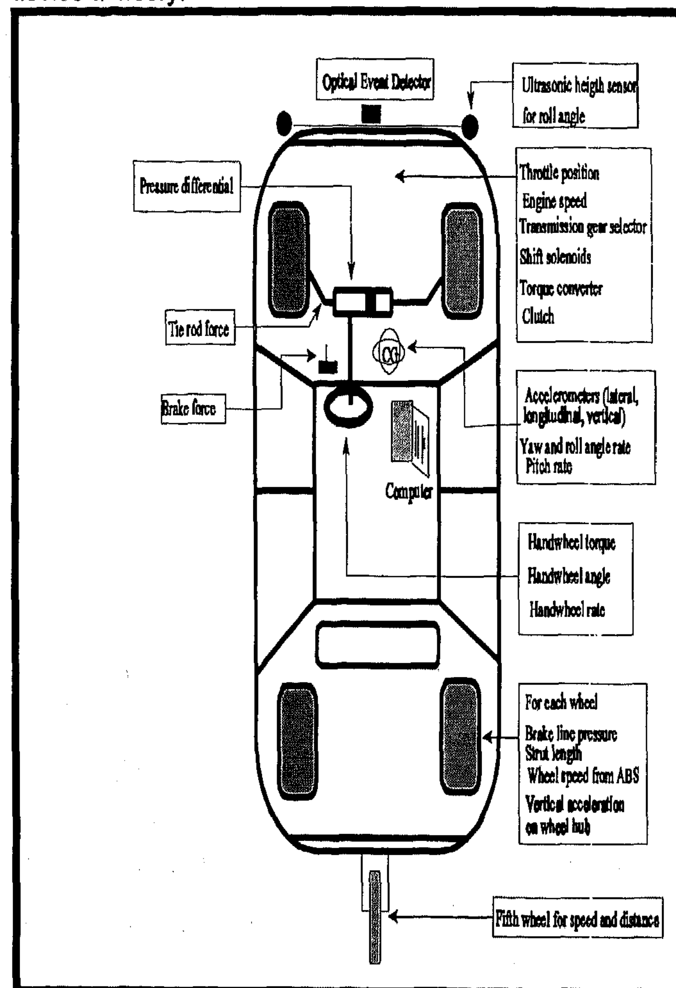


Figure 1 Vehicle Data Channels

VEHICLE DIRECTIONAL DYNAMICS

SLOWLY INCREASING STEER - The slowly increasing steer maneuver involved supplying the simulation with a slowly increasing handwheel angle, while maintaining a constant vehicle speed, until maximum vehicle lateral acceleration, rollover, or a spinout is achieved. A typical rate for handwheel angle that would eliminate dynamic effect and maintain steady state properties is on the order of *10 deg/sec.*

Figures 2-6 compare the simulation with the experimental results for the following variables: steering wheel input, roll angle, yaw rate, and lateral acceleration. These Figures show that the simulation predictions are close to the experiments including the severe cases. From Figure 3, the slope of the roll angle versus lateral acceleration is about *3.9 deg/g*, and the slope of the hand wheel angle versus lateral acceleration is about *235 deg/g*. Steering angle versus lateral acceleration gives an indication of vehicle linear understeer behavior which includes the effects of roll steer, compliance steer, and front to rear tire loading.

At higher g level (Figures 4-6), the front tire force saturation comes into play, and the steady state curve is asymptotic to some limit of lateral g level, which is an indication of limit understeer effect. The handwheel rates are slightly greater than *10 deg/sec* because experimentally the tires get too hot if the duration of this maneuver is long. The maximum lateral acceleration achieved was followed by a reduction, because the tire lateral forces saturated and the front tires 'wash out'. As a result, further increase in the hand wheel angle no longer increases lateral acceleration.

The results of this maneuver indicated that the vehicle model has the appropriate non-linear suspension properties, as well as the tire model is appropriate to model severe cornering. No statistical analysis was performed because the steering input is controlled by the driver which is not repeatable.

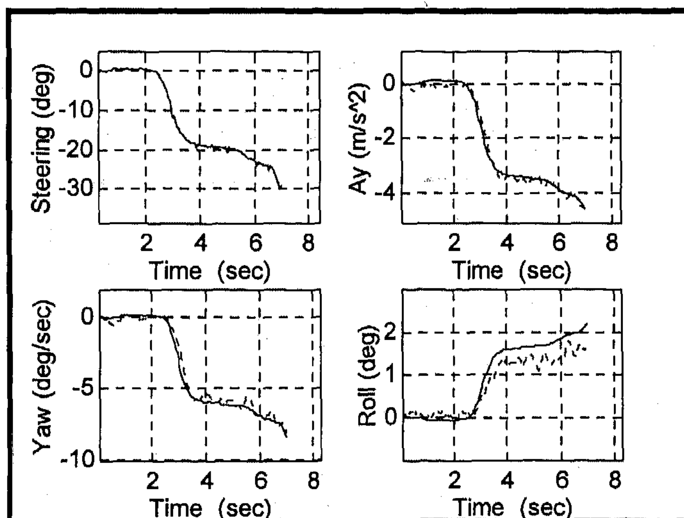


Figure 2. Slowly Increasing Steer at Vx = 32 m/sec
(solid: simulation, dashed: experiment)

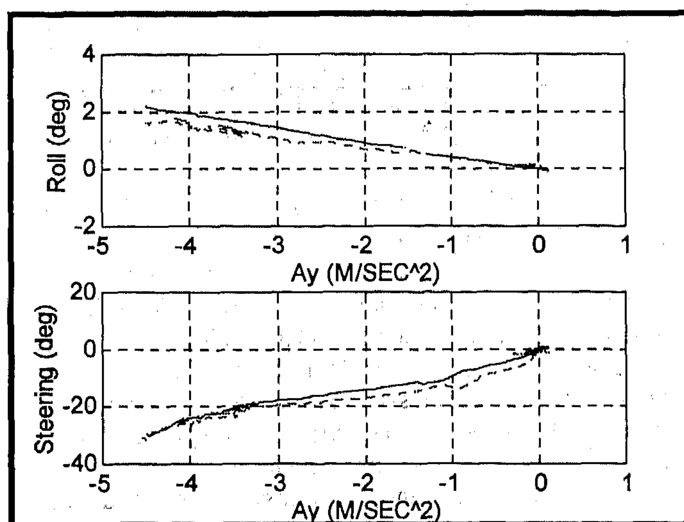


Figure 3. Slowly Increasing Steer at Vx = 32 m/sec
(solid: simulation, dashed: experiment)

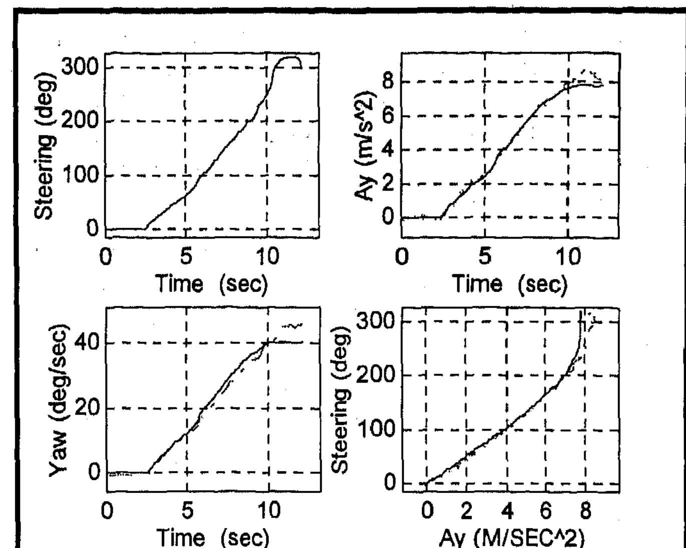


Figure 4. Slowly Increasing Steer at Vx = 11 m/sec
(solid: simulation, dashed: experiment)

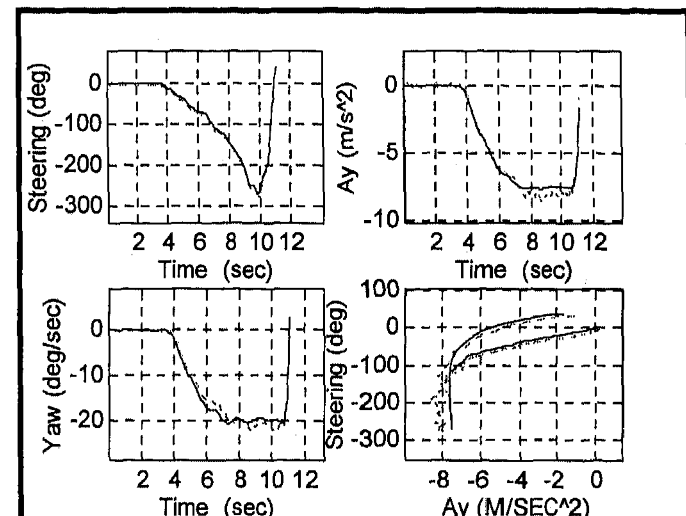


Figure 5. Slowly Increasing Steer at Vx = 22 m/sec
(solid: simulation, dashed: experiment)

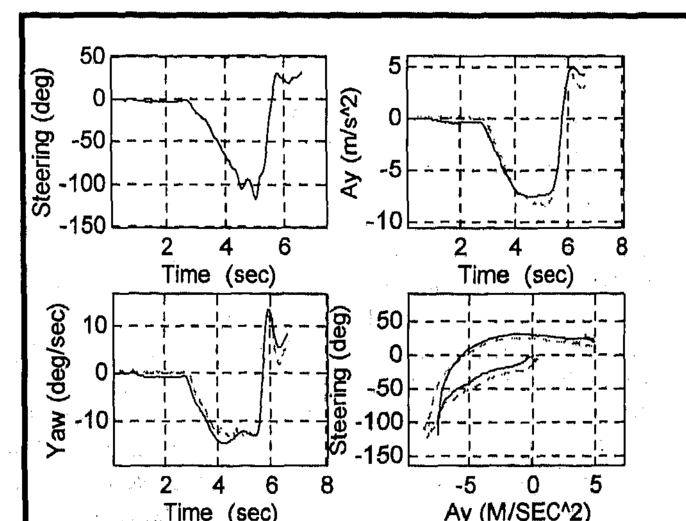


Figure 6. Slowly Increasing Steer at Vx = 32 m/sec
(solid: simulation, dashed: experiment)

STEP STEER INPUT - This maneuver (J-turn) is a pseudo-step input of handwheel angle. The J-turn maneuvers are simulated to determine steady-state as well as transient responses. A physically realistic input rate on the order of *500 deg/sec* is used. A pure step input is not even mathematically feasible for the multi-body formulations applied in this simulation, because kinematics constraints diverge.

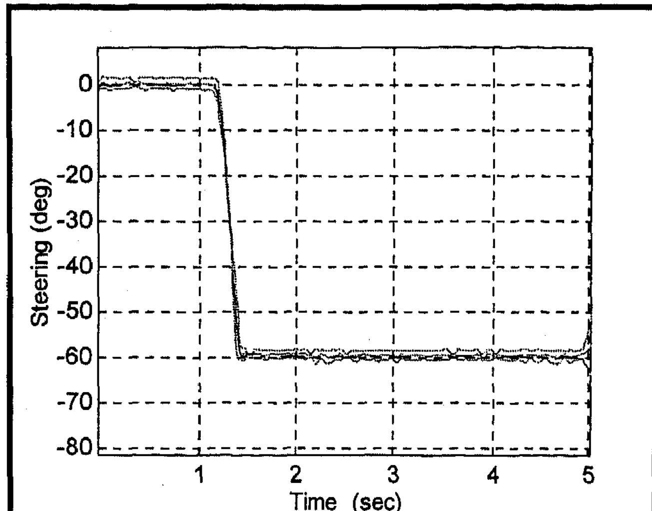


Figure 7. Step Steering Input at Vx = 32 m/sec
(Solid: experiment (95% confidence interval), dashed: Simulation)

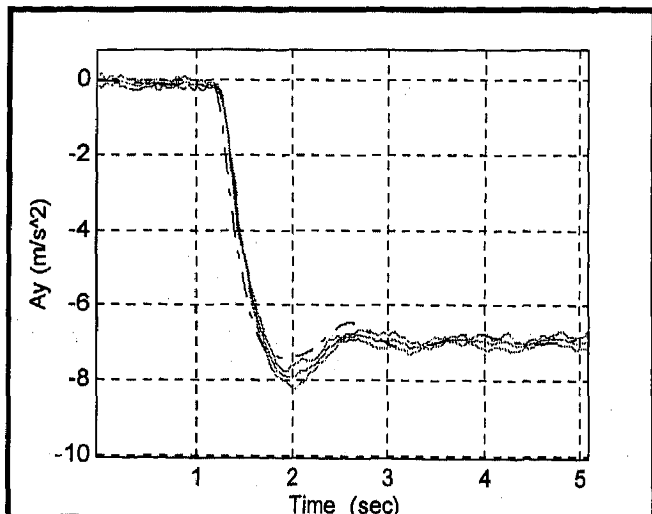


Figure 8. Lateral Acceleration at Vx = 32 m/sec
(Solid: experiment (95% confidence interval), dashed: Simulation)

For the simulated vehicle, the steering system dynamics from handwheel to steering rack are not yet modeled, and the tire force model is only quasi-static. However, lateral force dynamics are included using a first order lateral slip angle differential equation. The longitudinal lag is included in the differential equation of wheel spin rate, and has minimum effect within the frequency span considered in handling analysis. The complexity of the structure and dynamic behavior of the tires are such that no complete and satisfactory theory has yet been propounded. Figures 7-10 show a severe simulated step steering input compared with experimental results. The transient response comparisons between field test and computer model show good agreement in terms of timing of the transient events and peak levels of the variables that indicate limit performance conditions.

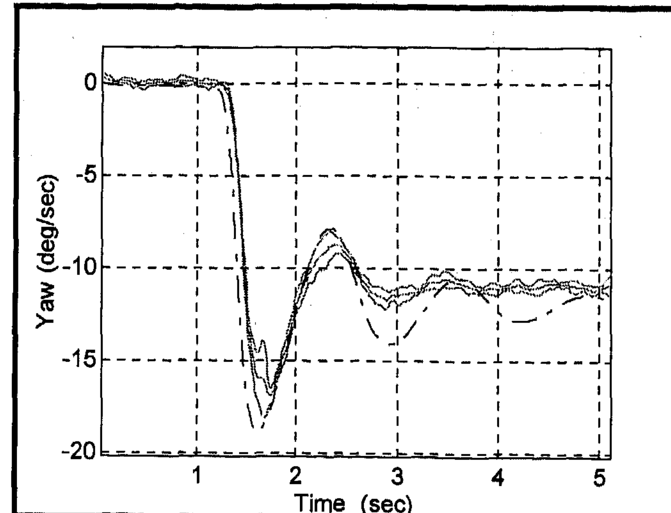


Figure 9. Yaw Rate at Vx = 32 m/sec
(Solid: experiment (95% confidence interval), dashed: Simulation)

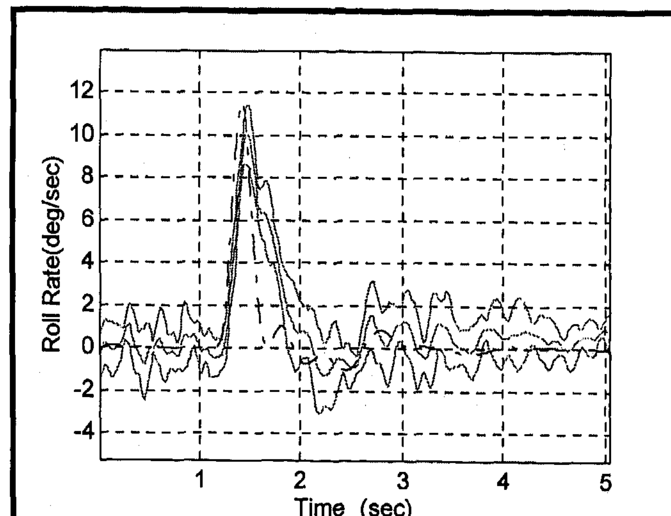


Figure 10. Roll Angle Rate at 32 m/sec
(Solid: experiment (95% confidence interval), dashed: Simulation)

PULSE STEER (DIRECTIONAL DYNAMICS FREQUENCY RESPONSE) - The frequency responses in vehicle dynamics are mostly computed from sinusoidal sweep input, random input, or discrete sinusoidal input testing, and only recently pulse testing was taken in consideration [9]. By judiciously selecting the proper short duration of pulse input to drive a system, it is possible to excite a full range of frequencies using a single experimental test or simulation run.

Linear frequency responses can be developed for linear, time invariant systems, or approximated for a nonlinear system about some linearized operating condition. For vehicle handling dynamics models, linear transfer functions can be derived only for constant speed

maneuvers, since handling responses are speed dependent. Additionally vehicle handling behavior becomes increasingly nonlinear as lateral acceleration level increases, and linear models for passenger vehicles are assumed applicable below lateral acceleration levels of *0.3-0.4* g's. Computer simulations of vehicle handling are typically more linear than actual vehicles because many of the physical non-linearities within the vehicle system are not precisely modeled in the simulation. If the input of the pulse has too little power for a particular frequency range, the frequency response will not be meaningful in that range. Without sufficient power, noise will dominate the signal and produce invalid numbers in the frequency response. An ideal pulse produces power over all frequencies; however, this is not possible in the physical world. In experiments, the pulse was generated by a sequence of step increase and step decrease steer with a total duration on the order of *0.2 sec*, which provides good steering power up to the 3 to 4 Hertz of frequency range [9]. Experimental frequency domain results from one pulse test to the next are considerably different unless the system is nearly linear and the data have little noise. In this case the test results are quite similar, and an ensemble frequency response is found by dividing the sum of the cross-spectral densities of the output channels by the sum of the power spectral densities of the input channels. The error bounds of the frequency responses are computed using the coherence between the averaged channels. This technique produces better results then by computing point-by-point mean of the response function, and provides an effective reduction of any random noise in the individual signals.

The time domain, Figure 11, of steering wheel input, lateral acceleration, yaw rate, and roll angle rate show that the simulation predicts these variables well. The frequency responses of lateral acceleration, yaw rate, and roll rate (Figures 12-14) , show that the simulation predictions are reasonably between the statistical bounds of the mean experimental results. The coherence function is plotted to show the validity of the experimental measurements for the frequency range considered. The closer the coherence to one, the better the experimental data are in terms of noise effect and system linearity. The directional dynamic properties of the simulation are reasonable as indicated by the agreement with the yaw rate frequency response field test data. For the phase angle plots, the experiments showed higher order than the simulation but close values.. This may be explained by the lack of dynamic steering system model between the handwheel and steering rack, lack of a complete dynamic tire model, and other systems that are modeled with lower order. The lateral acceleration frequency response shows reasonable agreement which gives some credibility to the composite lateral/directional dynamics. However, the roll angle measurements have a high noise signal ratio, and the dynamics trends are reasonably predicted. The bandwidth, peak amplitude ratio, and peak frequency of yaw rate, lateral acceleration, and roll angle frequency response to steering input are speed of response type measures. A wider bandwidth indicates the vehicle response characteristics are maintained for a higher frequency input. Yaw rate frequency response to hand-wheel angle gives an indication of the frequency where the vehicle starts becoming less responsive. The graphs confirm that the simulation predicted well these measures.

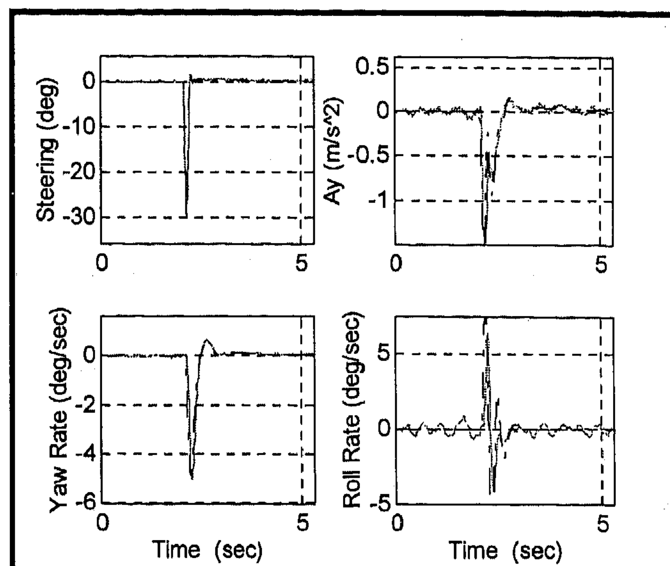


Figure 11. Time Domain Comparison
(Solid: experiment, dashed: Simulation)

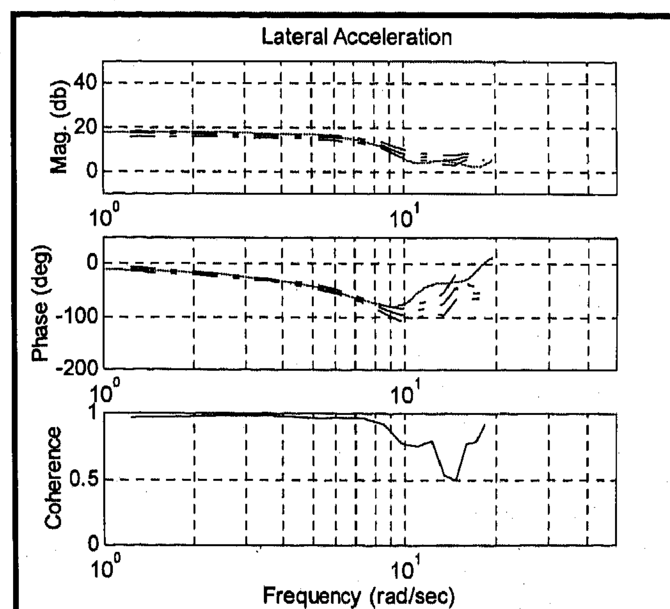


Figure 12. Lateral Acceleration
(Solid: experiment (95% confidence interval), dashed: Simulation)

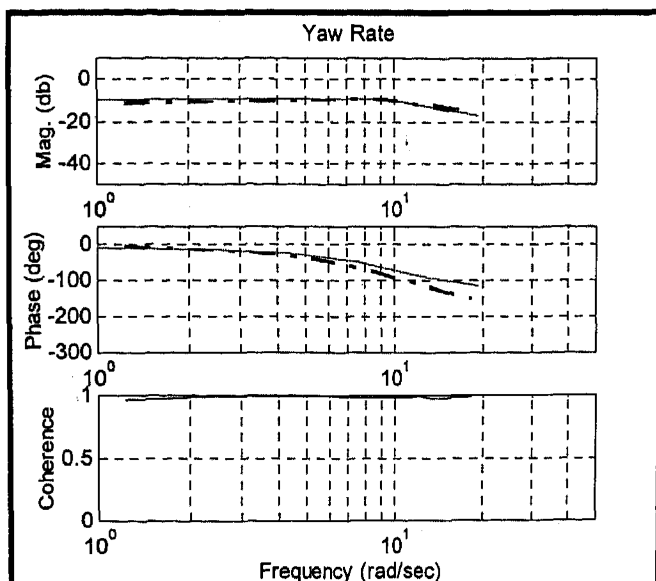


Figure 13. Yaw Rate
(Solid: experiment (95% confidence interval), dashed: Simulation)

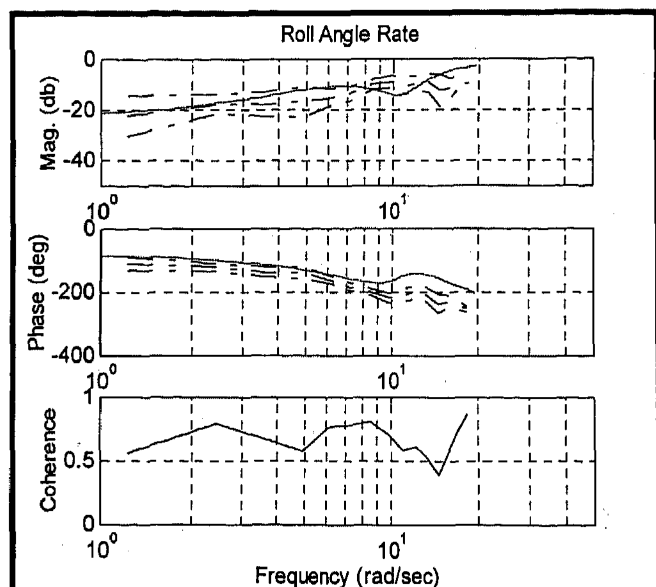


Figure 14. Roll Angle Rate
(Solid: experiment (95% confidence interval), dashed: Simulation)

LANE CHANGE MANEUVERS - Once the steady state, transient, and frequency response predictions of the simulation have been evaluated, the simulation is checked against more complex maneuvers designed to simulate real world driving conditions, like lane change maneuvers. In this maneuver the steering input represents the driver/vehicle closed loop following performance. These lane change maneuvers are applied with sufficient amplitude to produce limit maneuvering performance conditions. The timing of steering profile is such to avoid rollover for the test vehicle. Figures 15-16 show that the simulation predicted well the lateral acceleration, and the yaw rate. No statistical analysis was performed since only one run was performed for each maneuver.

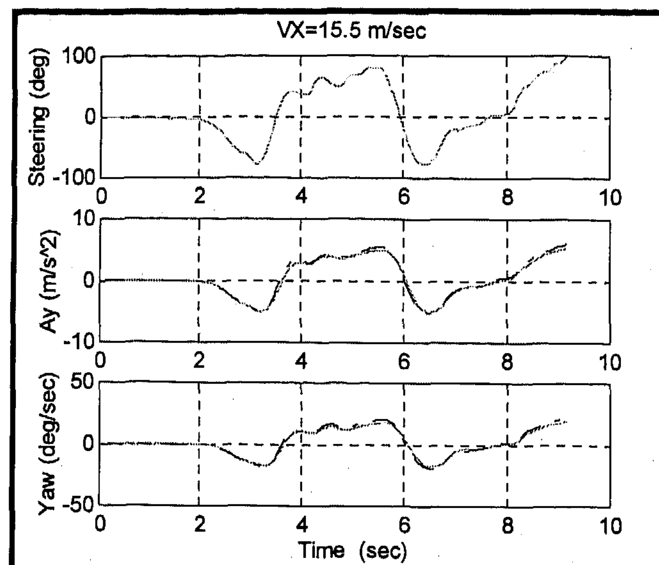


Figure 15. Lane Change Maneuver
(Solid: experiment, dashed: simulation)

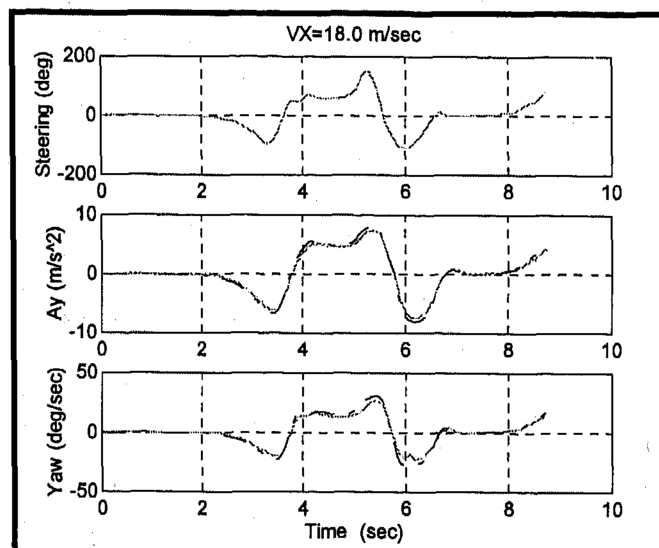


Figure 16. Lane Change Maneuver
(Solid: experiment, dashed: simulation)

VEHICLE BRAKING DYNAMICS

STRAIGHT LINE BRAKING - The braking maneuvers are performed at different levels of severity from low g up to the limit. Despite uncertainty in road friction and other parameters, Figures 17-20 show the simulation predicted reasonably the longitudinal velocity, longitudinal deceleration, pitch angle, and the wheel spin rate. The comparison between the simulated and the measured wheel spin rate shows that the wheel spin dynamics mathematical model is adequate. The brake model is adequate as can be seen from the relation between brake pedal force and brake line pressures. The tire force model is adequate to predict longitudinal braking as can be noticed from the close relation between simulation and experiment of vehicle deceleration.

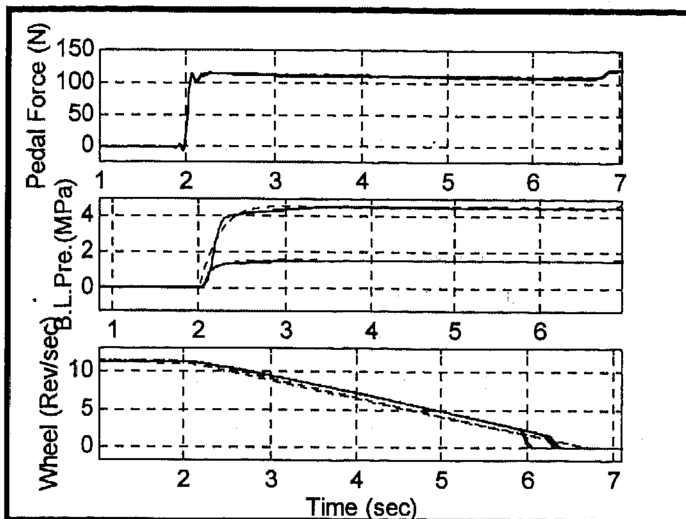


Figure 17. Brake C
(Solid: experiment, dashed: simulation)

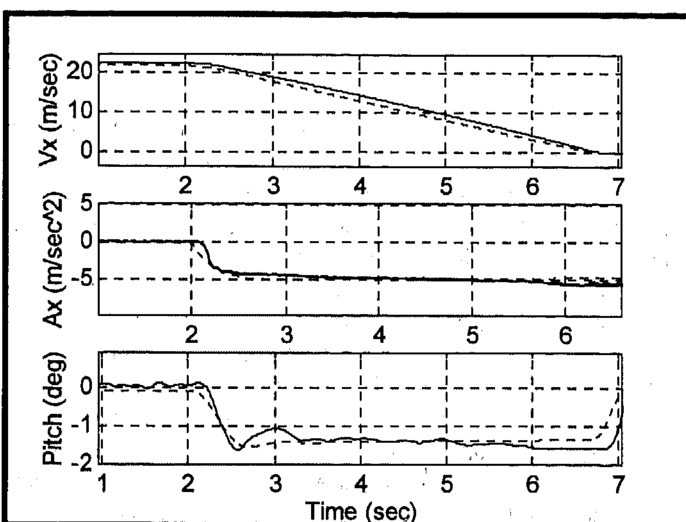


Figure 18. Brake C
(Solid: experiment, dashed: simulation)

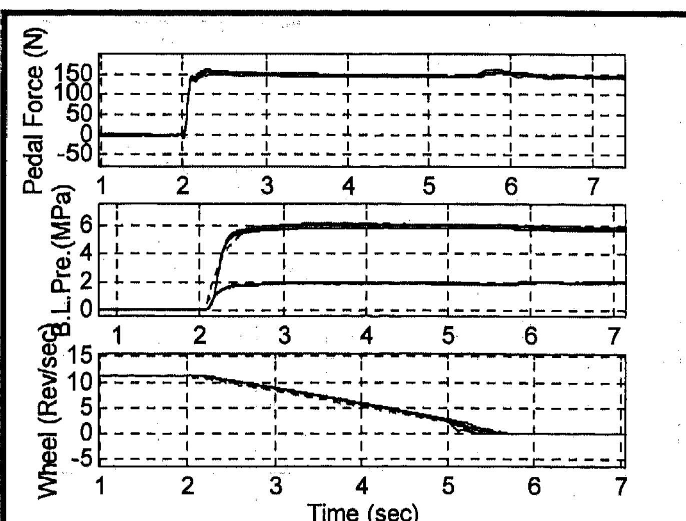


Figure 19. Brake D
(Solid: experiment, dashed: simulation)

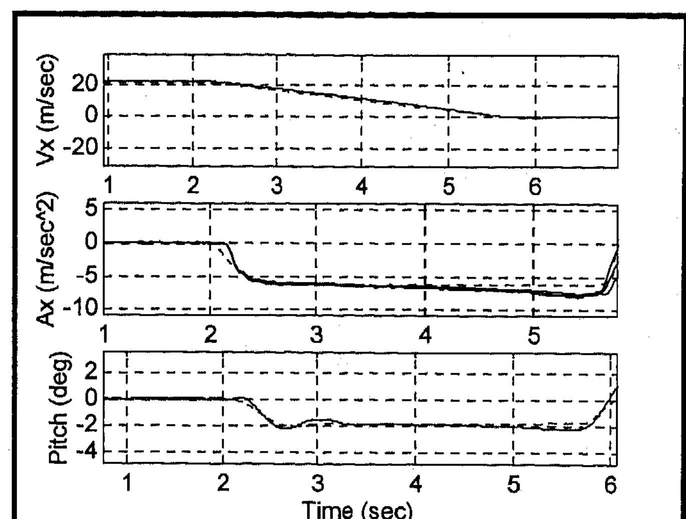


Figure 20. Brake D
(Solid: experiment, dashed: simulation)

BRAKING IN A TURN - Once steering and braking maneuvers are validated, a combined maneuver is not necessarily valid since the tire forces are extremely non-linear. Predictions for this maneuver require accurate tire force model for the particular tire/surface used. Despite the complications of this maneuver, the trends of the simulation predictions are reasonable compared to experiments. Figures 21-22 show a comparison between the simulation and experiment of brake pedal force, brake line pressure, steering wheel input, longitudinal velocity, longitudinal deceleration, lateral acceleration, yaw rate, pitch angle, front left wheel angular velocity, and front right wheel angular velocity. The tire model does not fit well the experimental data at composite slip conditions [15]. The tire model should be improved to better predict the composite longitudinal and lateral forces.

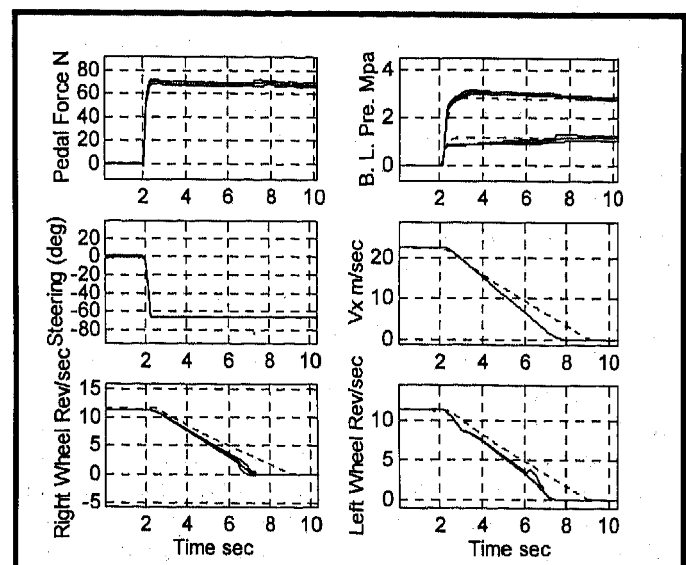


Figure 21. Brake and Steer
(solid: experiment, dashed: simulation)

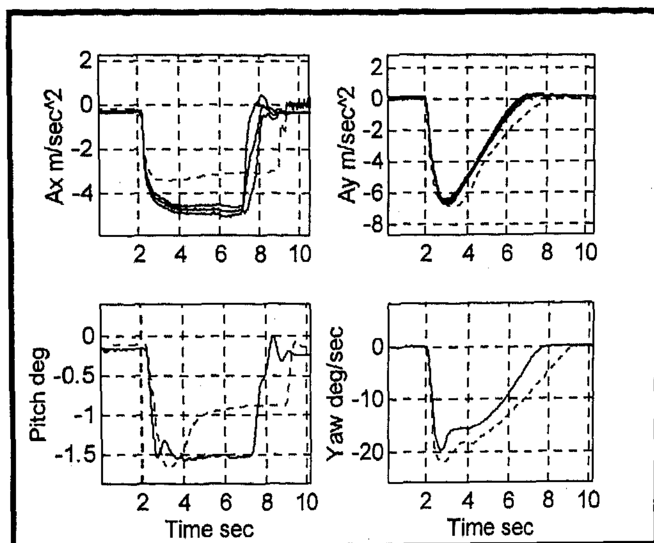


Figure 22. Brake and Steer
(solid: experiment, dashed: simulation)

RIDE ANALYSIS

Ride is a subjective perception, normally associated with the level of comfort experienced when traveling in a vehicle. The vehicle experiences a broad spectrum of vibrations in response to excitation inputs that include road roughness, tire/wheel, drive line, and engine. These vibrations are the source of road feel considered to be an essential feedback to the driver of a motor vehicle, and should not degrade his driving performance.

In this analysis, a simplistic approach is applied where road roughness is considered to be the input of vehicle vibration response. This roughness is described by the elevation profile that excited the bounce and pitch motions along the wheel tracks over which the vehicle passes. The most meaningful measure of ride vibration is the acceleration produced. The attenuation of this high frequency input is an important aspect of the ride isolation behavior obtained via vehicle suspensions. To determine the sprung mass transmissibility of road excitation, we determine the dynamic transfer function between road acceleration input and vehicle vertical acceleration at the center of gravity of the chassis. The acceleration spectrum at the vehicle center of gravity have high amplitude at the sprung mass resonant frequency, moderate attenuation through the resonant frequency of the wheel, and a rapid attenuation thereafter. The experimental transfer function is obtained from the vehicle vertical acceleration and road input acceleration. An analytical function is used to compute road elevation for the simulation. The road elevation function has amplitude and frequency as a linear function of time. The acceleration has the appropriate power spectrum density function (Figure 24), where the amplitude increases as frequency increases to fully excite the higher frequencies of vehicle's systems.

The simulated road input elevation, $h(t)$, is formulated as follows:

$$h(t) = A(t)\sin\left(Bt^2\right) \qquad (1)$$

$$A(t) = A_{max} - \frac{A_{max} - A_{min}}{t_{max}}\,t \; ; B = 2\pi\,\frac{f_{max} - f_{min}}{t_{max}} \qquad (2)$$

where:

A_{max} : Maximum road elevation
A_{min} : Minimum road elevation
t_{max} : Maximum time of simulation
f_{max} : Maximum frequency of road input
f_{min} : Minimum frequency of road input
t : Time

and the road acceleration input is as follows;

$$\ddot{h}(t) = \left(4Bt\dot{A}(t) + 2BA(t)\right)\cos(Bt^2) - 4B^2t^2A(t)\sin(Bt^2)$$

$$\dot{A}(t) = -\frac{A_{max} - A_{min}}{t_{max}} \qquad (3)$$

The measured road elevation is actually an approximation of road roughness. The road used has random bump height and spacing. One stream of vertical road deviation (vehicle left side) is measured and filtered, then differentiated to obtain road input acceleration. The power spectra of the road input acceleration is shown in Figure 23. The ride gain is determined from the chassis vertical acceleration to the road input acceleration. Figure 25 shows a comparison between the simulation and experiment, which is meaningful only up to *25 Hz*, since the experimental data are filtered with a cut-off frequency of *25 Hz*. The natural frequency of the chassis, and wheels, as well as the amplitude attenuation at higher frequencies are reasonably predicted by the simulation.

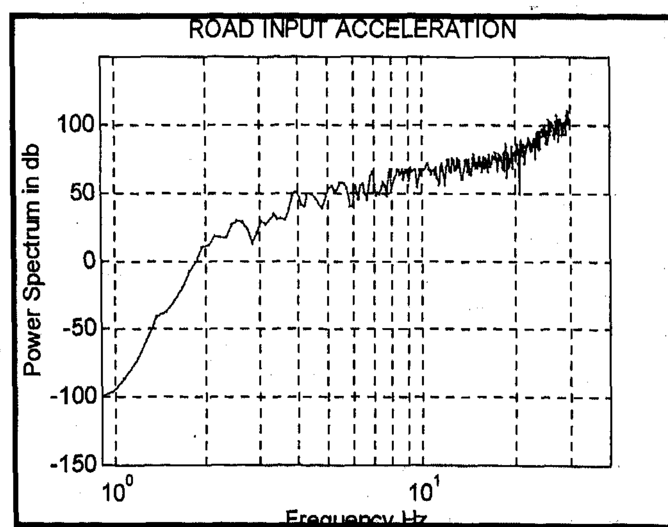


Figure 23. Experimental road disturbance

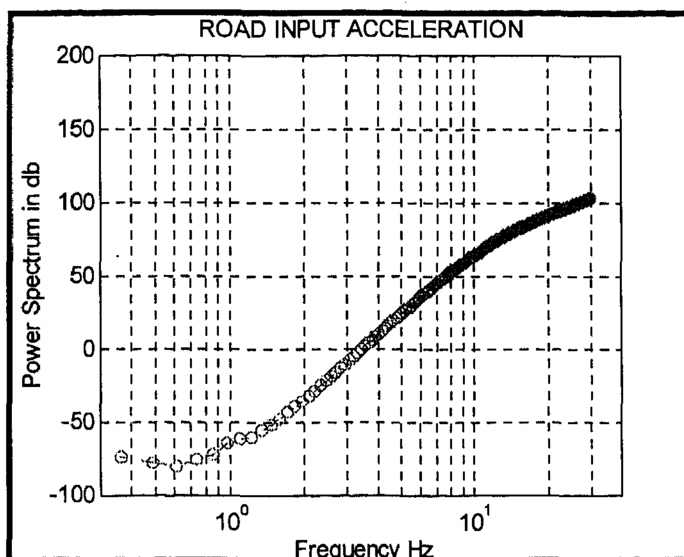


Figure 24. Analytical road disturbance

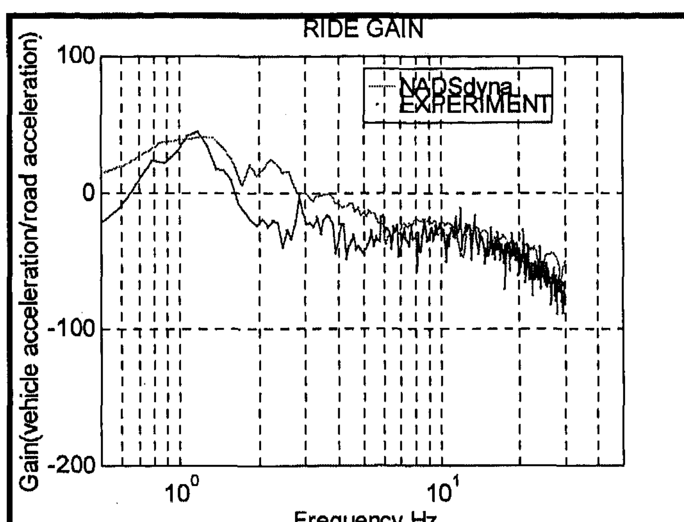


Figure 25. Experimental and simulation ride gain

CONCLUSION

The evaluation of the 1994 Ford Taurus NADSdyna model have shown that the simulation predicted the fundamental mechanics of vehicle handling responses in general. The steady state response is close to experiments for up to the asymptotic limit that characterizes limit understeer effects. The transient response show good agreement in terms of timing of the transient events and peak levels of the variables that indicate limit performance conditions. The directional dynamics properties are in agreement with experiments which give credibility to the composite lateral/direction dynamics. For braking, the longitudinal velocity, longitudinal acceleration, pitch angle, and wheel spin rate are close to experiments. For the combined steering and braking, the trends are reasonably predicted, but an improved tire model should be used. The transfer function of ride analysis shows that natural frequency peaks and attenuation at high frequency are predicted by the simulation.

The results presented show promise that NADSdyna will provide the fidelity required for the NADS vehicle dynamics simulation. In general, the vehicle directional, braking, and ride responses match those measured from the Taurus. This evaluation effort is being extended to include acceleration maneuvers, a complete dynamic steering system, as well as other vehicles including a tractor semi-trailer.

After the off-line evaluation of the model has been completed, the model will be subjectively evaluated once NADS is operational. This will be the most important test of the model, and the NADS must "feel" like an actual car.

REFERENCES

1. Heydinger, G. J., "Improved Simulation and Validation of Road Vehicle Handling Dynamics," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio, 1990.
2. Haug, E. J., et al., "Feasibility Study and Conceptual Design of a National Advanced Driving Simulator," Report No. DOT-HS-807-597, March 1990.
3. Clark, S. K., Mechanics of Pneumatic Tires, National Bureau of Standards Contract CST-386, U.S. Department of Commerce, November 1971.
4. Allen, R. W., et al., "Validation of Ground Vehicle Computer Simulations Developed for Dynamics Stability Analysis," SAE Paper 920054, February 1992.
5. Freeman, J. S., et al., "The Iowa Driving Simulator: An Implementation and Application Overview," SAE Paper 950174, February 1995.
6. Bernard, J. E., and Clover, C. L., "Validation of Computer Simulations of Vehicle Dynamics," SAE Paper 940231, February 1994.
7. NADS Vehicle Dynamics Simulation, Release 4.0, Center for Computer-Aided Design, National Advanced Driving Simulator, The University of Iowa, Iowa City, IA, August 1995.
8. Starkey, J. M., "The Effects of Vehicle Design Parameters on Handling Frequency Response Characteristics," International Journal Of Vehicle Design, Vol. 14, No. 5/6, pp. 497-510, 1993.
9. Heydinger, G. J., Grayier, P.A., Lee S., "Pulse Testing Techniques Applied to Vehicle Handling Dynamics," SAE Paper 930828, February 1993.
10. Chrstos, J. P., "Parameter Measurement and Computation for 1994 Ford Taurus GL," DOT-NHTSA Report in Progress.
11. Chrstos, J. P., Grygier, P. A., "Instrumentation and Field Testing of 1994 Ford Taurus GL for NADSdyna Evaluation," DOT-NHTSA Report in Progress.
12. Chrstos, J. P., "Simulation Evaluation of VDANL and VDM RoAD," DOT-NHTSA Report in Progress.

13. Salaani, M. K., "Vehicle Dynamics Modeling for the National Advanced Driving Simulator of a 1994 Ford Taurus GL," DOT-NHTSA Report in Progress.

14. Salaani, M. K., Heydinger, G. J., "Vehicle Dynamics Evaluation for the National Advanced Driving Simulator of a 1994 Ford Taurus GL," DOT-NHTSA Report in Progress.

15. Salaani, M. K., "Development and Validation of a Vehicle Model for the National Advanced Driving Simulator," Ph.D. Dissertation, The Ohio State University, Columbus, Ohio 1996.

16. Salaani, M. K., Chrstos, J. P., Guenther, D. A., "Parameter Measurement And Development Of A NADSdyna Validation Data Set For A 1994 Ford Taurus," SAE 970564, February 1997.

970569

Development of a Sled-to-Sled Subsystem Side Impact Test Methodology

Jaekoo Chung, John M. Cavanaugh, Matt Mason Jr., and Albert I. King
Wayne State Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A sled-to-sled subsystem side impact test methodology has been developed by using two sleds at the WSU Bioengineering Center in order to simulate a car-to-car side impact, particularly in regards to the door velocity profile. Initially this study concentrated on tailoring door pulse to match the inner door velocity profile from FMVSS 214 full-scale dynamic side impact tests. This test device simulates a pulse quite similar to a typical door velocity of a full size car in a dynamic side impact test. Using the newly developed side impact test device three runs with a SID dummy were performed to study the effects of door padding and spacing in a real side impact situation. This paper describes the test methodology to simulate door intrusion velocity profiles in side impact and discusses SID dummy test results for different padding conditions.

INTRODUCTION

Side impacts have been shown to produce a large portion of both serious and fatal injuries in automotive crashes. A new federal standard for car-to-car broadside collisions, FMVSS 214, was implemented as of September 1, 1996. Occupant protection in side impact has emerged as a most important issue in automotive safety. The new dynamic side impact regulation requires all new passenger cars sold in the United States to pass the new standard specified in FMVSS 214. Intensive research and much progress have been made in side impact safety in the last fifteen years. For side impact research the Heidelberg-type sled test set-up has been widely used. A test subject slides on a long bench seat during sled deceleration, then impacts a stationary side wall [1]*. Eppinger et al. conducted forty-nine side impact cadaveric tests using this type of test set-up and developed a side impact injury criteria, Thoracic Trauma Index (TTI), based on these results [2]. Cavanaugh et al. performed seventeen side impact cadaveric tests using a similar test set-up at the Wayne State University and developed another injury criterion, Average Spine Acceleration (ASA) [3, 4]. The viscous criterion was developed by Viano and Lau, which is utilized in the BIOSID dummy [6]. The controversy regarding acceleration and compression-based criteria has resulted in two different injury criteria in the United States and Europe, and three different side impact dummies, SID, BIOSID, and EUROSID. These issues make it difficult for safety engineers to determine how to design a safer car for side impact.

In a side impact, an initially stationary occupant is impacted by the struck car door. Lau et al. [7] and Deng [8, 9] made a good comparison of frontal and side impact using a spring-mass model and concluded that without simulating key side impact characteristics the results can be misleading. They found that Heidelberg type tests and free flight pendulum tests did not simulate the side impact punch, and may not be valid for the study of real-world side impacts. In the Heidelberg-type sled the test subject has a nominal constant velocity before impacting a stationary wall, while in real-world side impacts, an occupant is impacted by a door velocity pulse. Several dynamic side impact subsystem test procedures have been developed [10,11]. These procedures were based on a similar principle: A moving door, which was initially far away from the test subject, was accelerated to a desired velocity and impacted a stationary test subject. The door velocity was constant or decreased during the impact. Therefore these procedures did not simulate the whole door velocity pulse during contact with occupant, and also were difficult to use to study some basic side impact issues, such as effects of door padding and spacing.

Under the sponsorship of the Centers for Disease Control and Prevention, a sled-to-sled subsystem side impact test methodology was conceived and has been developed at the WSU Bioengineering Center. This test device simulates a pulse quite similar to the typical door velocity profile of a full size car in a dynamic side impact test. This device has been used to study the biomechanical response of side impact dummies, and injury tolerance of human surrogates.

* Numbers in parentheses designate references at end of paper.

Three runs with a SID dummy for different padding conditions, soft, stiff, and no padding, were performed using a simulated door velocity profile of a full size car based on FMVSS 214 full-scale dynamic side impact tests. The objectives of these early stage tests were to simulate side impacts similar to the real-world situation and to study the effects of door padding and spacing in this situation.

SIDE IMPACT MECHANISM

Injury potential of side impact is mainly related to the interaction between a struck car door and occupant. The intruding door velocity-time history, the shape and material property of the door padding and the space between door and occupant can have major effects on injury severity in side impact. A typical velocity time history of the door inner panel can be characterized by a velocity that rapidly increases, then quickly decreases and increases slightly again, then decreases, as shown in Figure 1. The struck car door strikes the stationary occupant directly. The overall door kinematics is influenced very little by the interaction between the door and the test subject and behaves like a high energy, limited-stroke velocity punch during this initial door intrusion. Occupant contact time to door depends on initial door spacing and most interesting events influencing occupant injury occur very early during side impact. In general, the door hits the occupant within 15 - 25 milliseconds and the most important events in occupant injury are over within 50 milliseconds. Hobbs made a good analysis of the sequence of events in side impact [12]. Energy absorbing materials like door padding occupy space and thus advance and prolong occupant contact with the door.

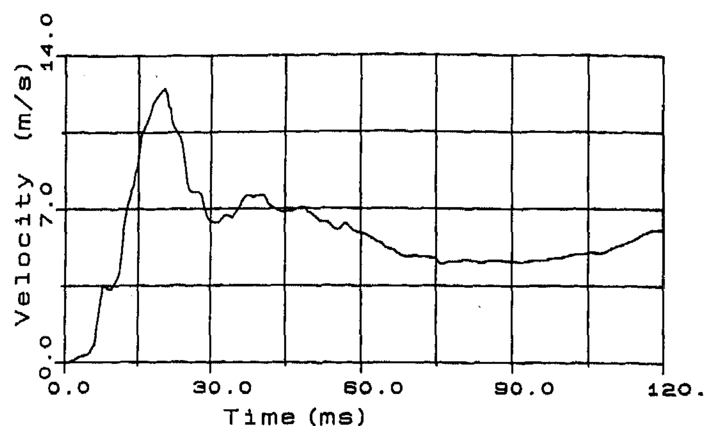


Figure 1. Example of a door velocity profile during side impact [7].

Normally padding does not significantly influence peak spinal acceleration but substantially reduces peak rib accelerations. However, padding can increase the chest deflection due to an increase of energy transfer to the occupant during the impact. These conflicting trends between TTI(d) and chest deflection have already been reported [7, 8].

TEST METHODOLOGY

Initially this study concentrated on tailoring a door pulse to match the inner door velocity profile from an FMVSS 214 full-scale dynamic side impact test. The test device consists of a striking sled and a struck sled with a movable door fixture and a low friction seat covered with Teflon™ sheet. The seating fixture is similar to that of a Heidelberg-type test set-up. The configuration of the WSU sled-to-sled side impact test set-up is shown in Figure 2.

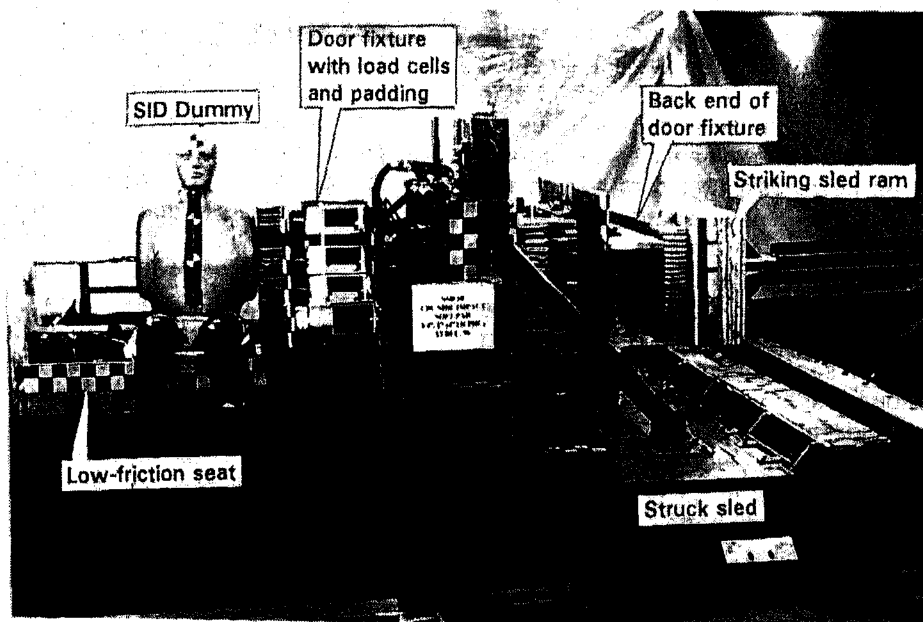


Figure 2. Photo of WSU sled-to-sled side Impact sled test set-up

The sleds run on the WHAM III track (Wayne Horizontal Accelerator Mechanism), a versatile experimental research facility for studying all types of collision environments. Each sled measures 2.0 m wide by 3.66 m long and is accelerated on the 40 m track. The striking sled is accelerated up to a desired velocity by a pneumatic propulsion device and then strikes the simulated door mounted on the struck sled and the struck sled itself at the same time. A hydraulic stopper system was implemented in the door system in order to prevent door rebound motion after maximum door intrusion. The door velocity profile can be tuned by using different dimensions of energy absorbing stiff paper honeycombs placed at three interfaces: (1) the door system, (2) striking-to-struck sled interface, and (3) impacting ram mounted on the striking sled, which impacts the struck sled door system. Five load plates instrumented with nine load cells were mounted on the door wall to measure interaction forces between test subject and door wall surface at each body region: shoulder, thorax, abdomen, pelvis, and knee. Two load cells were used at each load plate except for one at the knee plate. In the padded runs each load plate was covered with a piece of 4 inch thick paper honeycomb pad. The contact forces for each body region give very useful information for test subject responses during impact. Details of the load plates on the door wall are shown in Figure 3. With the SID dummy the shoulder and thorax load plates hit the rib cage as shown in Figure 4.

The first version of the test device was used in eight SID dummy and two cadavers tests. However, the SID TTI(d) and Average Spine Acceleration (ASA) were high and cadaver thoracic injuries were also severe in all test conditions.

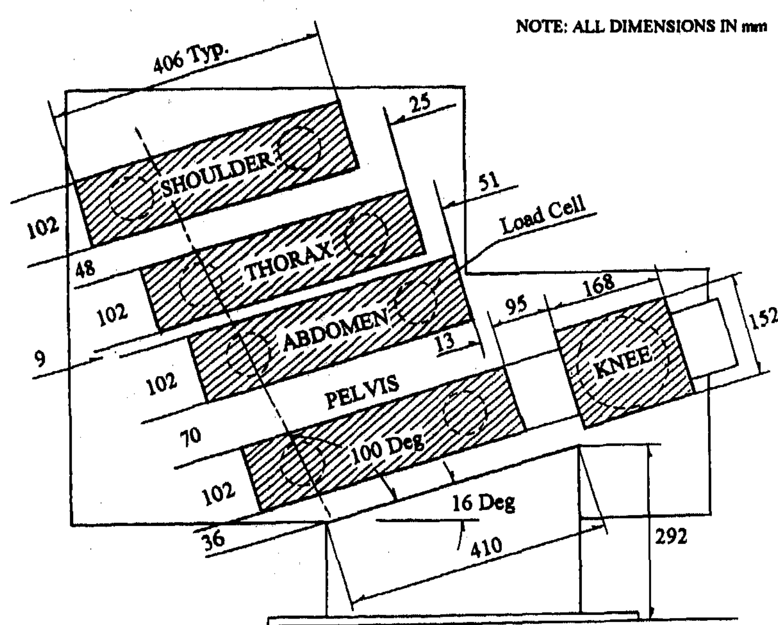


Figure 3. Diagram of impacting door wall showing load plates at each body region instrumented with nine load cells.

The 1st peak door velocity of previous tests was not high (10-11 m/s) but the peak duration time was somewhat long. We modified the fixture to obtain a new door velocity profile with the 1st peak door velocity close to the previous one but with a shorter duration. Newly simulated door velocity profiles look quite similar to a full size car inner door panel velocity from FMVSS 214 full-scale test. The weight of the simulated door wall assembly including all instrumentation is about 800 lbs. It can be seen that door velocity did not vary much with different door padding conditions as shown in Figure 6a at the end of this paper.

EFFECTS OF DOOR PADDING AND SPACING

It can be argued that energy dissipation in door padding can reduce injury severity in side impact. On the other hand, door padding can increase the amount of energy transferred to the occupant during impact and result in increased chest deflection. The spacing hypothesis suggests that increasing space between the inner door panel and the occupant will reduce injury due to the door velocity characteristics because the occupant can be spaced to avoid impact with the door during the high initial peak of door velocity. Intensive research into padding performance with various types of padding has been done at the WSU Bioengineering Center [4, 5]. Different properties of paper honeycombs (PHC) and closed cell foams were used in this research to determine the effect of crush force on dummy response and cadaver injury. Three different padding conditions were run in an initial evaluation of the newly modified sled-to-sled test device: (1) 4 inch thick soft pad, no space, (2) 4 inch thick stiff pad, no space, and (3) no pad, 4 inch space.

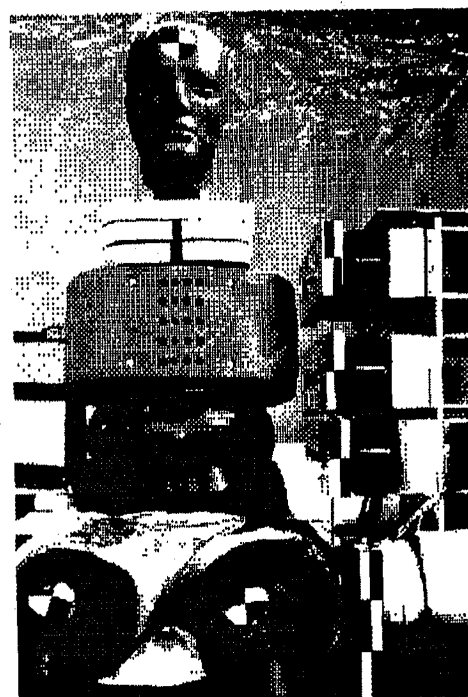


Figure 4. Photo of SID dummy (without jacket) adjacent to the impact door wall.

The crush strengths of the PHC used in these current tests are summarized in Table 1. For soft and stiff padding, two different combinations of PHC were used as shown in Table 2. The 4 inch thick PHC was precrushed up to 1/2 inch in order to get rid of initial force peaks. The PHC was then attached to the face of each load plate of the door wall. Figures 5a and 5b show 4 inch soft PHC before and after impact.

Table 1. Door padding properties

Paper honeycomb cell sizes (inches)	Tested crush strength (psi)
1	10.1
3/4	20.6
5/8	34.9

* Note: Padding crush strength at 35% of pad thickness

Table 2. Configuration of door padding combination

Body region	Soft padding cell size (inches)	Stiff padding cell size (inches)
Shoulder	3/4	5/8
Thorax	1	3/4
Abdomen	1	3/4
Pelvis	3/4	5/8
Knee	3/4	5/8

* Note: Padding size is 4 inches thick and 4 inches wide by 16 inches long for each load plate except for knee padding.

SID DUMMY TEST RESULTS

Comparisons of tests for three different padding conditions were made. A summary of parameters for three SID dummy tests is presented in Table 3. The striking sled velocity was measured with a velocity pick-up sensor and the door velocity was obtained from the 1st integration of acceleration measured at the moving door wall. The door intrusion was calculated by the difference between the struck sled displacement and the door displacement.

The comparison plots of door velocity and SID dummy acceleration-time histories are shown in Figure 6. It can be seen that the door velocity profiles for three tests were quite similar to each other (Figure 6a). The peak values are less than in the full-scale test shown for comparison. However, it was felt that since our door was rigid rather than deformable, lower peak velocities were justified in the initial studies. The peak accelerations to be used for TTI(d) evaluation are summarized in Table 4 and comparison of TTI(d) and ASA are presented in Table 5. The comparison plots of SID dummy chest pot displacement and contact force-time histories at each body region are shown in Figure 7 and the peak contact forces are summarized in Table 6. The contact force was computed by summing the load cell responses of the two load cells at each body region and correcting for inertia forces due to the added mass of load cell and load plate. Raw data of accelerations were filtered with FIR100 software and contact forces and chest pot deflections with a digital Butterworth filter at 300 Hz.

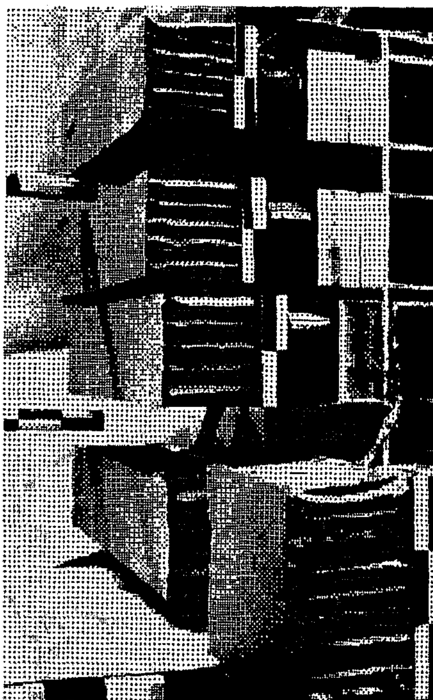


Figure 5a. Four inch thick soft paper honeycomb before impact.



Figure 5b. Four inch thick soft paper honeycomb after impact.

Table 3. SID dummy tests parameters

Padding type	Striking sled velocity (m/s)	Door peak velocity (m/s)	Max. door intrusion (in)
Soft	11.3	10.7	6.2
Stiff	11.1	10.4	6.1
No pad	11.1	10.3	6.8

Table 4. Comparison of SID dummy acceleration

Padding type	Upper rib-y (g's)	Lower rib-y (g's)	T1-Y (g's)	T12-Y (g's)	Pelvis (g's)
Soft	59.6	52.0	41.1	21.7	37.5
Stiff	67.2	59.0	40.0	55.5	50.2
No pad	108.2	93.6	49.2	53.8	54.7

Table 5. Comparison of SID dummy TTI(d) and ASA

Padding type	TTI(d) (g's)	Spine acceleration (g's)		
		ASA10	ASA15	ASA20
Soft	40.7	16.9	17.1	16.9
Stiff	61.4	38.1	38.4	37.6
No pad	81.0	29.0	36.5	43.0

Table 6. Comparison of SID dummy force

Padding type	Shoulder (N)	Thorax (N)	Abdomen (N)	Pelvis (N)	Knee (N)
Soft	4,708	4,355	2,030	6,063	4,206
Stiff	5,054	4,741	2,905	7,991	5,486
No pad	4,147	7,243	779	13,041	881

DISCUSSION

Door Velocity and Intrusion - Simulated door peak velocity and intrusion were limited to peak values within 11 m/sec and 7 inches, respectively, to prevent the possibility of test subject damage. In view of overall shape of velocity profile a quite similar door velocity-time history to one of a full size car was simulated (Figure 6a). A good correlation of velocity profiles can be seen within 30 milliseconds after starting to move the door wall assembly. The discrepancy after 30 milliseconds may have little effect on test subject impact responses for this basic study because most important events occur earlier in contact time. However better engagement of the door wall assembly to the struck sled after maximum door wall intrusion should improve this discrepancy.

SID Dummy Acceleration - Different padding conditions mainly effected rib cage acceleration rather than spine acceleration. TTI(d) is lowest for soft padding, intermediate for stiff padding and highest for no padding. ASA is also lowest for soft padding and more than double this for stiff and no padding. The ASA data suggested that stiff padding may be worse than no padding, while TTI(d) indicated the opposite.

Contact Force - Benefits of padding as force-limiting material can be noticed particularly in the thorax and pelvic regions (Table 6). The second peak force for soft and stiff padding was generated due to padding bottoming-out (Figure 7b, c). These were much less than in the unpadded case. Interestingly, the peak force of abdomen and knee region for no pad were quite lower than for padding conditions.

Chest Pot Displacement - Peak values of chest potentiometer displacement were 29 - 30 mm in all three tests.

CONCLUSIONS

1. A sled-to-sled subsystem side impact test methodology has been developed and used to simulate desirable door velocity profiles for the fundamental study of biomechanical response and injury mechanisms in side impact.

2. The simulated door velocity profile is repeatable and was not effected by the choice of padding or door spacing.

3. Based on the SID dummy tests, this device shows promise in demonstrating door padding and spacing effects that can occur in a real-world side impact.

4. The contact force data was useful in getting a better understanding of test subject response during impact. The soft padding was the best force limiting device at the thoracic level in this initial study.

FUTURE STUDIES

This paper describes the SID dummy test results with three trial runs. The door peak velocity can be increased up to 12 m/sec because SID dummy response was still below the 85 g's cut-off in all three test conditions. More runs with SID, BIOSID, and cadavers will be performed to analyze padding and spacing effects. Tests will also be run with and without the shoulder load plate to test the effect of door sill height on occupant response. It is hoped that this new device will clarify conflicting issues in side impact crash protection and help determined the best injury criteria for side impact safety.

ACKNOWLEDGMENTS

This work is supported by the Centers for Decease Control Grant No. CCR 502347. The authors wish to thank Tim Walilko for help with instrumentation and Richard Vogel from the Bioengineering Center. We also thank Dr. York Huang of First Technology Safety Systems Corporation for his advice of this test methodology.

REFERENCES

1. Marcus J.H., Morgan M.M., Eppinger R.H., et al: Human Response to and Injury from Lateral Impact. SAE Paper No. 831634, 27th Stapp Car Crash Conference. 1983.

2. Eppinger R.H., Marcus J.H., Morgan M.M: Development of Dummy and Injury Index for NHTSA's Thoracic Side Impact Protection Research Program. SAE Paper No. 840885, 28th Stapp Car Crash Conference. 1984.

3. Cavanaugh J.M., Zhu Y., Huang Y., King A.I.: Injury and Response of the Thorax in Side Impact Cadaveric Test. SAE Paper No. 933127, 37th Stapp Car Crash Conference. 1993.

4. Cavanaugh J.M., Walilko T., Malhotra A., Zhu Y., King A.I.: Biomechanical Response and Injury Tolerance of the Thorax in Twelve Sled Side Impacts. SAE Paper No. 902307, 34th Stapp Car Crash Conference. 1990.

5. Cavanaugh J.M., Zhu Y., Huang Y., King A.I.: Performance and Mechanical Properties of Various Padding Materials Used in Cadaveric Side Impact Tests. SAE Paper No. 920354, 36th Stapp Car Crash Conference. 1992.

6. Viano D.C., Lau I.V.: Thoracic Impact: a Viscous Tolerance Criterion. ESV Paper pp. 104-114, 10th ESV Conference, 1985.

7. Lau I.V., Capp J.P., Obermeyer J.A.: A comparison of Frontal and Side Impact: Crash Dynamics, Countermeasures and Subsystem Tests. SAE Paper No. 912896, 35th Stapp Car Crash Conference. 1991.

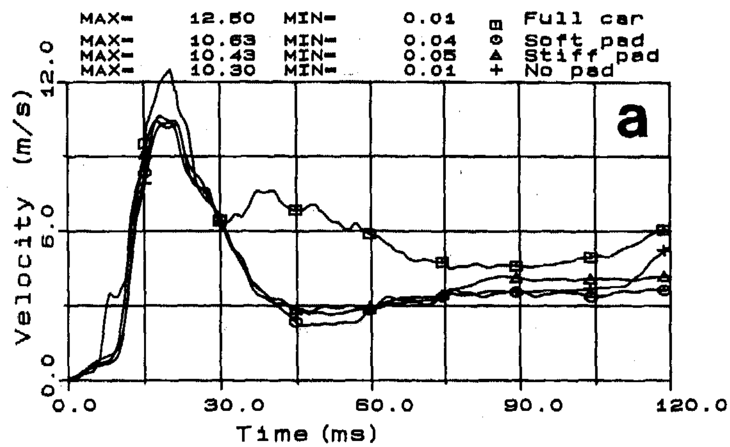
8. Deng Y.C.: Design Considerations for Occupant Protection in Side Impact - a Modeling Approach. SAE Paper No. 881713, 32nd Stapp Car Crash Conference. 1988.

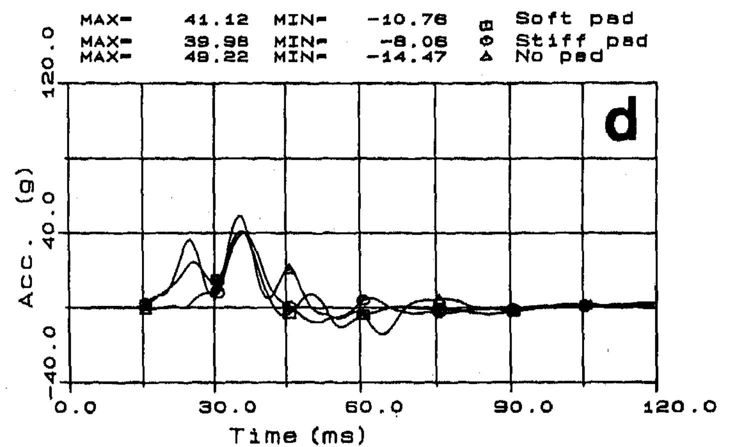
9. Deng Y.C.: The Importance of the Test Method in Determining the Effects of Door Padding in Side Impact. SAE Paper No. 892429, 33rd Stapp Car Crash Conference. 1989.

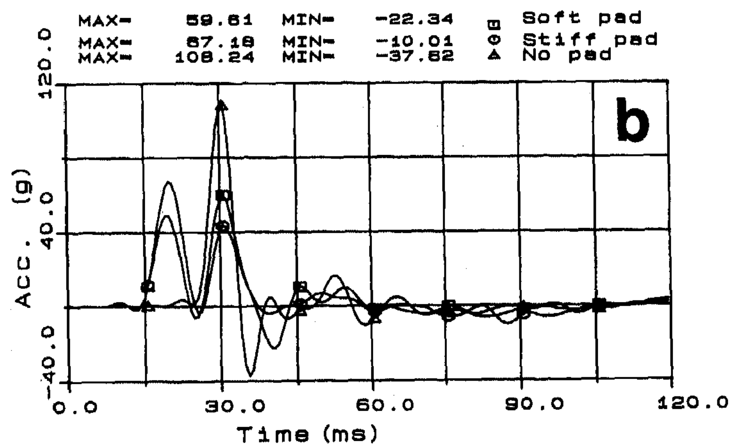
10. Haland Y., Pipkorn B.: The Protective Effect of Airbags and Padding in Side Impacts - Evaluation by a New Subsystem Test Method. ESV Paper No. S5-0-20, 13th ESV Conference, 1991.

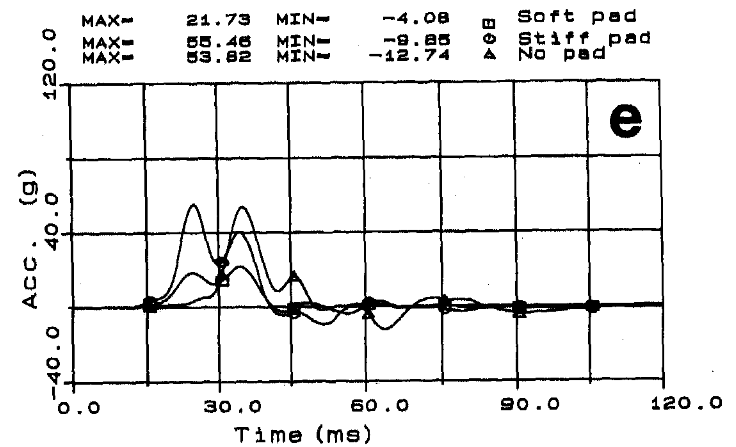
11. Lindquist M.: A simple Side Impact Test Method for Evaluating Vehicle Paddings and Side Structures. ESV Paper No. S5-0-18, 13th ESV Conference, 1991.

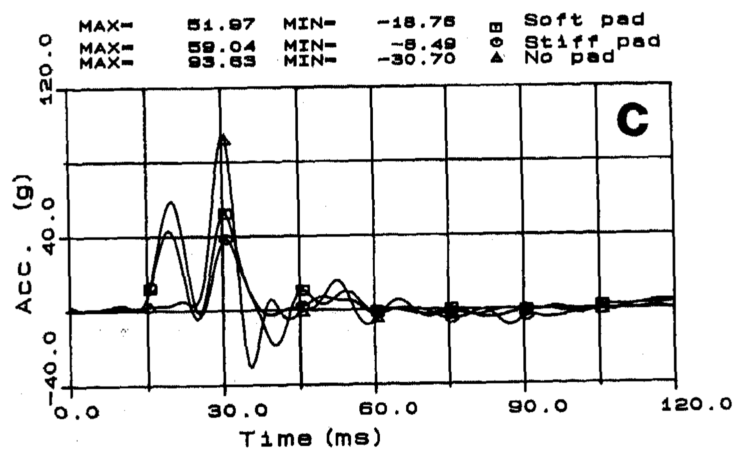
12. Hobbs C. A.: Dispelling the Misconceptions about Side Impact Protection. SAE Paper No.950879, 39th Stapp Car Crash Conference. 1995.











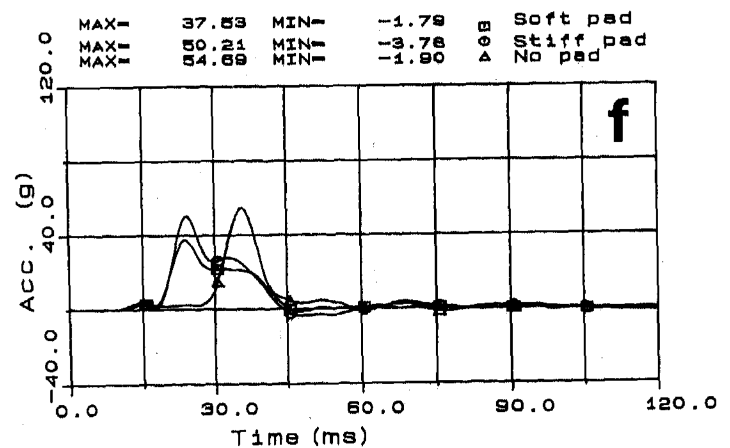
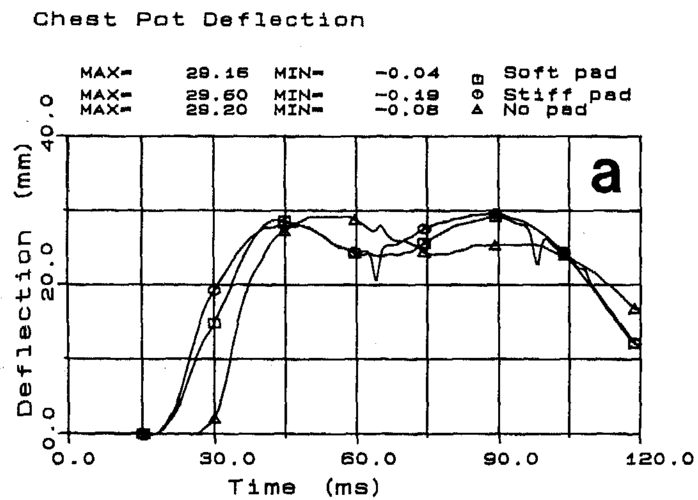
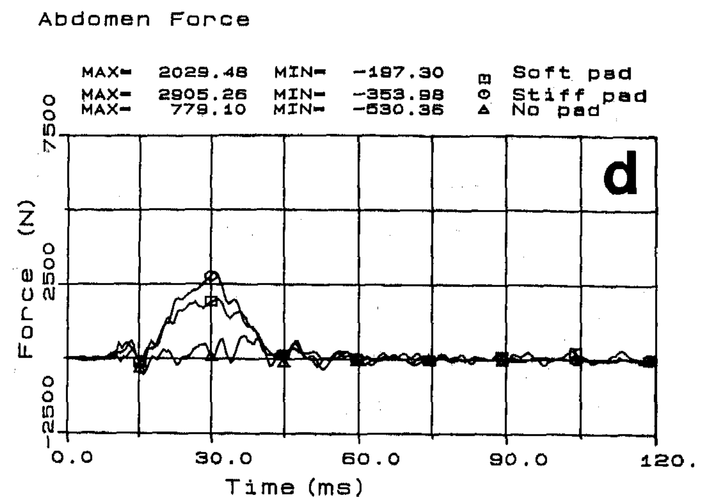
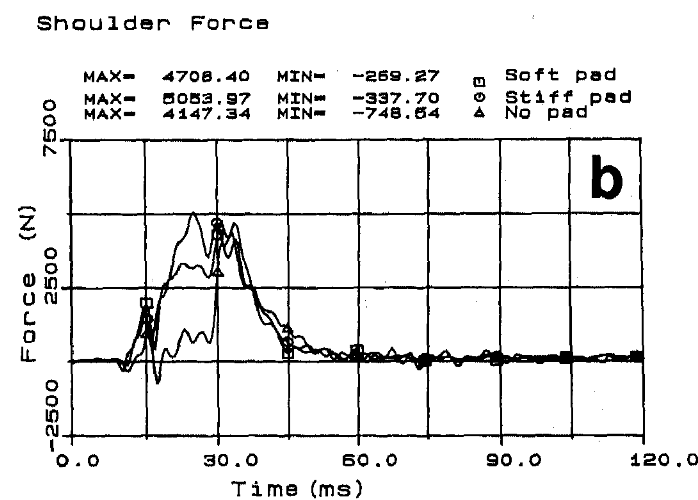
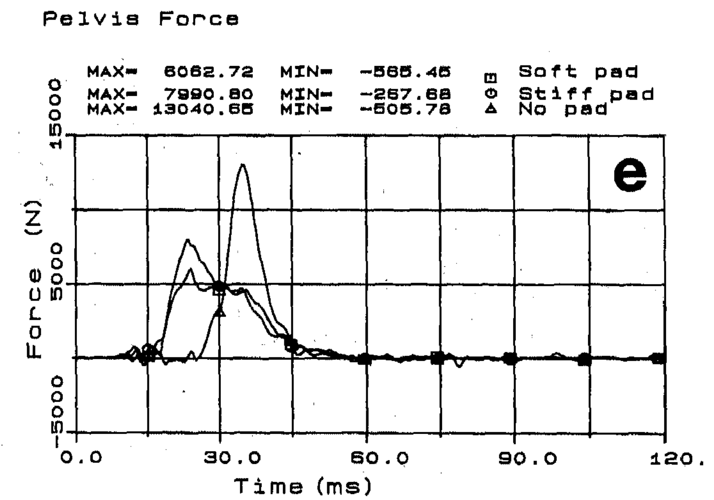


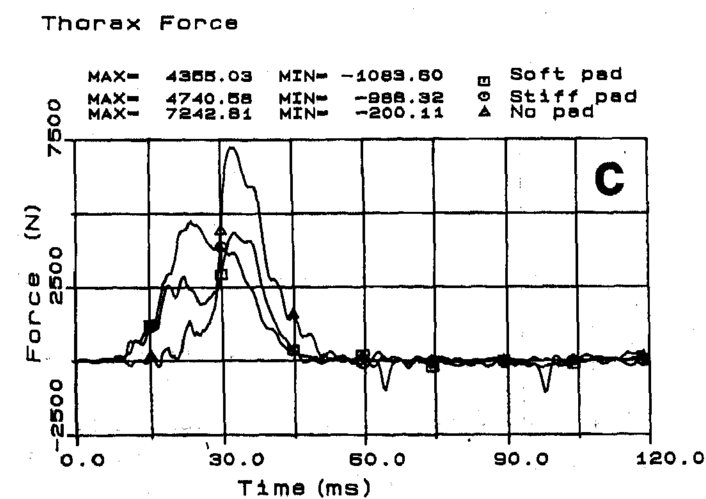
Figure 6. Comparison plots of door wall velocity (filtered with a digital Butterworth filter at 300 Hz) and SID dummy ribs and spine acceleration-time histories (filtered with FIR100 software).











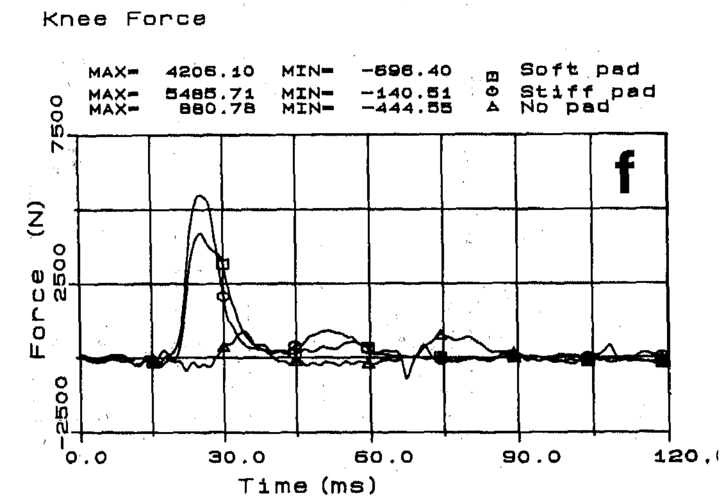


Figure 7. Comparison plots of SID dummy chest pot deflection and contact force-time histories (filtered with a digital Butterworth filter at 300 Hz).

970570

Sled Testing Procedure for Side Impact Airbag Development

P. Michael Miller II and Hai Gu
MGA Research Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Side impact crashworthiness development presents a unique challenge to auto safety engineers. One fundamental issue is how to evaluate side impact air bags with a component test that realistically simulates the kinematics of a full scale side impact crash test. This paper presents a test methodology that can be used to evaluate side impact air bags utilizing an accelerator-type sled typically used for frontal impact simulation. The approach uses a "two carriage" system, whereas the struck door and vehicle acceleration profiles are simulated. These acceleration responses are matched through a series of sled variables including thrust column setting, metering pin shape and an on-board pneumatic cylinder which controls the relative response between the two carriages. The mechanical operating principles behind this testing methodology, the approach used to conduct baseline matching tests, data showing the level of correlation achieved with this testing technique, and side impact air bag test results is presented. Correlation of this type of testing is based upon both the vehicle response and the acceleration responses of a Side Impact Dummy. Readers of this paper will gain a greater understanding of how accelerator-type sleds can be used to realistically simulate dynamic side impact conditions.

INTRODUCTION

Side impact crashworthiness presents a complex problem with regard to development due to the lack of commonly accepted component test methodologies. The automotive industry has been seeking effective methods for side impact testing of their vehicles and components. A system utilizing an accelerator type sled has been developed for side impact crash test simulation. This methodology is refered to as the Side Impact Test and Evaluation (SITE) sled technique (patent pending). The SITE sled system simulates the dummy and door interaction witnessed in full scale crash tests. Exact representation of the vehicle interior and dummy positioning is achieved with this system. The SITE test procedure has been developed as a basic methodology for side airbag development. The SITE procedure has many of the same advantages typically associated with accelerator sled testing including low test to test variability, ease of setup, high correlation with crash test results and realistic deployment of air bags.

As is the case with most simulation approaches, it is important to first establish a "baseline" test case. For the SITE system, the baseline case is a test condition where the SITE results correlate very closely with full scale side impact crash test results. Correlation is based on a number of test variables, including the door acceleration and Side Impact Dummy (SID) lateral accelerations. Once a baseline condition has been established, this test condition can be used to evaluate design changes, or more specifically, the effectiveness of a side impact air bag. The main challenge in the development of the SITE system has been in converting an accelerator-type sled system typically used for frontal impact simulation to be used for side impact simulation. When comparing the kinematics of a frontal impact case to the side impact crash test case, it is obvious that the vehicle accelerations produced in a side impact will be much greater than the frontal case due to the limited crush space for a side collision. Essentially, the closeness of the occupant to the impact area and crush zone is why side impact occupant protection is so difficult. In order to produce the high accelerations of a struck door, the basic operating conditions must be altered. There are a number of approaches that can be implemented, examples of which include an actual carriage to carriage impact, the use of a crushable material to control the response between two carriages, or increasing the distance between the occupant and the door panel, thus allowing the main sled carriage to reach the desired door velocity prior to contacting the occupant.

The approach used with the SITE system is to reduce the mass of the upper carriage so that higher accelerations may be achieved, and controlling these accelerations through the optimal design of a metering pin. In conjunction, the upper carriage (which simulates the door response) is connected to the lower carriage with a pneumatic cylinder. Likewise, the relative response between the upper and lower

carriage is controlled via the onboard pneumatic cylinder. The lower carriage response mimics the unstruck side of the vehicle. Thus, with the SITE system, both the intruding door and floor pan acceleration can be simulated at the same time. This results in a test which accurately simulates what occurs during a side impact crash test.

For side impact air bag development, an accurate side impact sled simulation is an essential tool to control both cost and time expenditures. The SITE system is one option which can be used for evaluation of side impact air bags. Subsequent discussion include an overview of the FMVSS 214 crash test, operating principals and SITE setup, baseline correlation testing, and air bag testing.

FMVSS 214 CRASH TEST

It is assumed that the reader is familiar with the basic requirements of the FMVSS 214 crash test; however, specific details are available in the literature (1, 2, 3). Here, only a very general review is presented in order to set the stage for for the rest of the paper.

The test configuration, as specified by the National Highway Traffic Safety Administration (NHTSA) is shown in Figure 1. Schematically, a Moving Deformable Barrier (MDB) is shown impacting the side of a stationary vehicle at 33.5 mph. The MDB is towed at a crabbed angle of 27° to its longitudinal axis. This configuration is intended to simulate a striking generic vehicle (MDB) moving at 30 mph, perpendicularly into the side of the struck vehicle (test item), which is traveling at 15 mph.

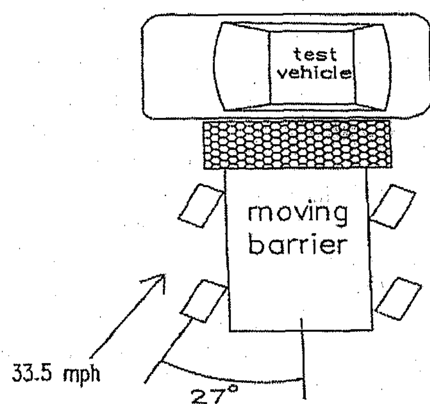


Figure 1. FMVSS 214 Crash Test Configuration

NHTSA Side Impact Dummies are placed in the front and rear seating positions on the struck side of the stationary test vehicle. An occupant injury criterion is based upon the lateral acceleration (Y axis) of the SIDs' upper and lower ribs, lower spine, and pelvis. A photograph of an actual crash test is presented in Figure 2.

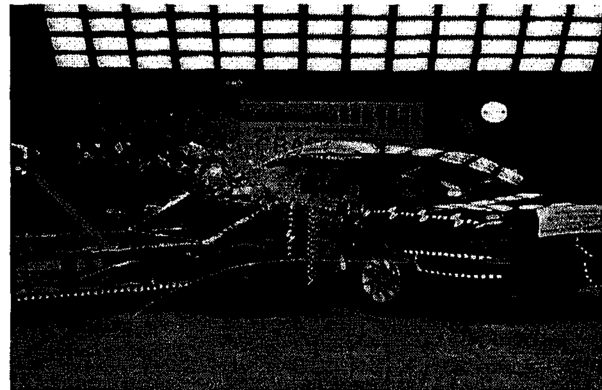


Figure 2. Photograph of FMVSS 214 Crash Test

It is emphasized that the injury criteria for the SID are based solely on the lateral acceleration responses measured in the dummy. In addition, it is also evident that the struck door and vehicle are accelerated primarily in the lateral direction, especially during the time when the dummy impacts the inner door panel. From a test simulation stand point, most of the effort can be directed toward controlling the lateral vehicle and door response and producing the desired SID response. Simulation of only one dimension also greatly reduces the complexity of any side impact sled simulation, since a typical acceleration type sled is designed to produce responses in only one direction. The problem is greatly simplified because now one-dimensional analysis and tests can be used to approximate the more complex FMVSS 214 test condition.

In a typical FMVSS 214 test, instruments will include accelerometers on the MDB, struck vehicle and SID. For the purposes of the approach outlined in this paper, it is important that further instrumentation be placed on the struck door and vehicle. This instrumentation is needed to set up the appropriate dynamic environment with the SITE system.

Accelerometers should be placed on the inner door skin behind the trim panel attachment. The placement of these accelerometers (longitudinally and vertically) should be in the general area of the dummy loading. Typically, accelerometers are mounted on the upper, center and lower positions very near to the dummy contact area. Accelerometers should also be mounted on the undeformed areas of the vehicle such as the seat attachment, floor pan, and unstruck side B-pillar. All of the accelerometers on the door and vehicle are in the Y direction. The data from these accelerometers are critical for setting up the SITE test and for precisely sensing when to and triggering a side impact air bag.

SITE OPERATING CONCEPTS

Presented in Figure 3 is a schematic drawing showing all of the basic components of the SITE system. As shown by this drawing, the upper carriage rests against the thrust column piston and rides on top of the lower carriage on a rail system. The lower carriage is attached to the sled rail system via a pressurized braking system. A pneumatic cylinder is used to

control the relative response of the two carriages. A door trim and other components (such as a pillar) that the occupant may interact with are attached to a flat surface on the upper carriage. A seat (with the dummy placed in it) is attached to the floor of the lower carriage.

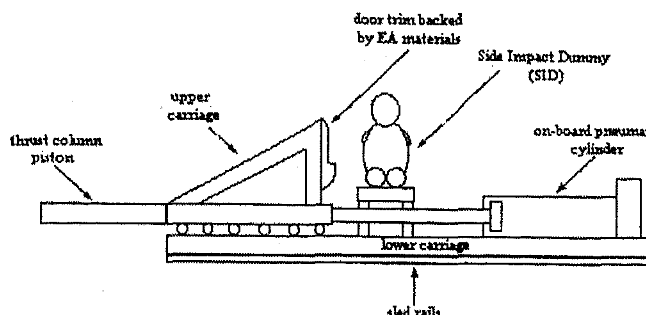


Figure 3. Basic Components of SITE System

When the sled is fired, the thrust column piston will push the upper carriage. The response of the upper carriage is controlled primarily by the shape of the metering pin and load pressure. Since the mass of the upper carriage is significantly lower than a normal carriage, the accelerations which can be produced are very high. The lower carriage will not begin to move until sufficient pressure is produced inside the onboard cylinder. Once the forces generated inside the cylinder exceed the braking forces produced by the braking system on the rails, the lower carriage will then begin to move. Thus, this approach allows you to produce different dynamic responses that simulate the struck door and vehicle simultaneously. The ability to control these responses results in an ideal tool for simulating a variety of vehicles including small cars through sports utility type vehicles. The overall capabilities of the system demonstrated through development are as follows:

- Upper Carriage Maximum G's ≈ 160 G's (with a 600 lb payload)
- Upper Carriage Maximum Velocity ≈ 30 mph
- Upper Carriage into Lower Carriage Maximum Intrusion ≈ 18 inches

The capabilities listed above are variables which are considered the most relevant for side impact simulation. The maximum values for the lower carriage are much lower, but the range is comparable to that seen in FMVSS 214 crash testing for the unstruck side.

An actual SITE test is conducted by attaching the seat to the lower carriage and placing an instrumented SID in it. A door trim is placed on the facing of the upper carriage and energy absorbing material is placed behind it. The role of the material is to attenuate the response of the SID so that it matches the crash test results. In essence, the crush of this material mimics the crush of the door during the crash test. Once the SID responses have been "tuned" to match the crash results, the simulation meets the key requirements (door, vehicle, SID responses) for an ideal simulation and that test condition can then be used to evaluate side impact air bags.

SITE COMPUTER SIMULATION PROGRAM

To conduct SITE testing it is important to be able to produce both the struck door and unstruck vehicle responses. A computer program was designed to simulate door and vehicle acceleration profiles from actual full scale side impact crash tests. There are a few parameters considered in the program, such as:

- pressures and volumes in the sled thrust column
- pressure in the on-board pneumatic cylinder
- lower carriage braking pressure
- total weights of upper and lower carriage
- fire pressure
- metering pin shape

The program works by accepting these inputs and predicting the upper and lower carriage responses by solving the governing equations for the system. The output of the program are the acceleration, velocity, and displacement response of each carriage. The pressures and volumes in the thrust column determine the total amount of energy which can be generated by the sled. The total mass of the upper carriage and the total mass of lower carriage are two variables which can affect the peak values of the acceleration profiles. The lower carriage braking pressure and sled fire pressure are secondary variables and have less influence on the acceleration profiles. The most important variable in the simulation is the shape of the metering pin. The shape which is typically used to match the door response is one which initially allows a large flow of gas through the orifice, and then chokes off the flow very quickly. This is what produces the very high acceleration response early and what causes the immediate decrease in G's.

BASELINE MATCHING

The first step which must be completed before airbag development work can be started is to establish a baseline test which correlates very well with full scale crash test data. Once this has been achieved, side impact air bags can be tested and the effects of the air bags on occupant injury levels can be evaluated.

Presented in Figure 4 is a flowchart showing the general steps involved in achieving a baseline test condition which correlates with the full scale crash test results. The first step (Step 1) is to determine specific information and data from the full scale crash test on the baseline vehicle. This step is somewhat crucial as this information is a direct input into the baseline and the accuracy of the input data is very important with regard to the overall validity of the model. Following is a listing of what is needed to conduct the baseline matching test series.

- struck door and unstruck side vehicle accelerations
- SID responses including upper rib, lower rib, upper spine, lower spine, and pelvis Y axis accelerations
- High speed films showing SID interaction with door and other vehicle structures.
- dimensional data including relative positions of the SID, seats, door, and other components

Initially, the only information which will be used is the door and unstruck side vehicle acceleration responses. As shown on the flowchart, the next task (Step2) is to utilize the SITE computer model to determine appropriate setup variables to match the door and vehicle accelerations with the upper and lower carriage accelerations, respectively. Using the computer model developed specifically to predict upper and lower carriage responses (as well as previous experience), the metering pin configuration, pressures, and volumes can be determined. Next, the thrust column and onboard backing cylinder should be set to these variables. Once this is done, sled pulse trials (Step 3) are completed until good correlation of the upper and lower carriage to the struck door and unstruck vehicle acceleration, respectively, is achieved. The most important response to match is the door acceleration, while the unstruck side of the vehicle is considered a secondary effect. It should be noted that for other vehicle types such as light trucks or sport utility vehicles, the response of the unstruck side of the vehicle may be of greater importance. The upper carriage acceleration should match very closely with the door acceleration through the dummy injury phase of the event. Correlation of these responses include both magnitude and timing.

Following this, the upper and lower carriages must be setup for the full system tests (Step 4). Most of the effort here is directed toward fixturing of the door trim, pillars, and seat so that the geometry and spacing are identical to the baseline vehicle. It is important that the exact geometry of these items is duplicated because of the implication of deploying a side impact airbag properly (once a baseline condition has been established).

Steps 5 and 6 consist of conducting full system tests while analyzing the dummy responses. The main variables influencing the SID response are the energy absorbing padding placed between the flat plate and the door trim. The padding choice of the first test is based mostly on previous experience. Based on the initial test results, the dummy responses can be tuned in by changing the location, thickness, and stiffness characteristics of the padding materials. The decision of what padding configuration to use is based mostly on the SID responses. To conduct this part of the baseline matching testing, a wide range of padding materials must be available. Also, the sled test films showing the dummy motion must be compared to the crash test films. If there are significant differences in the dummy kinematics, the setup variables must be carefully reviewed.

Following multiple test iterations, a reasonable baseline condition should be achieved. Once this has been done, an additional one or two tests should be conducted for repeatability purposes. It is thought that the baseline matching can be completed in approximately 6-8 full system tests.

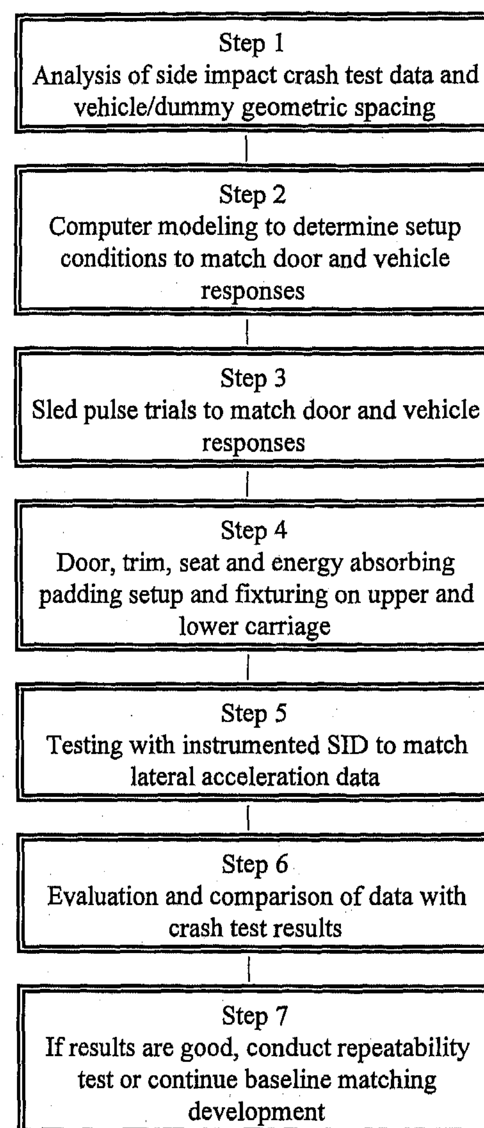


Figure 4. SITE Baseline Matching Flowchart

BASELINE TESTING RESULTS

As part of the overall development of the SITE system, a "baseline" test condition was established using a 2-door mid-size vehicle. The baseline crash test was a FMVSS 214 crash test that was part of a federal crash test program, of which the results are publicly available. The crash test was conducted using the government compliance test procedure. Following the steps listed in Figure 4, a baseline condition was established. Referencing Figure 4, the results for Step 3 (door and vehicle response matching) are presented in Figure 5. As shown by these over-plots, the inner panel struck door acceleration was matched reasonably well by the SITE system. The overall magnitude, timing, and general shape of

the response was reasonably simulated. The unstruck side vehicle response (which is not presented here) was not considered to be of significance for this simulation because the vehicle data showed only a very small amount of displacement during the injury phase. Also presented in Figure5 is the velocity-time overplot.

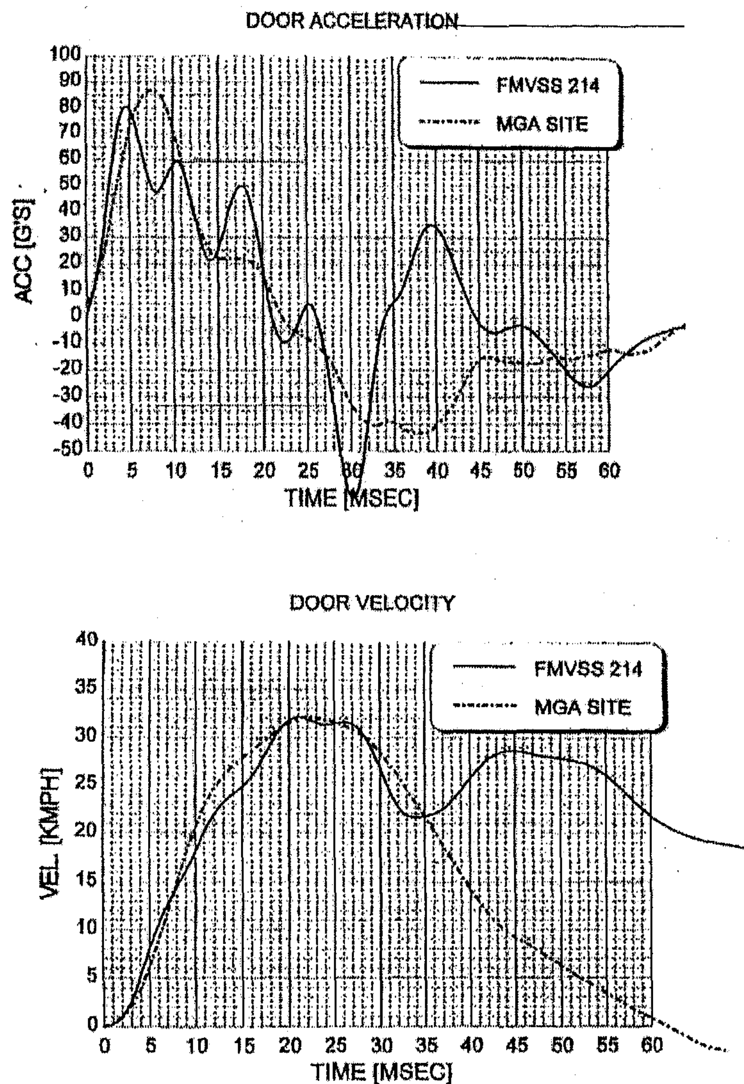


Figure 5. Door Acceleration Overplots

Once the upper carriage response was established, the setup for the first SID test was completed. The vehicle components for this test included the door trim and seat. Presented in Figure 6 are two photographs from this testing. Following an iterative testing process, a baseline was established. This baseline condition was based upon the door and SID lateral responses. Presented in Figure 7 is a series of overplots comparing SITE results with FMVSS 214 results. Based on the data, the SITE SID results were within approxmiatley 10% for the peak accelerations and 5 msec for the timing of the peaks. Likewise, the injury criteria values for the crash and SITE results showed the same level of correlation. Analysis of the crash and SITE films demonstrated that the SID's motion was very similar for each type of test.

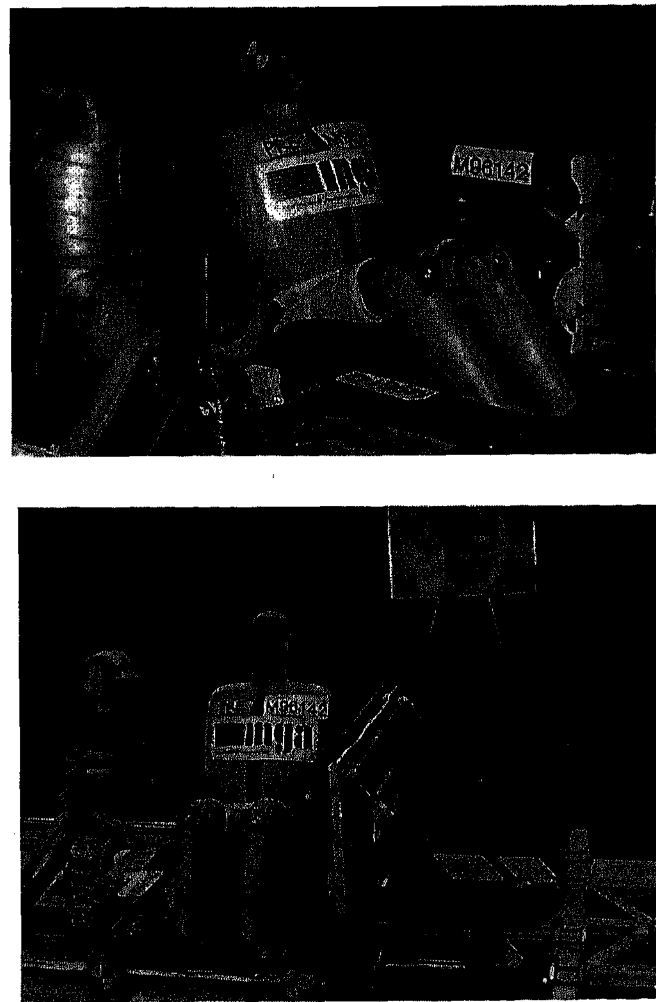
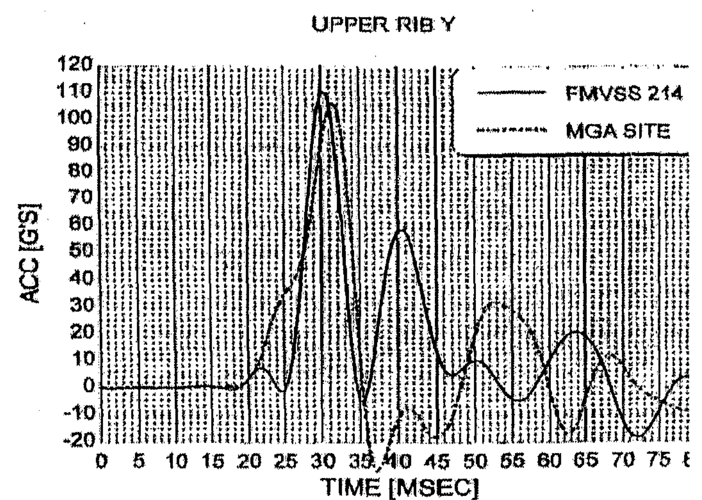
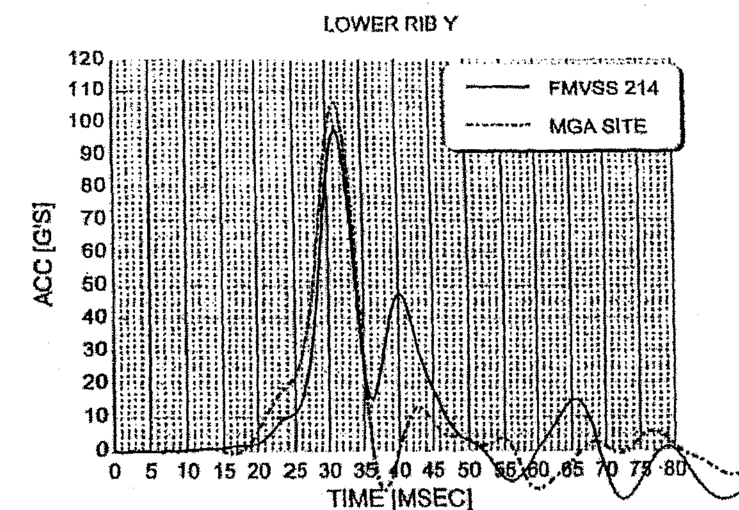
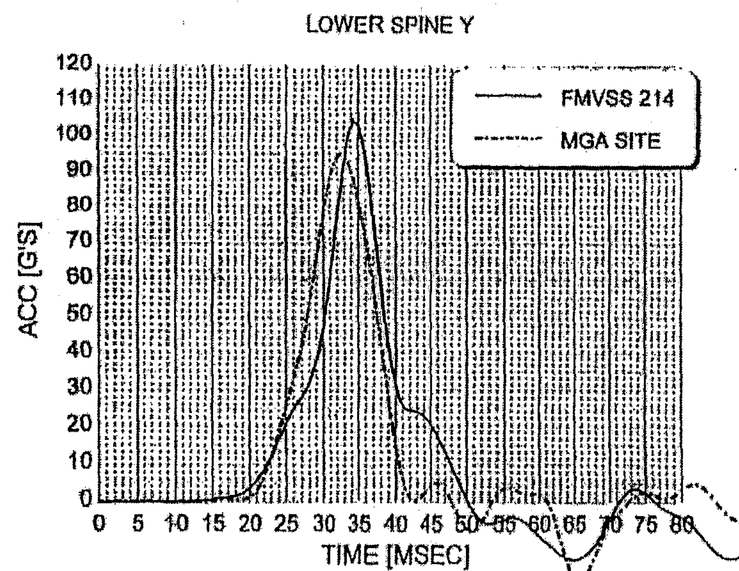


Figure 6. Photographs of SITE Sled Test







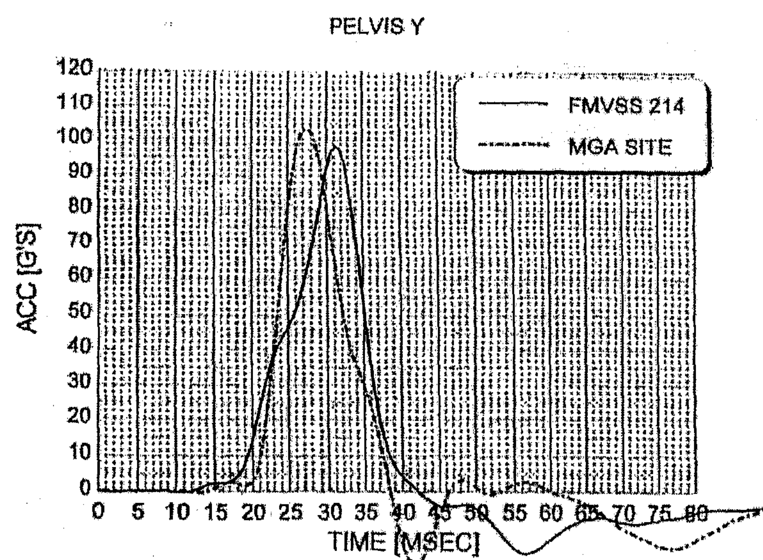


Figure 7. SID Acceleration Overplots

SIDE IMPACT AIRBAG TESTING

In order to further develop the SITE methodology, testing was conducted with and without side impact air bags. The air bags were provided by Breed Technologies to aid in the development of the methodology. It should be noted that the air bags which were used were "generic" in nature and not developed specifically for the baseline vehicle. The air bag was a seat back mounted unit which was attached to the seat back frame. The air bag was triggered based upon the upper carriage acceleration and was positioned to deploy so that it filled the space between the occupant and inner door. Presented in Figure 8 is a series of photographs taken from high speed photography, which show the deployment sequence.

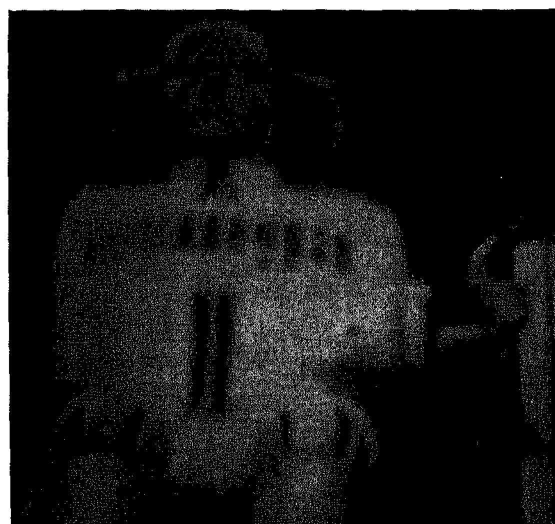


Figure 8a. Pre-deployment

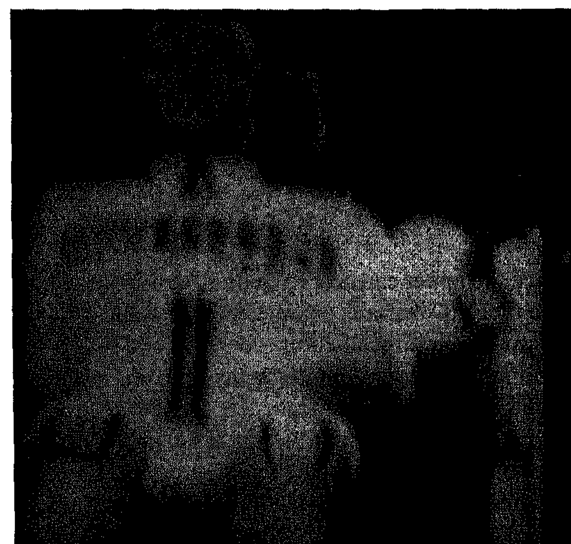
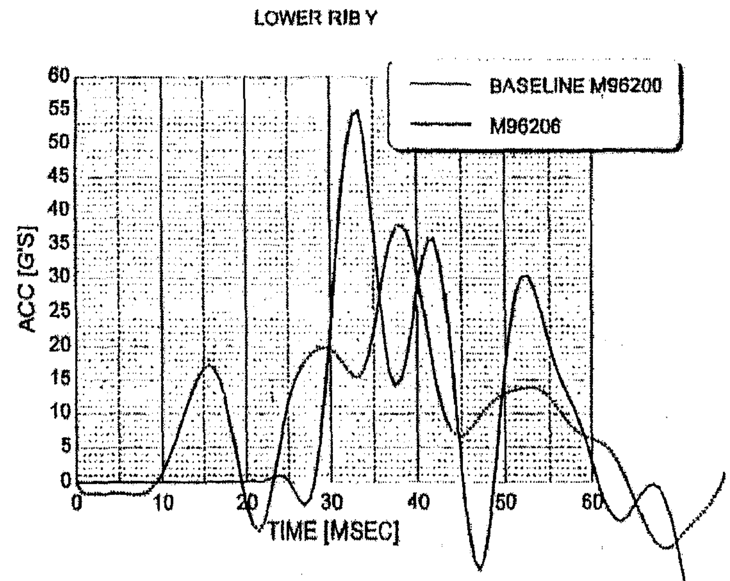


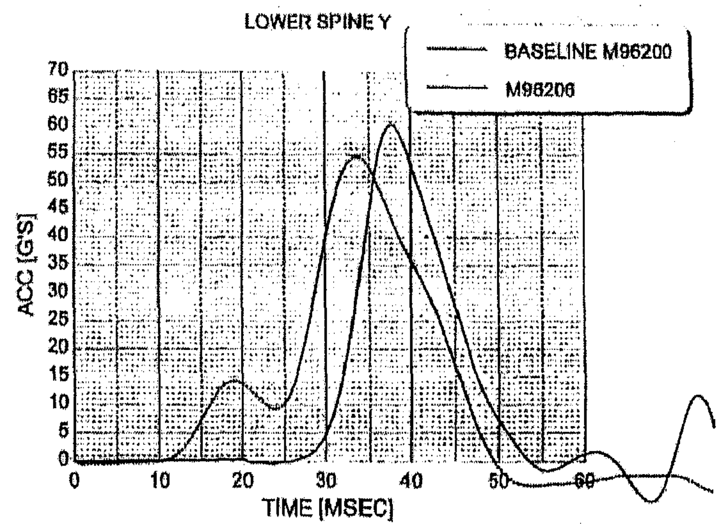
Figure 8b. Space between the SID and the door panel being filled up by the side impact airbag

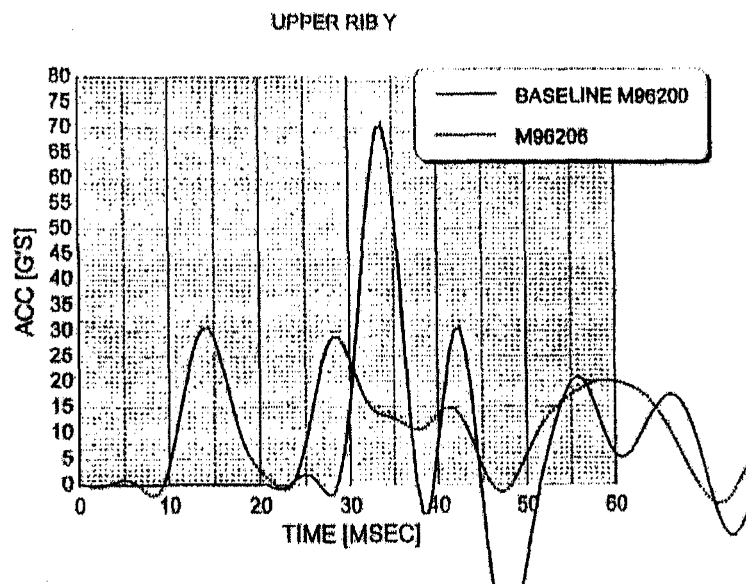


Figure 8c. Full Deployment

The overall effect of the air bag on the occupant response is shown through a series of overplots presented in Figure 9. As shown by this data, the air bag did provide some benefit to the occupant in terms of reducing the magnitudes of the peak accelerations and injury criteria values. In general, the loading on the dummy occurs earlier in the event and produces lower peak accelerations. The upper and lower rib responses were reduced by approximately 40%, while the lower spine and pelvis responses were reduced by approximately 10% with the airbag. It is at this point when the real benefits of the SITE are evident. This test methodology could now be used to further develop and fine tune the air bags without having to conduct a full vehicle crash test. Since this test requires so few components (door trim, seat, air bag), the costs to evaluate a new potential design are considerably reduced.







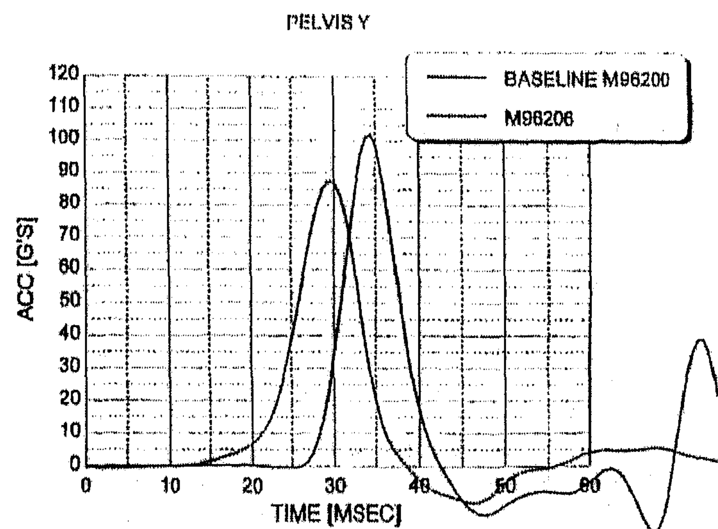


Figure 9. Side Impact Airbag Test Results (M96206)

CONCLUSION

This paper presents a component level test methodology which can be used for side impact air bag development which utilizes both technology and facilities commonly used for accelerator-type frontal impact sled simulation. The test which is presented is a very straight forward and simplified approach which requires very few parts. The level of correlation with actual full scale test results is reasonably high. Although results are not presented, the test to test repeatibility using the procedure is similar to what is shown for full scale FMVSS 214 testing.

The purpose for this paper was not to highlight specific strong points of the SITE system, but more to describe the procedure and indicate how this type of testing could be implemented into side impact air bag development. At this point, future development of this system is being directed towards including a complete door which is impacted with a moving barrier. It is thought that perhaps this approach may be more suitable for padding or door structure development. Results for this effort will be presented through future SAE proceedings.

ACKNOWLEDGMENTS

The authors of this paper wish to express appreciation to associates at MGA, as well as customers of MGA who have supported the development of the SITE system. In addition, a special thanks is given to Breed Technologies for providing air bags and technical assistance on this project.

REFERENCES

(1) NHTSA, Laboratory Test procedure for Federal Motor Vehicle Safety Standard No. 214 - Side Impact Protection-Passenger Cars. U.S. Department of Transportation, Washington, D.C. August 1991.

(2) NHTSA, Notice of Proposal Rule Making to Amend Federal Motor Vehicle Safety Standard No. 214 - Side Impact Protection - Passenger Cars. Docket No,. 88-06, Notice 8, U.S. Department of Transportation, Washington, D.C.

(3) Miller, Patrick M, and Miller II, P. Michael, Evaluation Methodologies For Automobile Side Impact Development, oral/written presentation (SAE Paper # 930483) at the SAE Conference held March 1-5, 1993 at Detroit Michigan.

970572

Apparatus and Method for Side Impact Testing

Douglas J. Stein
Morton International

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The introduction of the dynamic FMVSS 214 crash test requirement has prompted the need for an economical test method for the development of side impact occupant protection devices. Full scale side impact crash testing is too costly for the development testing of occupant protection components.

This paper describes a unique test method for side impact sled testing which utilizes a standard 12-inch HYGE type sled facility. It has been demonstrated, with several hundred tests, to provide high repeatability and good correlation with full vehicle side impact crash tests. An overview of the device and details on how it is used is provided. Typical data correlation between sled tests and crash tests is shown for FMVSS 214(dynamic) testing. EEVC side impact and lateral crash tests into a rigid pole have also been correlated with this device and will be made available upon request.

INTRODUCTION

Development testing of frontal impact protection components relies heavily on crash simulation tests. These tests are frequently conducted using devices like the CVC HYGE sled, which accelerates the occupant compartment of a vehicle in reverse with a controlled "crash pulse" that reproduces the acceleration and relative velocity of the vehicle during a crash test. The velocity change occurring between the occupant and vehicle interior during a side impact crash, however, occurs too quickly for most acceleration based sleds to simulate directly. Until recently, many believed that side impact tests could not be simulated using a HYGE type sled.

When modeling frontal impact tests, the occupant compartment structure is generally modeled as a rigid body due to minimal deformation rearward of the A-pillar in most frontal crashes. In side impact tests, however, the door, b-pillar, seat, roof rail, and floor pan of the struck vehicle all move different amounts at different times. The term "crash pulse" as it is normally understood for frontal impacts, does not apply for side impact. Instead, the dynamics of those components that directly effect occupant injury must be considered separately.

For the FMVSS 214 dynamic certification test condition, occupant injury is attributed primarily to dummy interaction with the intruding door of the vehicle. As the moving deformable barrier (MDB) crashes into the side of the stationary target vehicle, the struck door is crushed externally, and accelerated into the occupant. The inner door stiffness (a result of external intrusion) and the door velocity profile as it strikes the occupant are the primary factors resulting in specific occupant injury.

It is the door velocity profile that creates the greatest difficulty in simulating the side impact event on a conventional HYGE type sled test device. In a side impact crash, the door is accelerated from zero to approximately 10 m/sec in only about 20 ms. Over this 20 ms, the inner door only moves about 100 mm before impacting the occupant. The HYGE type accelerator sled, when used in the conventional way, is unable to accelerate the door to the needed velocity this quickly.

Consider a side impact crash event reduced to this simplified model:

An intruded door, with its deformed interior geometry and stiffness, is accelerated to some velocity before impacting a stationary test dummy.

If the door geometry, stiffness, and impact velocity profile from this crash can be reproduced in a component level test, the response of an impacted dummy should be nearly equal to that of a vehicle level test. Then, if some energy absorbing device, such as padding or an air bag, is placed between the door and occupant, the effect of the device on dummy response can be measured. It is this logic that provides the basis for the test device and methodology described in this paper.

SIDE IMPACT TEST DEVICE

Sled testing to simulate the side impact crash event has often been achieved on a HYGE sled by positioning the dummy far enough away from the door to allow the required velocity to be attained prior to impact (Heidelberg method). Although this method may be satisfactory for testing door trim and padding materials, it does not produce the geometry or timing correlation needed for dynamic side airbag testing. Relative positioning between the door, seat, and test dummy must be maintained to properly reproduce the air bag deployment space actually available during door intrusion. Timing must also be correlated between the sled and crash tests to enable air bag deployment times to be established and reproduced.

To fulfill these required test conditions, a unique test device and methodology has been developed and patented. The test device was developed to utilize a standard 12-inch bore HYGE sled facility. All facility support items, such as photographic equipment, lighting, data acquisition system, and sled braking apparatus are unchanged between frontal impact and

side impact test configurations. This commonality minimizes the complexity and time required for change-over between frontal impact and side impact sled tests. The side impact test device is fully contained on top of the existing accelerator sled carriage, as shown in Figure 1, and is attached using existing mounting holes.

The device consists of a ram, a dual guide-rail system, a secondary carriage or "dolly", and an aluminum honeycomb crush element. The dolly represents the occupant compartment of the struck vehicle. An aluminum base plate and vertical steel tube structures contain mounting hole patterns for universal mounting of test components. For FMVSS 214 test simulations, the dolly is typically outfitted with a trimmed door shell (without internal mechanisms), a B-pillar segment with trim, seat belts, and a seat which is mounted to the floor of the dolly using a sliding seat base to enable the seat to slide laterally. When properly configured for a test, the seat, door, b-pillar, and test dummy are positioned the same as in the subject vehicle. The dolly is mounted to the sled carriage through linear motion bearings to two guide-rails bolted to the top of the carriage in its longitudinal direction. The bearings and rails allow the dolly to slide in the subject vehicle's lateral direction (same direction as sled carriage motion).

A rigid ram is bolted to the front of the sled carriage in such a way that its flat impact surface is parallel to the leading face of the dolly. On the ram impact surface, a block of aluminum honeycomb is mounted as a crush element between the ram and the dolly. Typically, the honeycomb crush element is a block of Hexel Specification Grade aluminum honeycomb with a nominal crush strength of 1241 kPa (180 psi), a cross-section area of 0.1858 m^2, and a crush depth of 20 cm.

At the start of a test, the dolly is positioned approximately 75 cm away from the impact surface of the honeycomb. When the sled is activated, the dolly remains stationary (due to its inertia) as the sled carriage is accelerated to the desired velocity. When the honeycomb strikes the dolly's impact face, the kinetic energy of the sled carriage is used to accelerate the dolly while crushing the honeycomb. As the dolly (with the door rigidly attached) accelerates, the seat and dummy, which are on the sliding seat base, remain stationary due to their inertia. Therefore, like a side impact crash test, the door is accelerated toward the seat and occupant. The door will then strike the seat, pushing it from under the dummy until, finally, the door strikes the dummy. With this scenerio, the dynamic relationship between the door and occupant, and between the door and a seat mounted airbag module, can be simulated.

It must be pointed out, that this test device is not intended to simulate the entire duration of a side impact crash event. Only the time frame in which the occupant is subjected to the primary injury causing loads is evaluated. It is during this time, usually from 0 to 45 milliseconds, that an airbag must deploy into the available space and provide protection to the occupant. When the primary interaction between the door and dummy is complete, the onboard sled carriage brakes are activated. Two latches, activated during the crush of the honeycomb, hold the dolly to the carriage while braking.

TEST METHODOLOGY

For any side impact test device to be suitable for side impact air bag evaluation, the following elements of the crash event must be accounted for:

GEOMETRY - It is important that all components in the air bag deployment zone be positioned so that the target vehicle geometry is modeled. The shape of the air bag, the output of the inflator, and the mounting scheme used are based, at least in part, on the vehicle interior geometry and the reaction surface for the air bag as it deploys. For all side impact testing conducted with this device, the door, seat, air bag module, and test dummy are positioned in proper relationship to each other as defined by customers' vehicle layout drawings.

VELOCITY / TIMING - Because the air bag must be able to deploy into position within about 15 milliseconds after impact, it is important that the door intrudes upon the dummy at the correct velocity and timing. To protect the occupant from serious injury in most vehicles, the air bag module must be activated about 5 ms after impact and be into position to provide protection within another 10 to 12 ms. Geometry and timing work together to ensure the deployment environment, as a function of time, is properly simulated. Figure 2 illustrates the relationship between sled carriage, sled fixture door, and crash vehicle door velocities. Note the time it takes for the sled carriage to attain the required velocity.

DOOR CRUSH CHARACTERISTICS - If the door geometry and timing are correct, the only other variable that will significantly effect dummy response is the crush characteristics of the door surface. To establish the correct crush stiffness for the door, a door from the crash tested vehicle is examined for deformation from exterior intrusion and from dummy impact. Deformation to the door's exterior ultimately has the effect of reducing available crush space inside the door, particularly in the lower portion of the door where intrusion from the bumper shape of the moving deformable barrier (MDB) has intruded. The door targeted for sled tests can be modified by adding structure to the inside of the door. This structure is designed to model the intruded shape of the external door surface, thus reducing available inner door crush distance. Because the deformation to the door during a crash test occurs before the dummy is involved in the crash, pre-test modification to the door should not adversely affect dummy response correlation.

A component level test device is not intended to provide a perfect simulation of a crash. Crash tests contain many variables, both in vehicle structure and in test setup, which create test repeatability issues. These variables can be reduced significantly on a component test device, but not without some loss of model accuracy. This device is no exception. By understanding these limitations and concentrating on the primary vehicle to dummy interactions, the effects of these limitations can be tolerated.

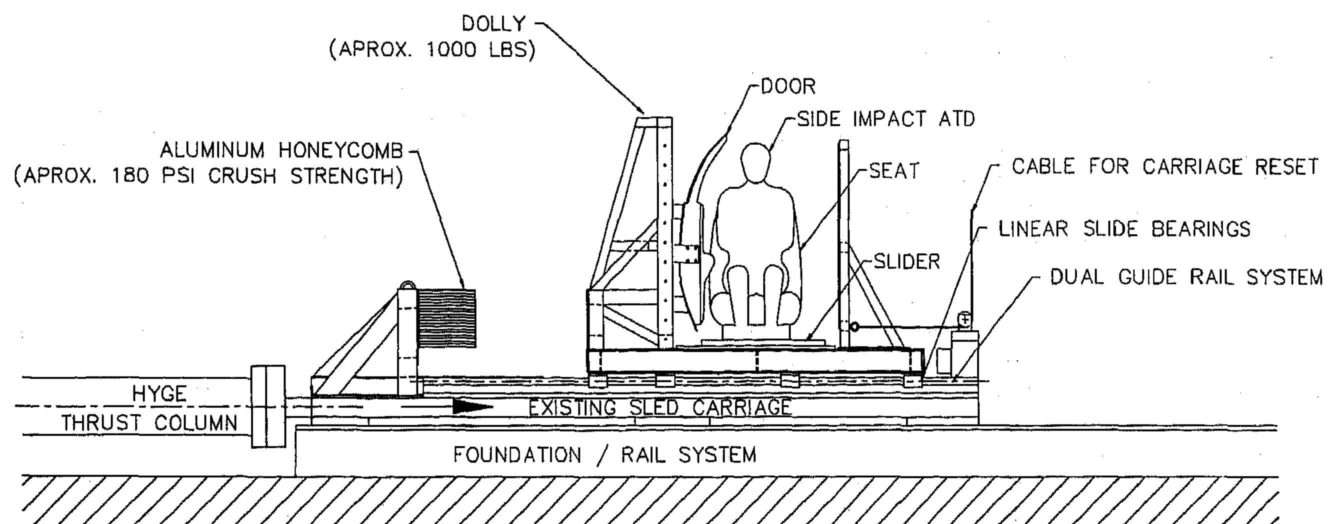


Figure 1. Side Impact Sled Test Device

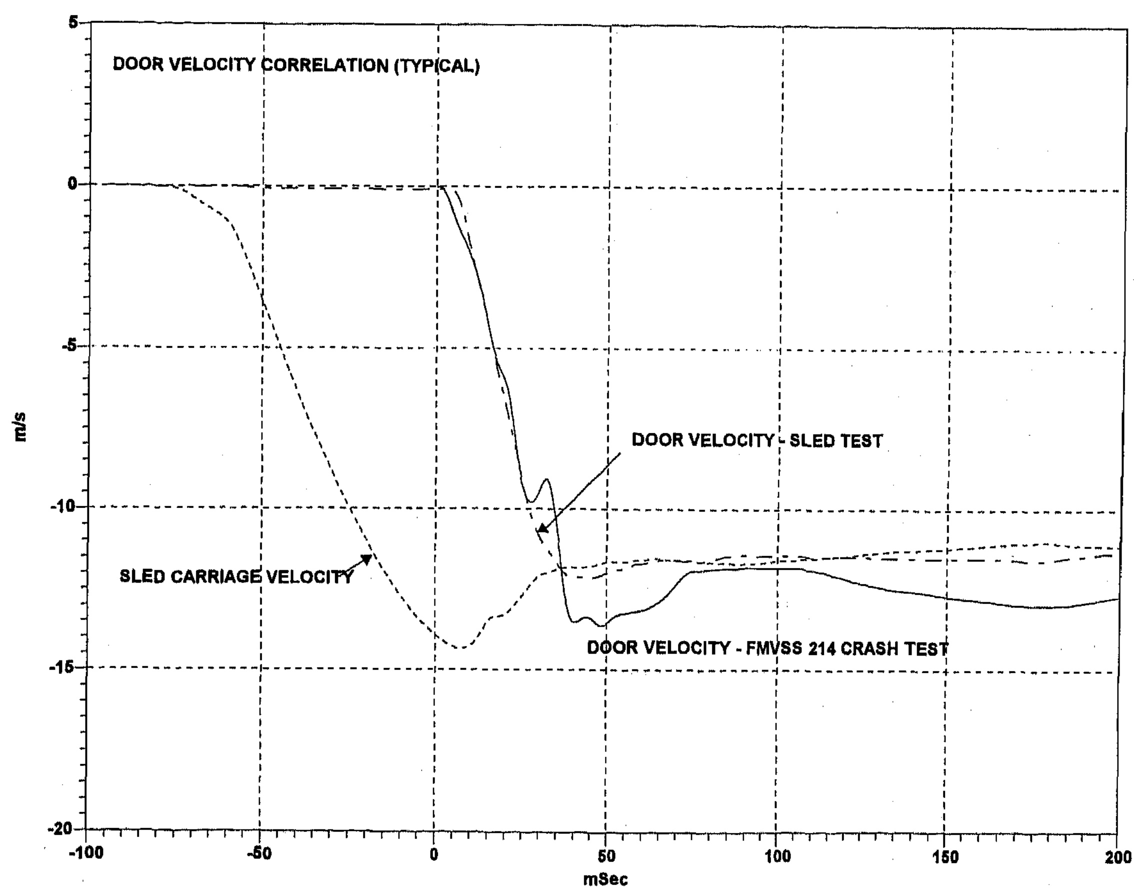


Figure 2 - Correlation of Door Velocity (typical)

Because the door and b-pillar are rigidly attached to the dolly, it is difficult to simulate door intrusion in cases where there is significant rotation (for example, when the bottom of the door intrudes more quickly than the top). If the rotation occurs prior to impact with the dummy, it can be simulated by pre-positioning the door in it's tilted position. Many tests have been conducted with the bottom of the door tilted inboard about 10 degrees. If the door rotation is still occurring at the time of dummy impact, pre-positioning will not be completely effective. Either the pelvis will be impacted too slowly in order for the rib impact velocity to be correct, or the rib impact will be too fast if the pelvis velocity is targeted. Because we are typically interested mainly in the effects of side air bags on thoracic injury, the rib impact velocity is our primary concern.

When simulating ECE side impact tests with the seat in full rear position, head to b-pillar impact correlation is also difficult to simulate. In this type of crash test, the upper b-pillar intrudes much more slowly than the lower b-pillar and door. This is because the upper b-pillar is attached to the roof, which only moves after significant intrusion of the MDB into the door and lower b-pillar has occurred. When simulating this test condition, the head impact velocity and the time of head impact must be addressed. Head impact timing can be accommodated by pre-bending the b-pillar to represent it's configuration at the time of head impact (through crash test film analysis). Unfortunately, in many tests there is still some difference in velocity between the upper and lower pillar at the time of head impact. Because the sled test device moves the upper and lower b-pillar at the same velocity, the head impact velocity is slightly higher than in the crash test. An over speed head impact means the head portion of a head/thorax side air bag will be over tested, which is not bad if the error is not great.

Side pole impact tests are often conducted to evaluate worse case head impact from the vehicle sliding laterally into a pole or tree. For most vehicles, the velocity at which the vehicle impacts the pole is only slightly greater than the velocity at which the pole hits the occupant's head. There is very little resistance to the initial pole intrusion into the vehicle. For pole impact simulation, some simplifications to this complex crash event are needed. The pole intrusion into the vehicle is quite severe, making the ability to simulate the event with this test device seem unlikely, due to its inability to simulate dynamic deformation of the door. But consider the actual impact with the dummy. Again there is simply a door, bent and stiffened by the pole, which strikes the dummy, and a pole which strikes the dummy's head. We could pre-deform the door to represent it's configuration at the time of dummy impact, but that would create an unrealistic deployment surface for the air bag to deploy against. Normally, the pole impact test is configured so the pole is aligned with the center of the dummy's head. With the door deformed as it is, the pole becomes the primary dummy loading device, with only the trim panel and crushed door sheet metal between the pole and dummy.

To simulate this on our apparatus, we start with the same setup as used for FMVSS 214 simulations. A pole segment is mounted to the fixture in the door window opening, and extends from the door belt line to the roof, with it's leading edge at the door-to-trim panel interface line. Below the pole, the door shell is rigidized to simulate the heavy intrusion of the door by the pole in a crash test. Only the area of the door directly impacted by the pole is rigidized, leaving the remaining door structure unsupported. The loading to the dummy from the unsupported door surface is relatively insignificant compared to the loads imposed by the reinforced segment below the pole. With the test configured in this way, we can closely reproduce the significant dummy loading paths while maintaining the door trim surface for air bag deployment. This method has been correlated successfully on many vehicle programs and used as an aid in head protection air bag development. Remember, the primary intent of this test setup is to evaluate an air bag's ability to protect an occupant when impacted by a rigid pole at a given velocity. This method will obviously not be of any value in evaluating door or vehicle structural changes.

EFFICIENCY - In the development of any air bag module, a large number of tests are needed to optimize the system and verify performance for a variety of conditions. A test method which is simple, quick, easy to setup, consumes minimal test hardware, and is highly repeatable is a real asset. This test device provides this efficiency. A primary goal in the design was to simplify the crash event to its main injury causing interactions. Only those vehicle components that interact with the dummy during the critical, injury generating time of the event are consumed during a sled test. Typically, the only parts consumed during a sled test are the door, door trim panel, and the seat. Any other vehicle parts utilized are used as part of the fixture to generate the proper geometry and air bag reaction surfaces. Because minimal parts are consumed, there is little effort required between tests. Turnaround time between tests is typically less than for frontal impacts due to the reduction in replacement parts.

Setup change between frontal impact sled tests and side impact sled tests is minimal, requiring no more time than a frontal impact changeover. The side impact device is contained in two sub assemblies that bolt onto the sled carriage just like a normal sled "buck". All support equipment, such as photographic lighting and cameras, data acquisition wiring, sled carriage braking, and system operational devices are left unchanged.

As with most testing, repeatability of data is directly related to the number of uncontrollable variables. The HYGE sled system, if well maintained, has proven to be very repeatable when used for frontal impact testing. Side impact testing with the HYGE demonstrates the same high level of repeatability, minimizing the number of repeat tests required to filter out test variability. Repeat tests are typically conducted only for verification testing, or when expected variance from a design change is small. Sled testing continues to be a much more efficient tool for sorting design parameters than full barrier frontal or side impact tests.

The following two pages contain data plots showing the correlation attained in an actual sled test series (Figures 3 through 6). This sled test was conducted to establish baseline (no air bag) data correlation at the onset of an air bag development program. The solid line in each plot indicates the dummy data recorded during an FMVSS 214 crash test, while the dashed line denotes sled test dummy data. Correlation to this level can typically be attained in two to three tests by an experienced test engineer.

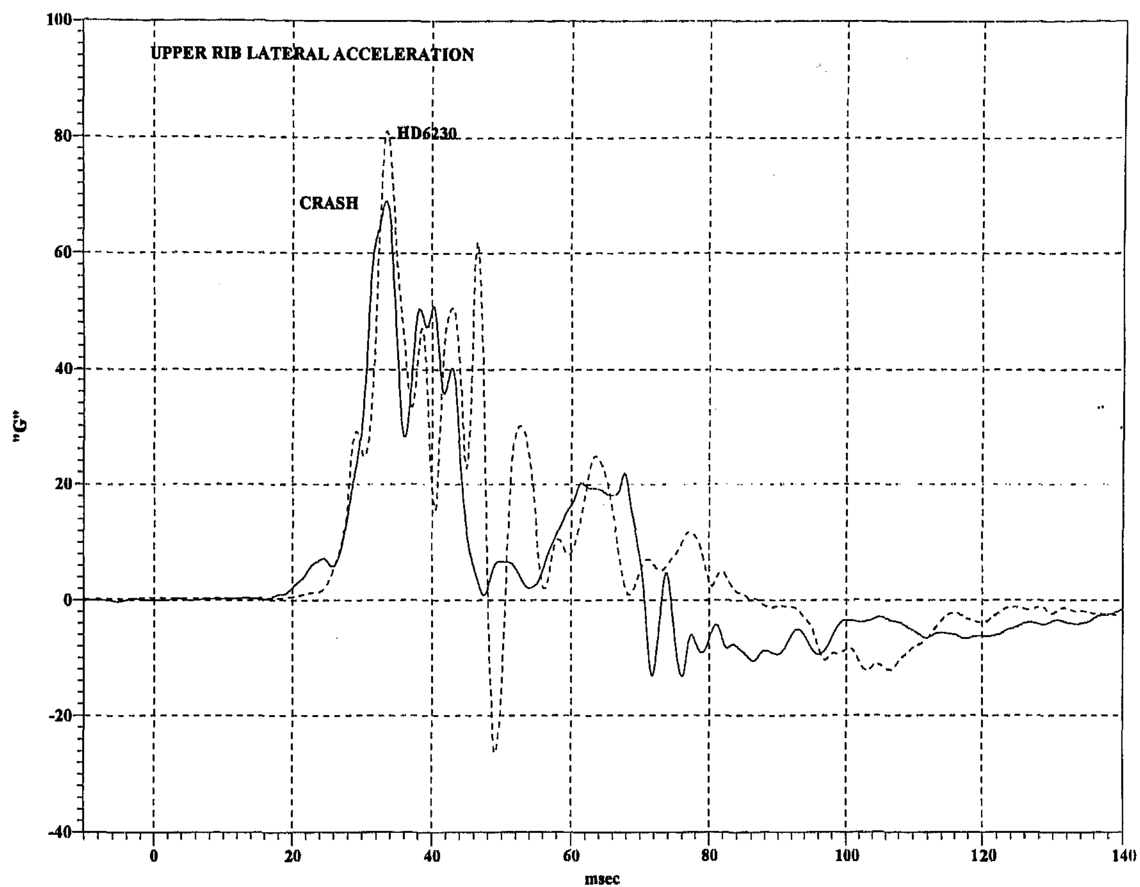


Figure 3. Correlation of Lateral Upper Rib Acceleration

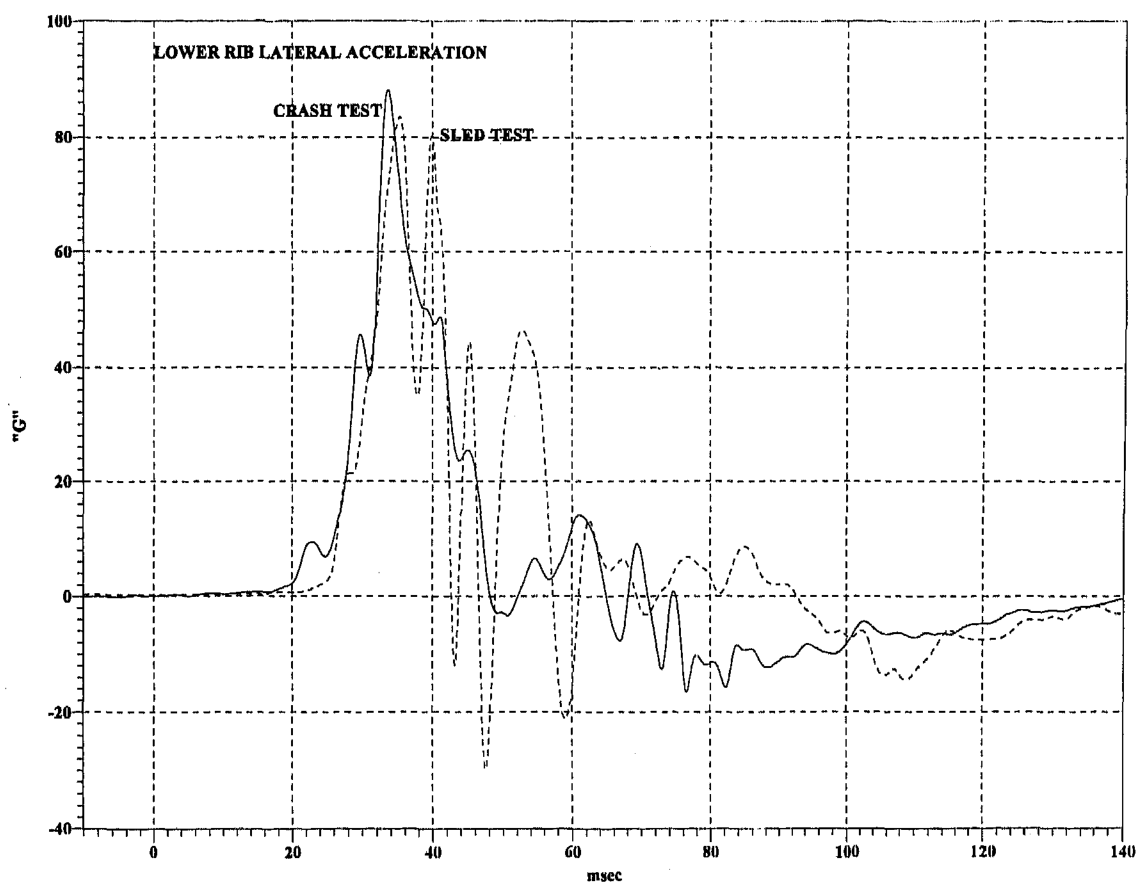


Figure 4 - Correlation of Lateral Lower Rib Acceleration

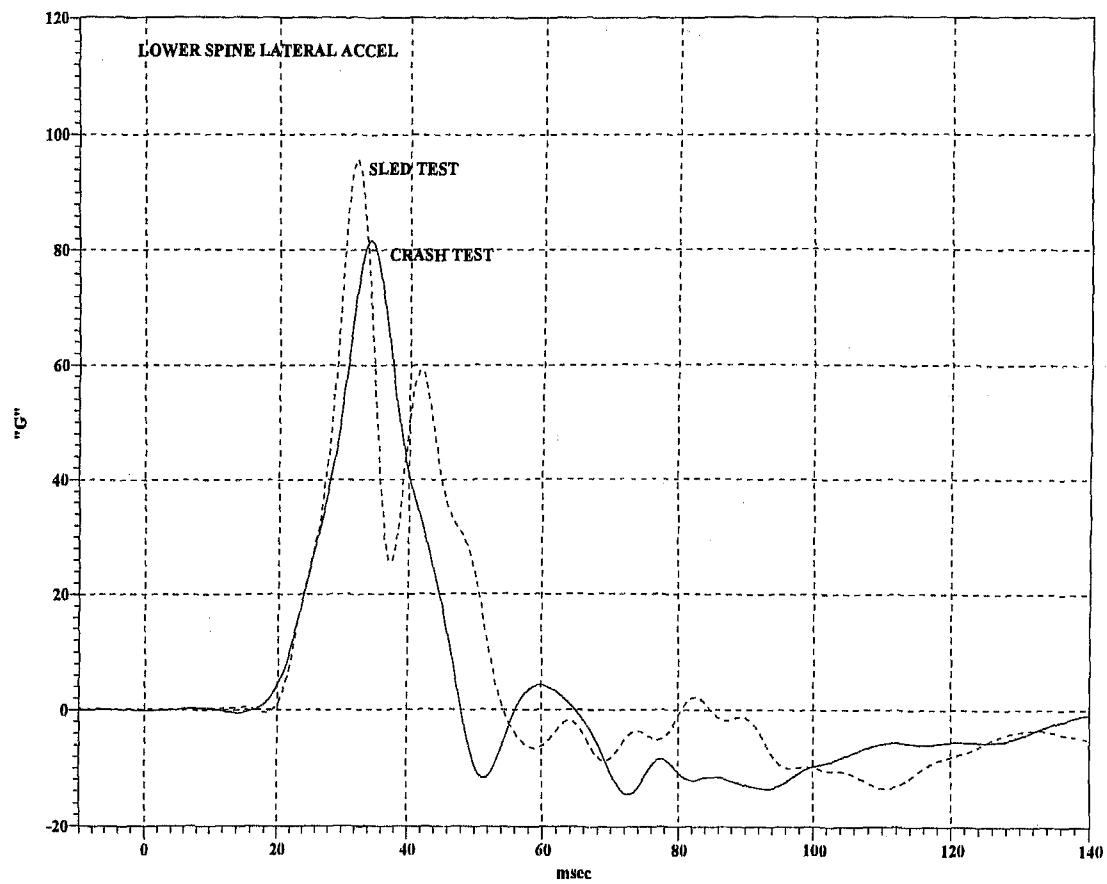


Figure 5. Correlation of Lateral Lower Spine Acceleration

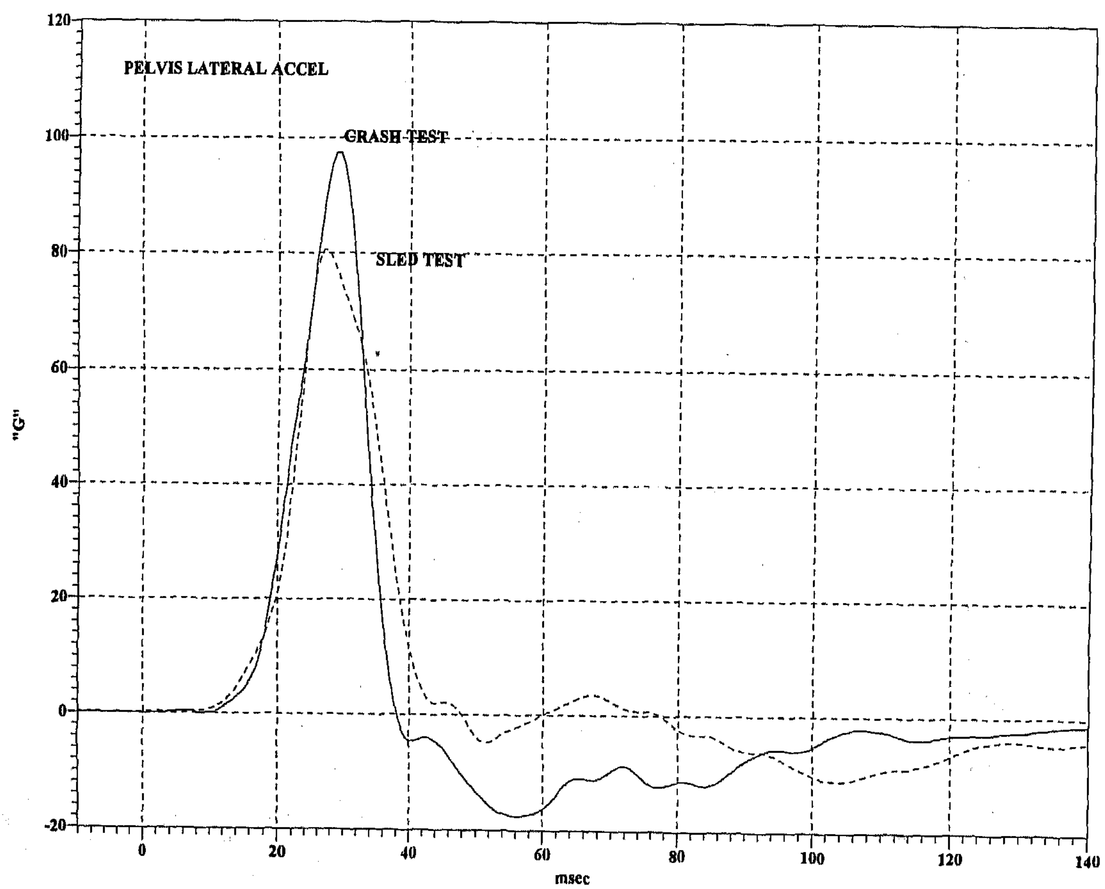


Figure 6. - Correlation of Lateral Pelvis Acceleration

SLED TEST CORRELATION:

The typical procedure for side impact air bag development programs is shown below:

- Receive vehicle and occupant test data from a baseline crash test (no air bag).
- Baseline sled correlation is obtained in 2 or 3 tests.
- The side air bag module is developed using the sled test as the primary system performance measure.
- The module is tested to other conditions such as EEVC or Pole impact requirement and "Out of Position" tests using adult and child dummies.
- One or more crash tests are conducted by the customer using the new air bag module to verify system performance.

Occupant data from the baseline crash test should be provided along with vehicle data. The data provided by accelerometers mounted to the struck vehicle's b-pillar or door sheet metal may not provide usable data for establishing sled correlation. Oscillations and noise in the data caused by the flexibility and deformation of these components during a crash test often make the data unusable for velocity computations.

Dummy response provides a good means for door velocity correlation. The initial accelerations of the test dummy ribs, being light by comparison, are affected to a high degree by the velocity of the approaching door. The total change in velocity of the larger dummy masses, obtained through integration of the spine and pelvis accelerations, indicate the energy transfer into the dummy over time. These velocities are not so much indicative of the inner door sheet metal velocity, but of the average velocity of the intruding barrier.

It is also helpful to have the dummy data available when tuning sled performance for optimum correlation in occupant response. If the velocity is correct, and the time to impact is correlated based on dummy impact timing, then the door stiffness may need adjustment to improve correlation. The best way to tune the door stiffness is to analyze dummy response as a result of impact with the door, and adjust accordingly. Once some experience is gained, a satisfactory correlation can be attained in 3 to 4 tests. It is possible to correlate the door stiffness by measuring the force vs deflection characteristics from a door taken off a crash tested vehicle, then matching that stiffness in the sled test door. This option, however, requires more testing to obtain correlation than tuning the door through sled testing.

Dolly velocity is adjusted by changing the "load" pressure in the HYGE cylinder. The particular pulse generated by the thrust column is not important here, as the acceleration of the dolly is controlled only by sled carriage velocity at the time of impact and by the crush characteristics of the honeycomb crush element. It is important that enough gap between the honeycomb and dolly is provided to ensure the sled carriage attains constant velocity (zero acceleration) before impacting the dolly. Otherwise, the actual velocity at which the carriage strikes the dolly becomes dependant also upon the spacing between the honeycomb and dolly, adding one more variable to the test conditions.

The rise rate for the dolly velocity curve is directly proportional to the stiffness of the honeycomb crush element. Therefore if a 20 percent increase in rise rate is needed to correlate with door velocity, simply increase the honeycomb cross section area by 20 percent. Aluminum honeycomb has a relatively constant crush force for a given section size. As it crushes, the force (and therefore the acceleration) on the dolly is relatively constant.

CONCLUSION

The side impact test device described in this paper has been used in over 1000 tests. It has been demonstrated to provide the high level of correlation and repeatability necessary to develop side impact air bag modules and demonstrate air bag system performance in the test conditions required. To date, this device has been used on CVC 12-inch HYGE sleds, 24-inch HYGE sleds, and on a Schenk servo-hydraulic sled using variations of the mounting scheme. It has been used to simulate FMVSS 214, EEVC, and lateral pole impact crash tests using the US-SID, BIOSID, and EUROSID side impact dummies. The HYBRID III 95th% frontal impact ATD has also been utilized for air bag integrity and bag coverage in lateral pole impact tests. Because the interior geometry of the vehicle is simulated, the setup is also used for static testing to evaluate air bag interaction with child and adult occupants.

ACKNOWLEDGMENTS

The author gratefully acknowledges the contributions made by Fred Peters, Randy Kelly, and Chad Ivan of Morton International, for their assistance and support in the development of this device.

970573

Intrusion Factors and Their Effects on Steering Column Movements During Vehicle's Frontal Impact Testing

Bahig Fileta and Xinxin Liu
Ford Motor Co.

ABSTRACT

Significant dashpanel intrusion is seen in some cars after severe frontal crashes at high speeds or after offset impact with rigid barrier or both. This intrusion may also result in severe steering column displacements and rotation. Knowledge of both responses is critical for designing an efficient vehicle front end that will respond well in crash. The intrusion has an effect on deciding the car front end length, while the column movements have an effect on the driver dummy's response. For reasons of developing efficiency and safety in vehicles and due to lack of published research, studies were conducted to understand the nature of the intrusion phenomenon as well as the mechanics of the steering column movement in the presence of intrusion.

This paper describes an experimental investigation on intrusion and steering column movements. The different factors causing intrusion modes are identified, and the contributions of these factors to the movement of the steering column are determined. A front part of a midsize front-wheel drive vehicle was used as the design for test structures. A crusher testing approach was used in accordance with a comprehensive test matrix that included eight different tests. In each test the dashpanel of the test structure was crushed at single or multiple points to generate a different intrusion mode, thus simulating the sources of motion that occurs in frontal barrier tests. Intrusion as well as movements of the steering column were among the significant data collected during these tests. The paper presents the results as well as comprehensive implications of the findings on evaluating the capacity of the vehicle structure and steering column support for resisting intrusion.

INTRODUCTION

During severe frontal barrier impact testing, some vehicles exhibit significant upward displacement and rotation of the steering column relative to the motion of the passenger compartment. Examples of the types of severe vehicles' frontal impact tests are full barrier impact tests at speeds higher than 56 kph and offset impact tests at speeds of 56-64 kph. This upward column displacement is a result of the increased intrusion into the vehicle's passenger compartment. At this point, the term intrusion may be inadequate to explain how it affects the movement of the column. But it is easy to show that during these tests there is an increase of demand on the vehicle's front structure to absorb more energy and consequently the vehicle requires a longer crush distance to stop. That increase in crush translates into added deformation of the passenger compartment structure and the dashpanel, toe-board, and cowl. It is the deformation of these closure panels, mainly by primary structural parts at the passenger compartment, that is referred to as intrusion in this paper. In addition, how this intrusion influences the upward displacement of the steering column will be investigated.

The above definition of intrusion is quite different from what was published in the literature. For example, the term intrusion has been used in References [1][1] and [2] to describe the displacements of the steering column inside the vehicle. In another reference [3], the authors indicated that intrusion is a response (primarily due to powertrain) that is responsible for moving interior surface, such as knee bolsters as well as the steering column, toward the dummy driver. Yet, in a different paper, the author used intrusion implicitly to refer to the deformation of the dashboard while studying the effect of steering column movement on the occupant [4]. But regardless of how the term intrusion was used in the past, it involved negative implication, which is not totally true. The front closure panels of the passenger compartment may have some deformation without an effect on the column movement, as will be shown later in this paper.

Besides the above mentioned severe impact testing, excessive intrusion and upward displacement of steering column may be also observed during the development of vehicles for standard impact testing. For example, a steering column may show excessive upward movement when crash testing a vehicle that is fitted with an alternative larger size powertrain or when testing a vehicle with heavier optional equipment before properly engineering the vehicle's energy absorption capabilities.

Although vertical displacement of the steering column is not restricted by FMVSS 204 or 208 (Federal Motor Vehicle Safety Standard), for years most vehicle safety engineers preferred designing vehicles with some limits on the upward displacement of the steering column during frontal impact testing. Some of the reasons are: (1) to allow the steering column to collapse axially when it is impacted by the torso of driver's dummy, (2) to

1. Numbers in parentheses designate references at end of paper.

provide a stable base for the airbag as it restraints the impacting dummy's head and chest, and (3) to decrease the possibility that the dummy's neck or head will be caught by a moving steering wheel rim. In fact, recognizing these benefits, the proposed European regulation on offset barrier testing of vehicles has a requirement to limit the upward displacement to a maximum of 80 mm [5].

To manage the upward movement of the steering column, vehicle engineers devised different approaches. One approach was to design the various components of the steering system with improved capacity in resisting the modes of intrusions into the passenger compartment. In other words, they focused on improving the component's strength to withstand the loads from the deforming dashpanel and primary structural members at the front section of the passenger compartment. Toward that goal, component's engineers have been experimenting with several new steering system designs and innovations including improved steering columns, collapsible steering drive-shaft, and modified column support to the vehicle structure.

Another approach that was used to manage the upward displacement of the steering column involved the fundamental design of the front structure and the crash energy management of the vehicle front end system. This approach, however, is difficult and involves the complex behavior of the primary structural members and their influence on the kinematics of the steering column during vehicle frontal crashes. Consequently, for the purpose of implementing this approach and providing design guidance to the vehicle crash engineers, basic information on and understanding of the interrelationship between the deformation of the primary structural members, the dashpanel, and steering system are required. Toward this goal and because of the lack of information about this topic in the research literature, an experimental study was conducted in which several vehicle's front ends were crushed by using Ford Energy Management System (crusher) methodology [6].

In addition to providing basic information that fills up the void in the intrusion phenomenon, one of the objectives of the study was to develop a methodology for identifying critical behavior of the structure and determining the strength of the steering system's design to resist dashpanel intrusion. Furthermore, it was of extreme importance to determine which mode of intrusion has the most significant effect on the upward displacement of the steering column.

Remember, this study was conducted for only one type of vehicle design: A mid-size vehicle that was driven at the front wheels by a powertrain, which was installed along the transverse direction of the vehicle (as opposed to a rear-wheel drive vehicle type with a powertrain installed along the longitudinal direction of the vehicle). The body design of this vehicle was a unitized construction type. The powertrain was fully mounted on a longitudinal type subframe that was bolted to the body structural members. Also in this vehicle, the steering column was supported by a bracket which was bolted to the cowl structure.

For the experimental study, a test structure, which consisted of a vehicle body equipped with few components (including the cowl-mounted steering column and surrounding parts), was crushed from the front at different places to simulate different types of intrusion. The tested structure was limited to the structure near the dashpanel and front part of the passenger compartment. The test structures were extensively instrumented to measure loads and displacements while testing, particularly the intrusion and steering column displacements.

This report describes the experimental study of the intrusion modes into the passenger compartment. The results showing the effect of each of these modes on the upward movement of the steering column are explained. Then, the intrusion mode that is most critical to column movement is discussed. These sections are followed by an example depicting the implication of the findings in developing a simplified experimental procedure that can be easily extended into analytical methods for assisting vehicle design. But first, a section is presented explaining the intrusion modes during vehicle frontal impact tests.

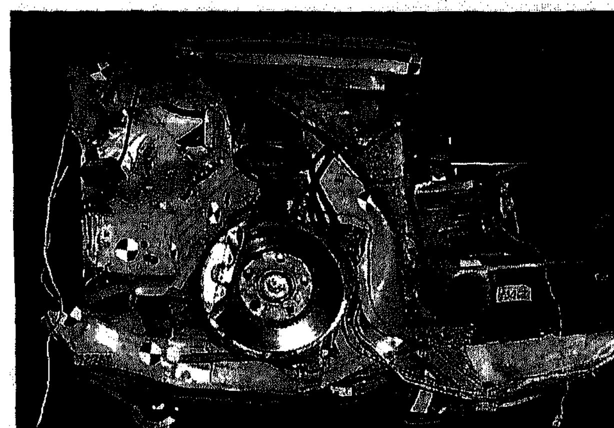
INTRUSION MODES

The way the primary front structure near the passenger compartment, and the panels such as dashpanel, cowl top and front floor deform during crash influences the intrusion mode of that vehicle. In fact, the vehicle's intrusion mode during frontal impact is highly dependent on the structural design concept of that vehicle. Therefore, to determine the possible intrusion modes in the vehicle type used in this study, films, photos and test data of crashed vehicles were reviewed and resulted in the following findings:

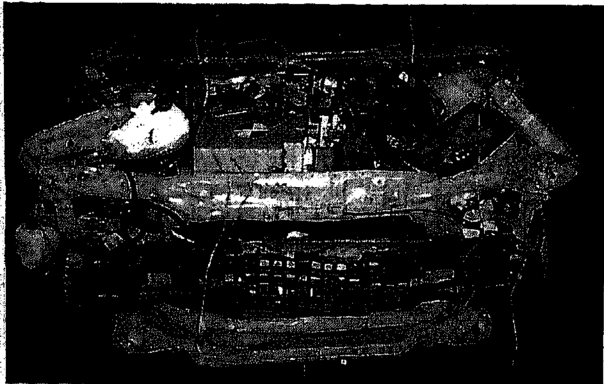
DEFORMATION AND COLLAPSE MODE OF PRIMARY STRUCTURE - A review of photos of a crashed vehicle (Figure 1a and Figure 1b below) shows that the mode of intrusion into the front space of the passenger compartment is influenced by the relative displacement of powertrain and deformation of primary structural members as follows:

- Crushing of powertrain crush into dashpanel and toe-board
- Bending of the aft-rail
- Crushing of the subframe into dash crossmember
- Bending of the sled runner
- Bending of the shotgun, twisting of the shocktower, and pushing at the cowl.

The final intrusion profile after impact is a result of the combined effects of crushing the above members. This profile was easily determined by implementing an intrusion mapping technique.



a: Side view showing collapse of aft-rail, subframe and sled runner



b: Front view showing collapse of shotgun

Figure 1: Collapse mode of the primary structural members after frontal barrier impact test

INTRUSION PROFILES OF CRASHED VEHICLES- As a way to determine the intrusion as well as the collapse mode of the vehicle structure near the dashpanel of a crashed vehicle, mapping plots are constructed using techniques developed at Ford. By these techniques, the deformed surface of the dashpanel and final position of the powertrain relative to the dashpanel are plotted and superimposed on the original position. Also shown on those plots are the initial and final relative positions of the steering column, brake pedal bracket, and other components that change the driver's space environment. Examples of these plots are shown in Figure 2.

The intrusion maps (also referred to as intrusion profiles) depict very detailed and reasonably accurate information about the behavior of the structure and components during crash. This high quality results are achieved because the plots are constructed by: (1) compiling information from design layout drawings, (2) measuring positions of several points on the tested vehicle before and after crash, (3) analyzing the motion of several points marked by targets using the crashed films, and (4) analyzing the kinematics of some selected points where accelerations were measured during crash. Intrusion profile plots show a section taken in a vertical plane that passes through the driver's center, front passenger's center or the vehicle's center.

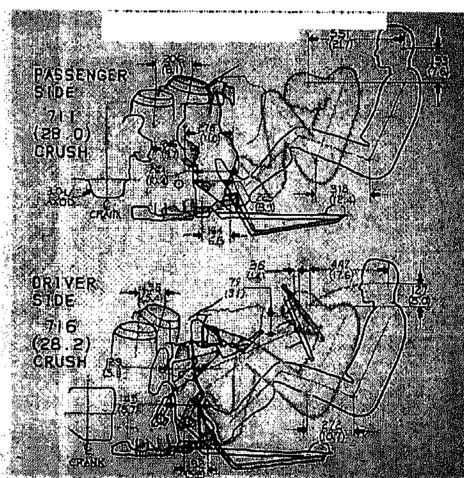


Figure 2: Intrusion profile and relative position of powertrain, shock-tower, steering column after frontal barrier impact test of a midsize front wheel drive vehicle

INTRUSION MODES AND STEERING COLUMN MOVEMENTS - By examining the intrusion profile plots shown in Figure 2, it can be easily seen that the deformed structures and panels move whatever components are mounted on them. For example, the brake pedal bracket moves as a rigid body and follows the motion of the dashpanel as it deforms. Similarly, when the cowl moves, the steering column support that is attached to it also moves. Profile plots also reveal that gaps between different components packaged under the instrument's panel and near the lower end of the column (the end that is near the dashpanel) may be closed and the components may be pushed by intrusion forces.

In fact, to determine the components that may be stacked between the dashpanel and steering column which can force it to move during intrusion, the packaging layout drawings of the vehicle interior and exterior near the steering column were examined. A compiled list of these components is shown in Table 1.

Table 1: List of component's packages that may offset steering column movement during vehicle frontal crashes

Item	Description
1	steering column and wheel installed on the vehicle structure
2	steering shaft
3	steering gear installed in the vehicle
4	brake pedal and bracket installed in the vehicle
5	other rigid components, which were not listed as secondary contributors to column movement are listed here because they influence the intrusion at cowl and upper dashpanel: windshield wiper motors, brake booster &housing, air conditioning evaporator core (inside instrument panel)

INTRUSION EXPERIMENT

The above brief introduction of the intrusion phenomenon and the column movements during frontal vehicle crashes is used as a background to explain the experimental study presented in this paper. The parameters of the experiment and the structural testing are described next.

FACTORS CAUSING STEERING COLUMN MOVEMENTS - A list of the factors that could have an effect on the movement of the steering column inside the vehicle was easily compiled from the background information given above such as: (1) the factors which may influence the shape of the intrusion mode while crushing the vehicle primary structure during frontal barrier impacts, and (2) the effects of the intrusion mode on the stacking of the components near the steering column and make it move during crash. Figure 3 shows this list of factors. The figure shows two major groups. The first group is on the outside of the ellipse and involves the vehicle's primary front structure and powertrain. Therefore they are referred in this report as primary factors or intrusion factors since they influence the intrusion mode shape in the vehicle.

The second group in the figure is given on inner ellipse, and the factors in this group are referred to as secondary factors. Some of these factors must exist to cause steering column

movement, but they are originally driven by the primary factors. Of course, various combinations and interactions between the various factors are additional possible causes of steering column movement during frontal crashes.

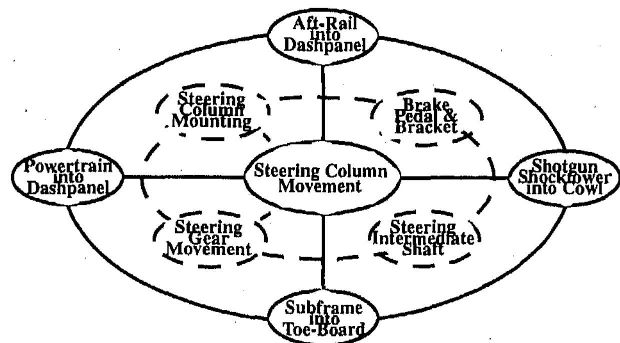


Figure 3: Primary and secondary factors causing steering column movement during vehicle frontal crashes

TEST MATRIX - Since the major objective of this study is to determine the effect of the different intrusion modes on the upward movement of the steering column, only the primary factors identified earlier were investigated. However, to actually see the effects on the column, all the secondary factors must be present. In other words, all the components listed in Table 1 were mounted on the tested structures.

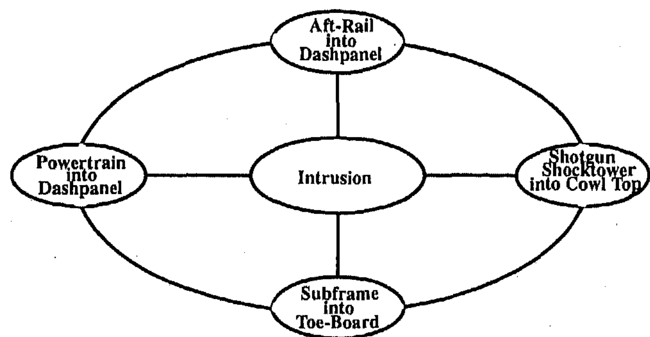


Figure 4a: Primary factors causing intrusion into passenger compartment front closure panels

Primary (Single) Factors	Combined Two Factors	Combined Three Factors
1 Rail into Dashpanel	1 + 2 Rail into Dashpanel + Subframe into Toe-Board	1 + 2 + 3 Complete aft-frame Structural System
2 Subframe into Toe-Board	1 + 3 Rail into Dashpanel + Shotgun/Shock into Cowl	
3 Shotgun/Shock Tower into Cowl	2 + 4 Subframe into Toe-Board + Powertrain into Dashpanel	
4 Powertrain into Dashpanel		

Figure 4b: Matrix of intrusion tests

Four primary (or main) factors were selected in order to study their effects separately. These are shown in Figure 4a and are also shown on Figure 4b as single factors. The interactions between these primary factors were selected by conducting tests which have the combined effects of two factors. Three cases were studied and shown in Figure4b (combined two factors). Finally, a test case was set up to study the effects of combining three factors, also shown in Figure 4b.

RESPONSES PLANNED FOR ANALYSIS - To analyze the effects of the primary factors mentioned above, the amount of intrusion in every test mode and the resulting steering column upward movement would constitute the most important responses to be measured during testing. However, for the needed understanding of the intrusion phenomenon as well as the mechanics of the steering column movement during intrusion, several key responses were also measured for additional analysis. These responses are discussed further in the experiment's instrumentation section which is discussed next.

TEST STRUCTURE- Literally, a passenger compartment was cut off a complete car. The body was cut off at the center of the shocktower from the front, and slightly rear of the B-pillar from the back end. Parts that had no effect on the test were removed. But primarily, the compartment left with the needed components for the test. These components are the parts listed in Table I. Additional modifications were made to this structure as described at the section on preparation of test structure and intrusion testing.

INSTRUMENTATION- Several unique measurements were taken while conducting the intrusion tests. Highlights of the more important ones include:

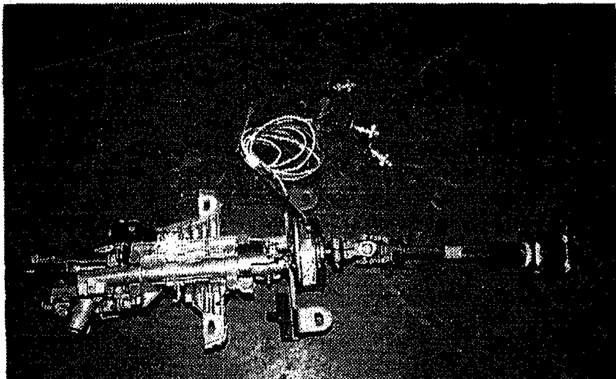


Figure 5: Loadcell at lower end of steering column to measure intrusion loads on steering column support system

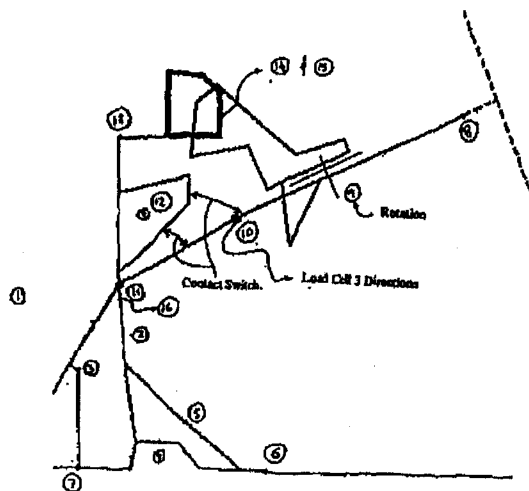


Figure 6: Instrumentation for intrusion tests

1. Steering Column Loads from Intrusion: One of the clever instrumentations used in the intrusion tests was the measurement of the loads imposed by the intrusion on the lower (also forward) end of the steering column (marked #10 in Figure 6). A 3-axis load cell, specially designed to be short, was attached (as shown in Figure 5), to the lower end of a modified steering column. The column modifications involved simple adjustments in the length of the column shaft to maintain the clearance space with the brake pedal bracket. Also, to signal the contact of the parts located under the instrument panel as they move with the dashpanel, an electrical switch was used (Figure 6). The signal from the switch helps in determining the intrusion that can occur without any effect on the column. The measured loads proved, as will be shown later, to be invaluable in explaining the interaction between the intrusion at the deformed dash panel and the moving steering column.

2. Steering Column's Displacements & Rotations - For capturing the kinematics of the column during crush, both ends (points marked # 8 and #10 in Figure 6) were instrumented by string pots to measure displacements in at least two directions. Additionally, an inclinometer was used as a check on the column rotation (point #9 in Figure 6) in a vertical plane.

3. Stack-up Between Dashpanel and Column - To monitor the progress of the contact between components and their stackup near the lower end of the column, string pots were used on some components under the cowl. These components were: brake pedal bracket, steering intermediate shaft, and steering gear. The measurements were taken at points marked by numbers 10, 11, 12, and 3 in Figure 6.

4. Intrusion Profile - To analyze the intrusion mode during crush, string pots were used at the cowl, toe-board, front floor, and dashpanel to record the deformation history. Measurements were taken at several points on these parts as shown in Figure 6. The results will allow plotting the intrusion profile (or mode) for each intrusion test.

5. Load- Crush Distance of Primary Structure - While conducting the crush tests, load cells were used to measure structural crush loads in addition to the crush distance (Figure 7). These load measurements were necessary to monitor the collapse of the structural members. The significance of knowing the collapse load is that its occurrence indicates the phase at which a large displacement and possibly of tear and fracture occurred in the test structure.

PREPARATION OF TEST STRUCTURE AND INTRUSION TESTING - At least one or more tests were conducted for each of the cases listed in the test matrix table shown in Figure 4b. Before each test, the structure in front of the dashpanel was modified by cutting the parts that are not used for loading. For example, in the case of crushing the aft-rail, the part of the apron, shotgun, shocktower and subframe were removed ahead the dashpanel's plane. This leaves only a small part of the structural member (406 mm to 508mm long) protruding at the dashpanel (see figure 7a). A steel tubular structure, approximately 1.5 meter long, was attached at one end of the member where the crush is desired. A steel end plate, 18 mm thick, was welded at the front of the member for that purpose. The other end of the loading structure was set to contact the crusher loading ram. The loading ram was equipped with load cells for measuring the intrusion loads during crush. Schematic of the set up for the four tests on primary (single) intrusion factors are shown in Figure 7a to 10. Additional schematic of set ups for tests on the two and three combined factors are shown in Figure 11 and 12. Actual photos of these test setups are given in Appendix A for reference.

In addition to modifying the front of the partial passenger compartment for proper loading, some welding was performed to restrain the body from moving during crush (Figure 7b). This was achieved by welding the rear edge of roof rails, roof panels, rockers, and floor panels to the test fixture. Also welded to the test fixture were plates that restrain the rocker at hinge pillar from moving. The hinge pillar, too, was held from moving rearward by welding two tubes, as shown in Figure 7b.

The intrusion tests were conducted by moving the ram toward the stationary test structure at a rate of approximately 50 mm per minute. The ram was stopped every two inches for observing the intrusion profile, the contact between stacked up components, the steering column, as well as for taking photographs. Another important reason for frequently stopping the test was to observe the crush mode of the primary structural member being crushed. To ensure a satisfactory test, this mode must be like the mode occurring in the actual crash test of the vehicle under investigation. To illustrate this point, the collapse mode of the shotgun during intrusion testing shown in Figure 13 was triggered to be the same as the mode shown in Figure 1b, the mode occurring in the vehicle crash test. A similar example can also be shown for the collapse mode of the sledrunner, particularly in the case of intrusion by subframe loading.

All test data from instruments were recorded electronically and stored on disc files for processing after the test. However, while running the test, data from four channels (crush loads, crush distance, and steering column displacements) were monitored on a screen to observe any unusual behavior and ensure satisfactory progression of the test.

Most of the conducted tests were stopped between 305 mm to 406 mm of intrusion or of deformation at the dashpanel, depending on the extent of damage or tear occurring at the thin walled panels and the spotwelds. Photographs were taken of the test structure and column before and after completion of the test. These photographs are shown in Appendix-A.

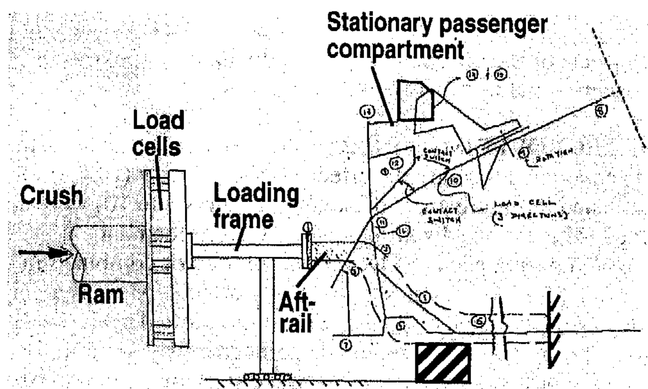


Figure 7a: Typical set-up of intrusion tests - shown aft-rail crush into dashpanel

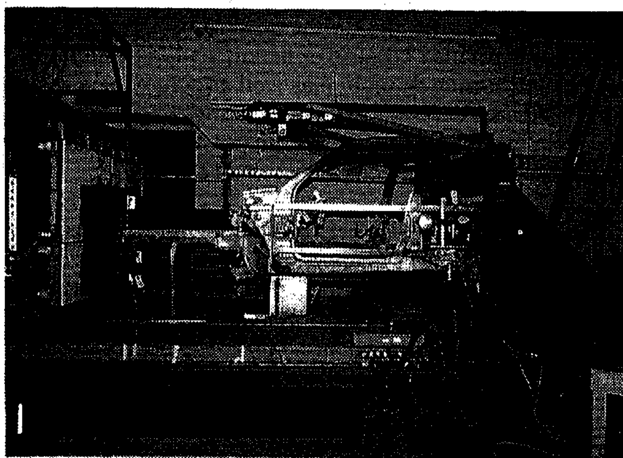


Figure 7b: Typical tie-down of partial passenger compartment structure to test fixture

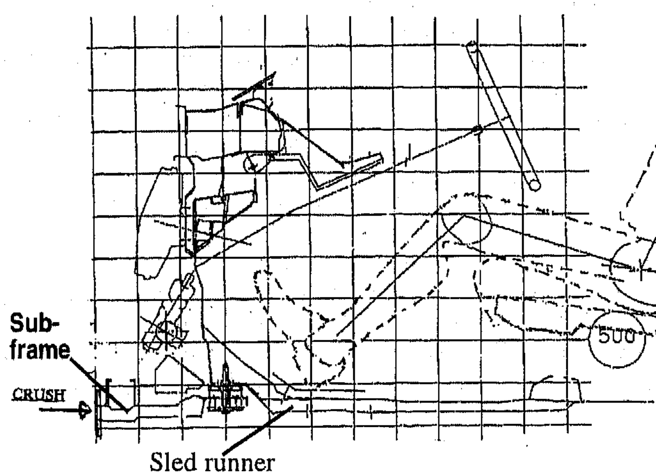
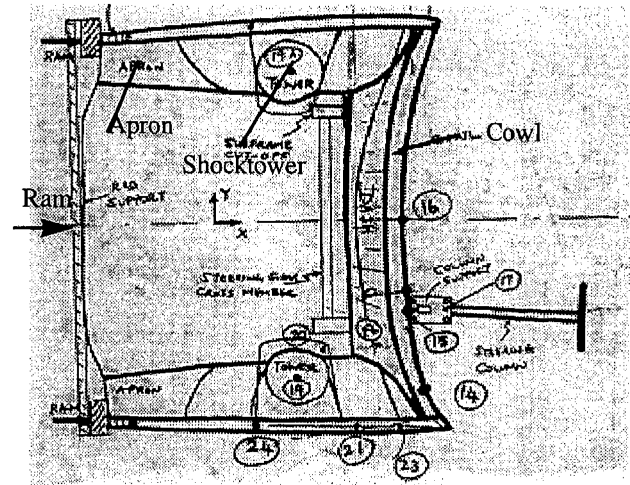
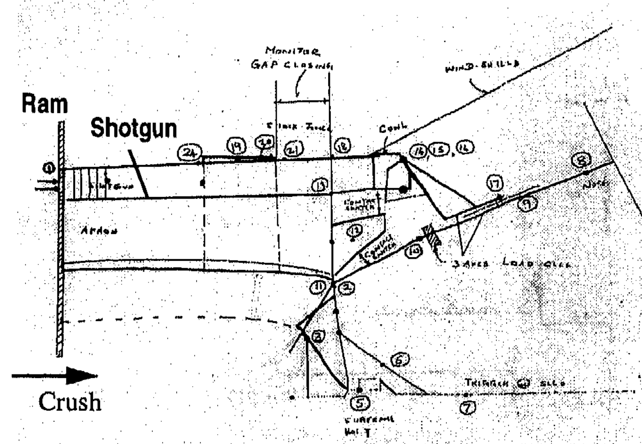


Figure 8: Subframe crush into toe-board crusher test



a. Plan view



b. Side view

Figure 9: Shotgun/shocktower crush into upper dashpanel/cowl test set-up

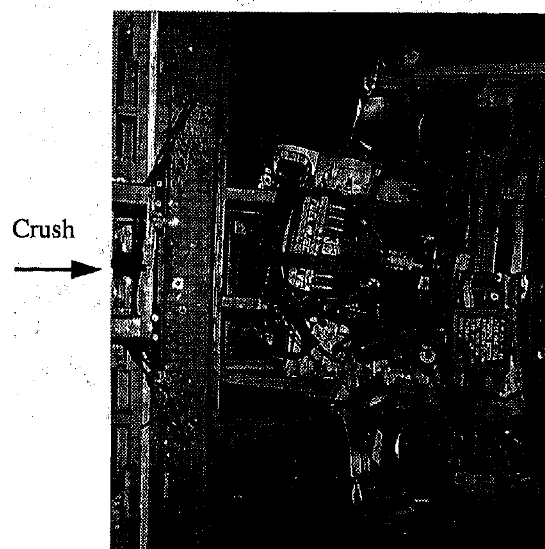


Figure 10: Test set-up for powertrain crush into dashpanel

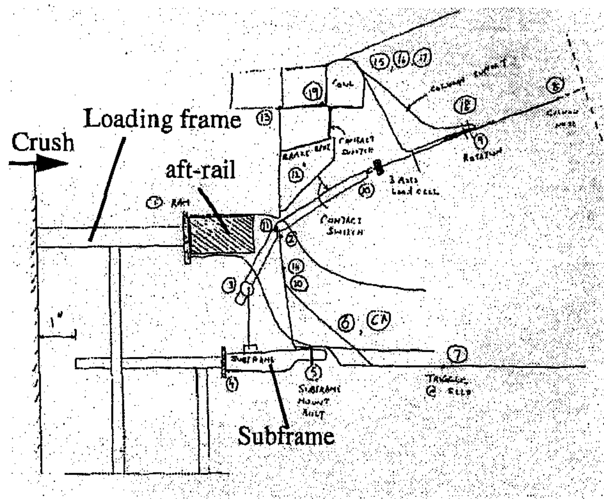


Figure 11: Loading setup for combined two factors intrusion test - shown crushing at aft-rail and subframe

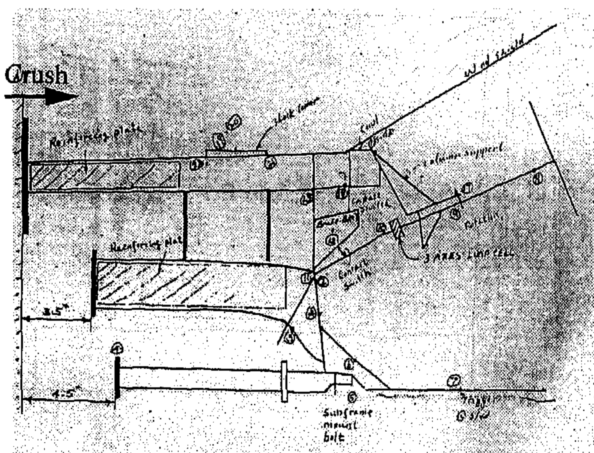


Figure 12: Schematic on setup for loading the front of the passenger compartment by the three primary structural members

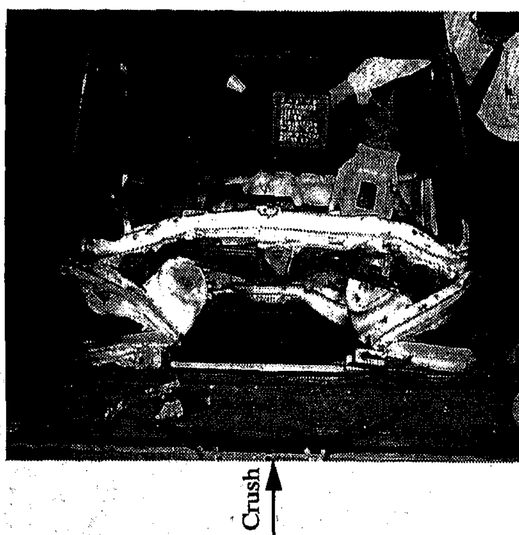


Figure 13: Shotgun and cowl crush modes after crushing shotgun and shocktower into cowl and upper dash intrusion test

TEST RESULTS

The data collected from more than ten intrusion tests conducted during the course of this investigation were used to calculate the effects of the different factors on the upward movement of the steering column as follows:

EFFECT OF PRIMARY FACTORS -Results of the intrusion tests conducted to study the effects of the primary (single) factors (listed in Figure 4b, and shown in Figures 7 to 10) were analyzed. The various displacements as well as load data measured during each test were cross plotted or combined in a way to determine any unusual or strong effects. Also, as a part of this analysis, some responses were cross plotted against a displacement of a common reference point as a way to normalize the data. For instance, the results on the steering column movements from various tests were plotted against the corresponding displacement of dashpanel near rail (point #2 in Figure 6) as a common variable. In addition, the results of these tests on the displacement of closure panels (dashpanel, cowl top, toe-board, front floor), were used to plot the deformation mode.

In this section, some results from only the structural intrusion tests will be presented. As for the powertrain intrusion test, no data will be provided because of the difficulties which were found in controlling its motion as well as achieving a representative dashpanel deformation while testing. Explanation is given later on, and an alternative test on powertrain intrusion is described under the section of combined factors.

Structural Intrusion Tests -In each of the three intrusion tests on the structural primary factors, the steering column moved differently. During the aft-rail crush into the dashpanel, the column was steadily rotating as well as moving upward while crushing the aft-rail into the dashpanel. The appreciable sensitivity in the column movement for each small increment of aft-rail crush was a classic cause and effect behavior, especially after the first few inches of crush. Figure 15 shows clearly this effect in that the steering column was almost vertical after 406 mm of aft-rail crush.

Contrasted with this excessive upward displacement and rotation noted above was the downward movement of the column while crushing the shotgun and shocktower into the cowl and upper dashpanel (Figure 9). Yet, there was another behavior of the column that fell between the extreme upward and downward motions. The limited and slow upward movement of the steering column were the results of the intrusion test in which the subframe was crushed into the toe-board. All these three results are shown in Figure 14. The vertical movements of the steering column were plotted against the longitudinal displacements measured at a common point in the three tests, a point at the dashpanel near the joint with aft-rail. It may be noted from Figure 14 that a small rearward displacement of the dashpanel at the rail is an indication of the weak effect between that particular intrusion factor and the middle area of the dashpanel.

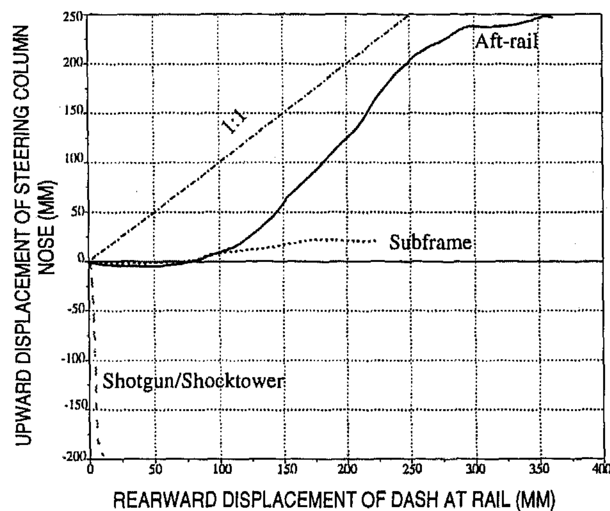


Figure 14: Effect of primary factors on the upward displacement of steering column

The crush mode (as the intrusion profile) for each of these three tests was plotted from the displacement results. The deformation (or mode) plots were found to be very effective in depicting the main characteristics of the behavior, as well as in explaining the intrusion effects on the steering column movements. Beginning with the aft-rail intrusion test, the intrusion mode is shown in Figure 16 after 200 mm and 400 mm of crush. Additional profile from that test was plotted from the results after 406 mm (Figure 17) illustrating the behavior of the column at an extreme intrusion condition. The important feature in this deformation mode is the membrane effect induced at the parts of the dashpanel alone and below the line of crush. The panel membrane forces pulled down the cowl from the top and pulled up the cross member reinforcing the toe-board from the bottom.

The next intrusion mode (shown in Figure 18) illustrated the effect of crushing the shotgun/shocktower into the cowl and upper dashpanel. This mode is distinguished by bending the upper dashpanel and cowl above the joint of aft-rail. It was observed from the shape of this deformation that there was not adequate stiffness in the upper cowl structure to pull the dashpanel below the joint with the aft-rail. The rearward displacement of the dashpanel at rail in this test was very small.

Continuing with the results on intrusion modes, the intrusion profile due to crushing the subframe into the toe-board was plotted and given in Figure 19. The effect of this primary factor was weak as is evident by the position of the steering column after 356 mm. Again the structural stiffness (in this case the toe-board cross member and underbody reinforcement sled runner), was not high enough to produce the necessary forces to push the column. The upward displacement of the steering column and the rearward displacement of the dashpanel at rail were small, as shown in Figure 14.

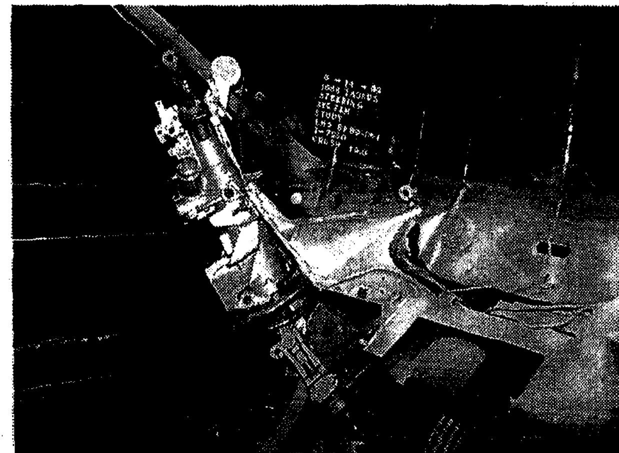


Figure 15: Excessive column upward displacement and rotation shown by the final position after 406 mm of crush of aft-rail into dash-panel

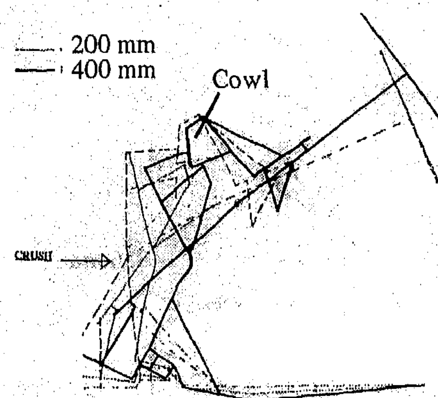


Figure 16: Deformation made after 100 mm and 200 mm of at-rail crush into the dashpanel intrusion test

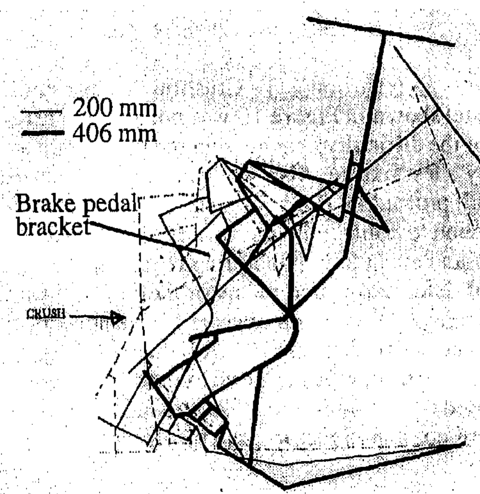


Figure 17: Intrusion profile and column position after 200 mm and 406 mm of aft-rail crush into the dashpanel intrusion test

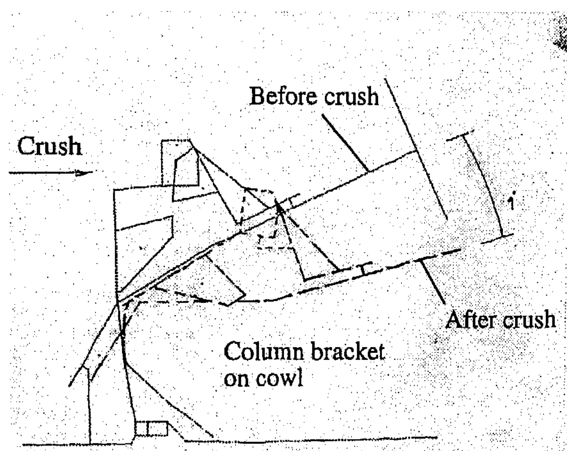


Figure 18: Effect of cowl and upper dashpanel intrusion on the steering column movement

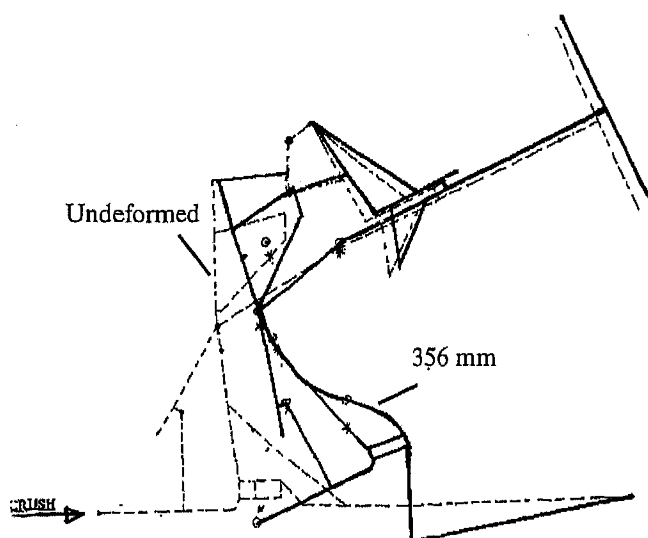


Figure 19: Deformation mode after 256 mm of subframe crush into toe-board intrusion test

Powertrain Intrusion test - Crushing the powertrain into the dashpanel shown in Figure 10 was not acceptable test set up because of the difficulty in achieving the powertrain motion by using only a test fixture. Also while conducting that test, some of the small protruding rigid parts on the engine and transmission were found to contact the dashpanel surface and caused local denting and eventually puncturing that surface. Because the mode of deformation and the tearing of the dashpanel were different from what has been observed in vehicle crash tests, this test was considered unrealistic. A better way of showing the effect of powertrain crush relative to dash is to push on it as it is positioned on the subframe. This alternative is discussed under effects of combined factors, and the set up is shown in Figure 21.

EFFECT OF COMBINED FACTORS - The results of intrusion tests to study the effects of two factors and three factors were analyzed similar to the primary (single) factor analysis. The effects are shown in Figure 20. It is very obvious from this figure that the strong effect of the primary aft-rail intrusion factor on the upward displacement of the steering column was also dominant in the combined effects that involve the aft-rail.

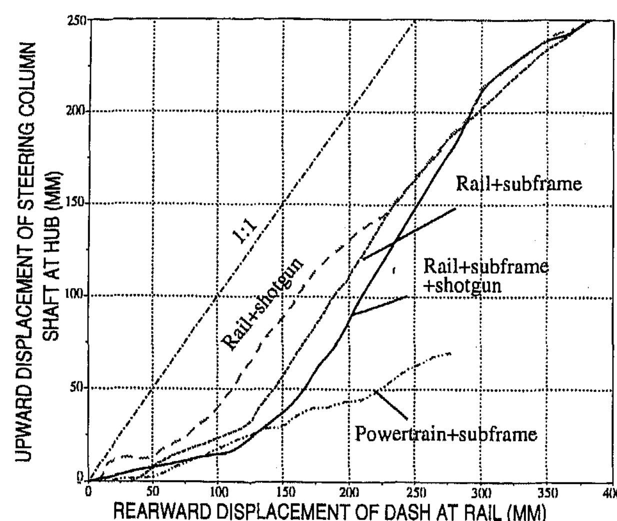


Figure 20: Effect of combined factors of intrusion on the upward displacement of steering column

However, the effect of crushing the powertrain which was mounted on the subframe, was not as strong as the structural intrusion on the upward displacement of the steering column (Figure 20). Appendix A shows the powertrain position relative to the dashpanel before and after crash. A significant rearward displacement of the dashpanel at rail, 275 mm but resulted only in a 70 mm upward displacement of the column (see Figure 20).

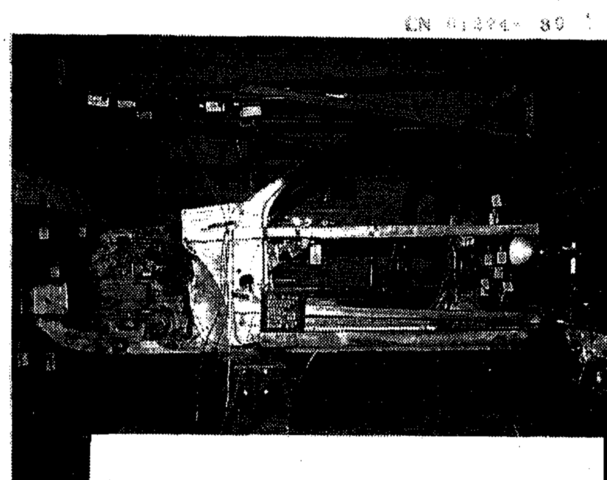


Figure 21: Powertrain intrusion into dashpanel new test setup

DISCUSSION

DOMINANT INTRUSION MODE - The intrusion tests discussed above showed very interesting and surprising results. For example, by investigating the effects of the primary factors, crushing at the aft-rail revealed a very strong effect on the upward displacement of the steering column. This was a surprise finding because of two reasons. First, this finding disproves the common perception among vehicle safety engineers that the

powertrain is the main reason of intrusion, in that it pushes on the dashpanel, and hence becomes the principal driver of the steering column movement. Figure 20 shows the weak effect of powertrain crush on the upward column displacement compared with aft-rail crush. Second, the strong effect when crushing at the aft-rail shows on Figure 14. The finding is a surprise because by crushing only at the aft-rail, the response of the steering column was magnified by the resulting intrusion mode. This finding was evidenced by the steering column's displacement results in this single case, which exceeded those occurred when crushing at all the structural factors simultaneously (Figure 20). In fact, when combined with other factors, the aft- rail crush dominated and overpowered the other effects (also shown on Figure 20).

The strong effect of the intrusion mode which was generated during aft-rail crush can be explained by a closer examination of the test results. Specifically, the mechanics of loading the steering column was analyzed to the extent of reaching a large displacement and loss of stability stage. Our methodology involved investigating the roll of the components' stack-up near the lower end of the column and its effect on the nature of loads as well as displacements at that end of the column. With this understanding as a background, the loads and deflections at the lower end of the column measured while testing with the different primary intrusion modes are presented and compared.

MECHANICS OF LOADING THE STEERING COLUMN DURING AFT-RAIL CRUSH - The reasons behind this strong effect can be explained by first investigating the effect of the intrusion mode on the stack up near the lower end of the column, i.e, secondary factor.

Roll of Secondary Factors on Loading the Column During Intrusion - To determine the effect of the secondary factors on the way the steering column is loaded during the aft- rail intrusion test, the intrusion profile plots shown on Figure 17 from the intrusion test and Figure 2 from the actual vehicle crash test are examined. In both plots, the position of the brake pedal bracket is rotated and moved as the upper part of the dashpanel rotates and moves. The similarity of the motion in both the vehicle crash test and crusher test indicates the validity of the simulation by the crusher method. This motion of the brake pedal bracket resulted in closing any gap that existed between it and the lower end of the steering column. Furthermore, as the crush and intrusion continued the brake pedal bracket and the steering intermediate shaft contacted and pushed the end of the column rearward. The events after the load on the column continued to build-up are divided into four stages as described below in this report. The photograph in Figure 22a shows good view of the brake pedal bracket pushed against the lower end of the steering column.

To validate the critical effect of the brake pedal bracket and the steering intermediate shaft on the upward displacement of the steering column, an aft-rail intrusion test was conducted after removing these two parts. The result of that test was compared with the results of a similar test before removing the parts and shown in Figure 22b. The figure shows a negligible upward displacement of the column after brake pedal bracket and the steering intermediate shaft were removed a definite proof of their effect in loading the column during an aft-rail intrusion test.

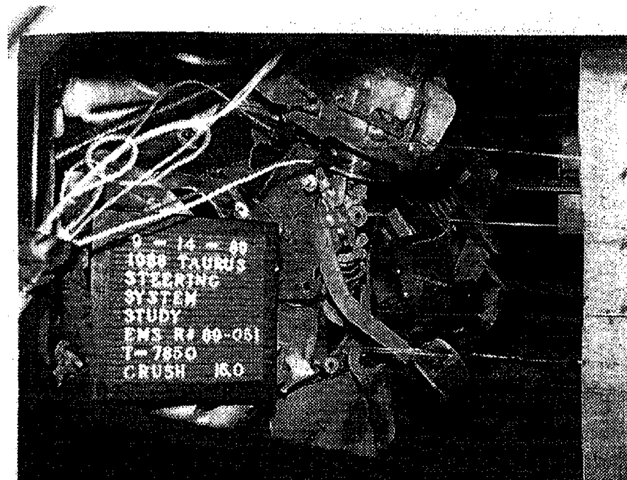


Figure 22a: The brake pedal bracket is pushed against the lower end of the steering column during an aft-rail intrusion test - shown at 406 mm crush

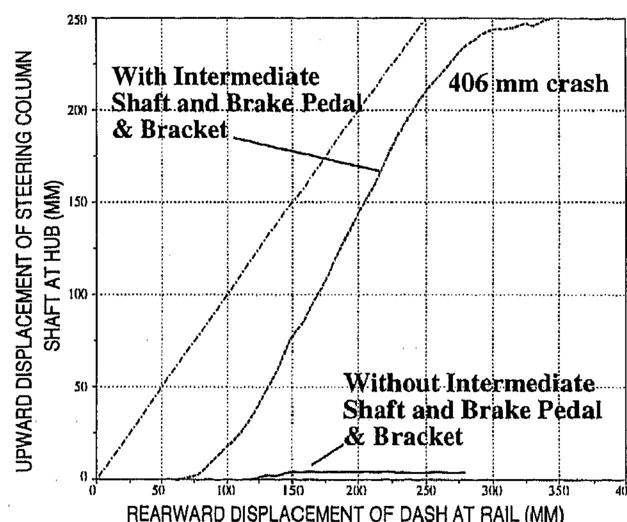


Figure 22b: Effect of brake pedal bracket on the upward column movement of the steering column

Stages of Loading Steering Column during Intrusion - We then examined the load deflection at the lower end of the column and identified the different stages in column loading. Referring to Figure 23, the loading and movement of the steering column can be divided into four stages described as follows:

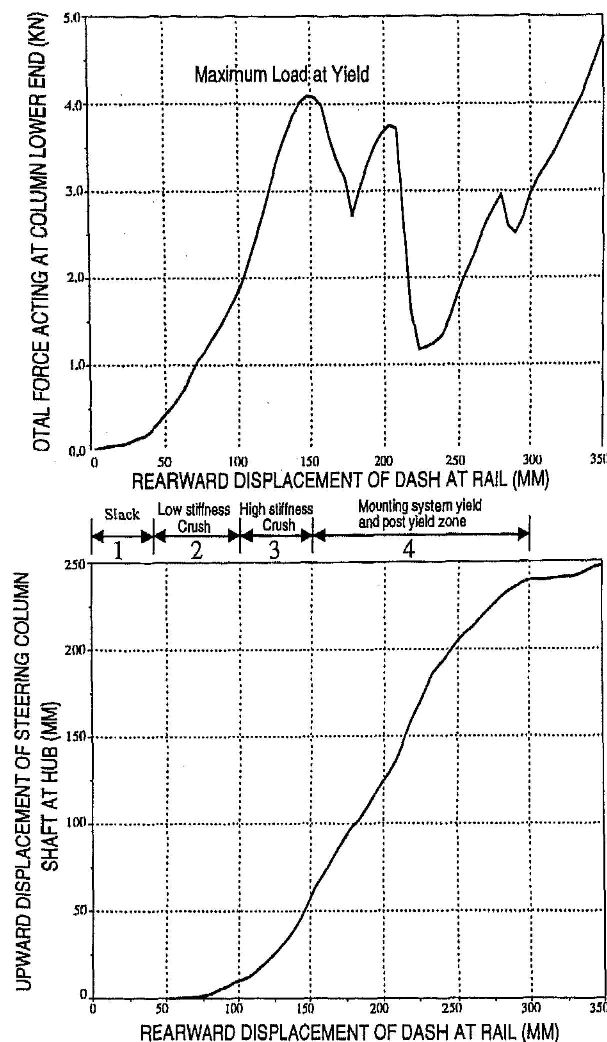


Figure 23: Different stages of column loading during crush

1. Slack- It is the rearward displacement of the dashpanel at rail during which the gaps or clearances in the package under the instrument panel are closed. During this motion there are no forces developed at the lower end of the column to cause any movement. For this investigation, the tested design showed 38 mm of slack.

2. Low stiffness crush - During this stage, as the dashpanel moves rearward from 38 mm to 89 mm, contact and local crush take place between soft edges of the components packaged between the dash and lower end of column. The load on the column builds up gradually (up to 1.33 KN), but the magnitude is still too small to cause any noticeable movement of the column.

3. High stiffness crush to peak load - Crushing the dashpanel from 89 mm to 152 mm causes the load at the lower end of the column to increase rapidly to 4.89 KN lbs with small increase in upward column displacement (0 to 50 mm)

4. Post peak load and large column displacement - crushing the dashpanel beyond 152 mm causes the column to move upward at a rate of approximately one-to-one.

EFFECT OF INTRUSION FACTORS ON THE STEERING COLUMN LOADS - The loads and displacements at the lower end of the column due to these different intrusion factors were examined and the results are given in Table 2. Shear loads are given in the table to show their effect on the moment acting on the support structure. A downward shear has the effect of adding the bending moment from the shear to the bending from the axial compression, while an upward shear has the effecting of substracting the bending due to shear from the bending due to the axial compression. The results clearly indicate that intrusion mode by the aft-rail has the capacity to load the lower end of steering column by a high load that exceeded the load capacity of the support system and made it yield. However, in the other cases, the loads generated at the lower end of the column were not high enough to weaken the support and move into the large displacement stage.

Table 2: Effect of intrusion factors on the resultant loads measured at the lower end of the steering column

Intrusion Factor	Initial peak load - KN		
	Axial	Shear[a]	Resultant[b]
aft-rail	4.67	1.82	5.01
powertrain & subframe	3.38	1.67	3.77
shotgun & shocktower	1.6	1.69	2.33

a. in a vertical plane
b. in a vertical plane

SUMMARY AND IMPORTANT FINDINGS

Based on these intrusion tests conducted on a partial passenger compartment structure of a medium size front wheel drive vehicle equipped with a cowl bracket type design to support the steering, the significant findings are summarized as follows:

1. The intrusion caused by crushing the aft-rail into the dashpanel had the strongest effect on the upward displacement of the steering column. Because of the stiffness and strength of the combined stretched dashpanel and aft-rail, this intrusion mode had the capacity to force the brake pedal bracket and steering intermediate shaft and consequently pushed the column to the point of yielding its support.

2. The intrusion caused by crushing a powertrain mounted on the subframe proved to have a significantly weaker effect than the effect of aft-rail intrusion on the upward displacement of the steering column.

3. The intrusion caused by simultaneously crushing at the aft-rail and at another primary structure (case of combined two actors shown in Figure 4b) was found to have stronger effect on the upward displacement of the steering column than the effect of the intrusion caused by the combined three structural factors.

4. The column support structure and the components packaged

under the cowl (near the lower end of the steering column) are critical factors in influencing the effects of intrusion.

APPLICATION EXAMPLE

To demonstrate the benefits and power of the key finding described above, an evaluation of alternative design concepts of steering column support was made. In other words, based on the finding that the intrusion mode from the aft-rail crush into the dashpanel is the most important factor causing the upward displacement and rotation of the steering column, this test case was used as a method for evaluating different designs of column support. These designs were: a cross-car beam that has a uniform cross-section, a cross-car beam that has a variable cross-section, and lastly a cantilevered half-car beam. The hardware representative of these designs was installed in the passenger compartment test structure as shown in Figures 24, 25 and 26 respectively. The base-line cowl mounted steering column system is shown in Figure 27 for reference.

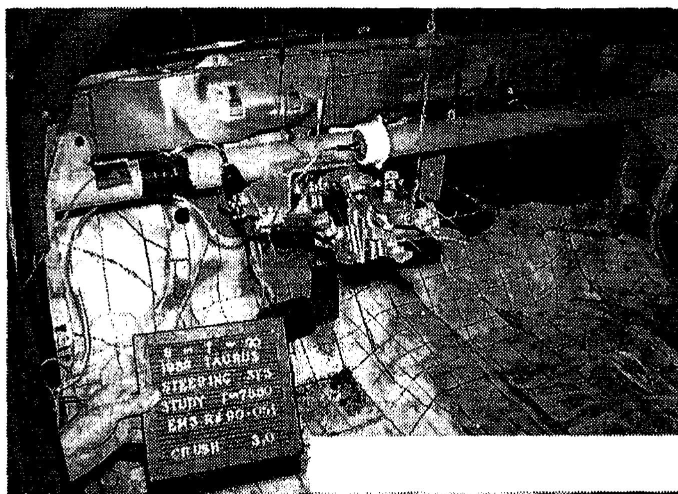


Figure 24: An experimental uniform section cross-car beam column support system

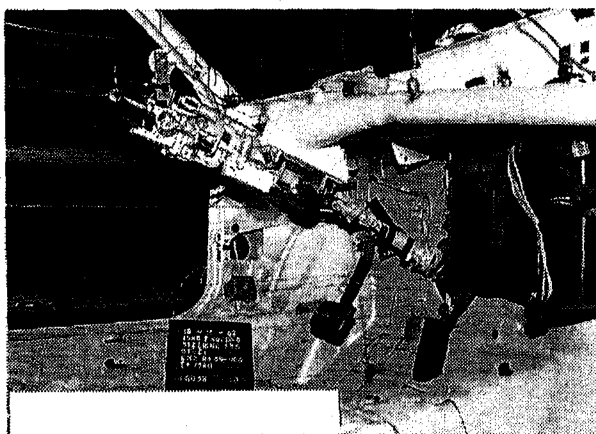


Figure 25: An experimental variable section cross-car beam column support system

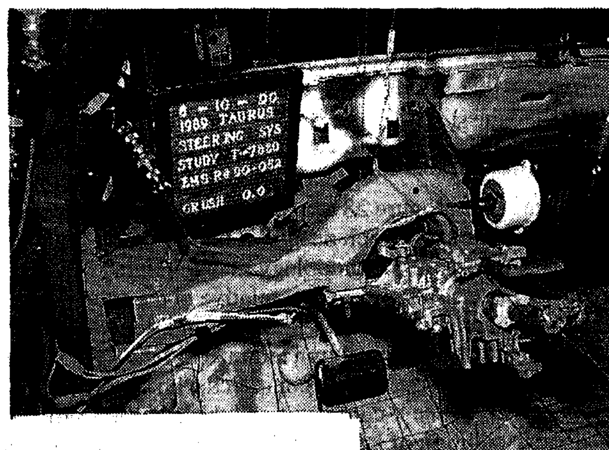


Figure 26: An experimental half-car beam column support system



Figure 27: An experimental cowl mounted column support system

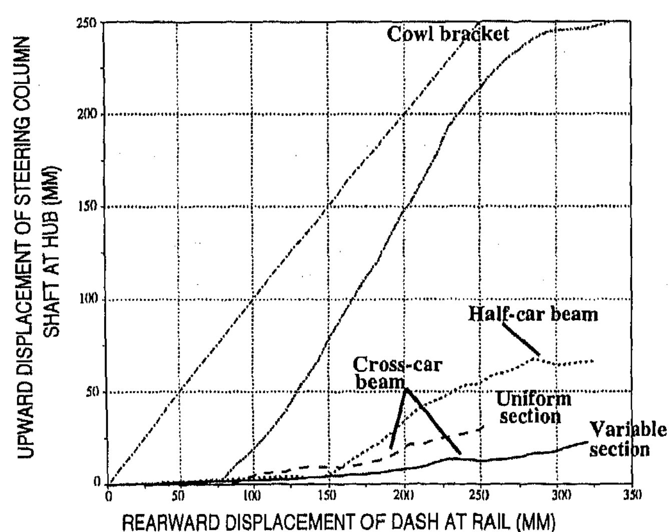


Figure 28: Comparison between the steering column upward displacement during the aft-rail intrusion testing of different column support designs

The test methodology of aft-rail crush into dashpanel was used for each of the three column support designs. The results of these tests showed the improvement in the sensitivity of these designs to the steering column's upward displacement as the intrusion at the dashpanel was increased (Figure 28).

The results in Figure 28 are very informative and provide the needed evaluation of each design. It illustrates the capability of aft-rail intrusion test methodology in highlighting the advantages of alternate column supports relative to the cowl mounted support. For example, at 254 mm of aft-rail longitudinal displacement the steering column mounted on the cowl moves upward approximately 200 mm compared with 12 mm and 38 mm for the cases off the x-car beams and the half-car beams respectively. In addition of knowing the column displacement at the expected intrusion in the vehicle, from the information in Figure 28 a guideline on the intrusion limit at which the column has a negligible upward movement can be established. For example, the alternative column support design can actually allow more intrusion without influencing the upward column displacement. Figure 28 shows that the benefit of the alternative column design can be translated into sustaining additional 100 mm of intrusion in that vehicle without seeing any effect on the upward column displacement. Historically, this consideration, (which can result in shortening the car's front end and making it more efficient) has traditionally not been included in studies that are conducted during early phases of vehicle development.

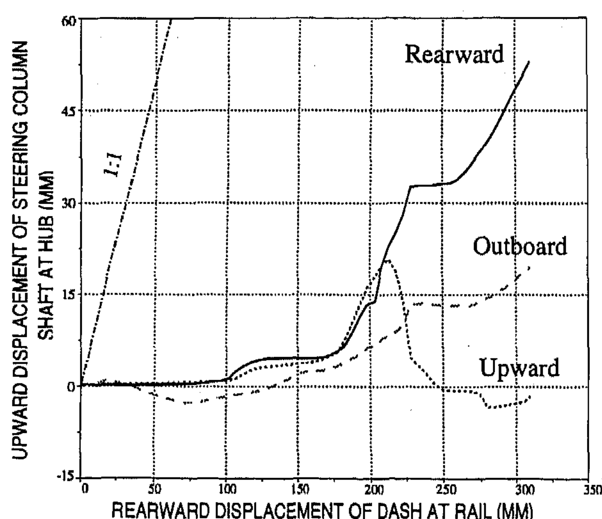


Figure 29: Steering column displacement at hub during aft-rail intrusion test of a partial passenger compartment equipped with cowl bracket column support

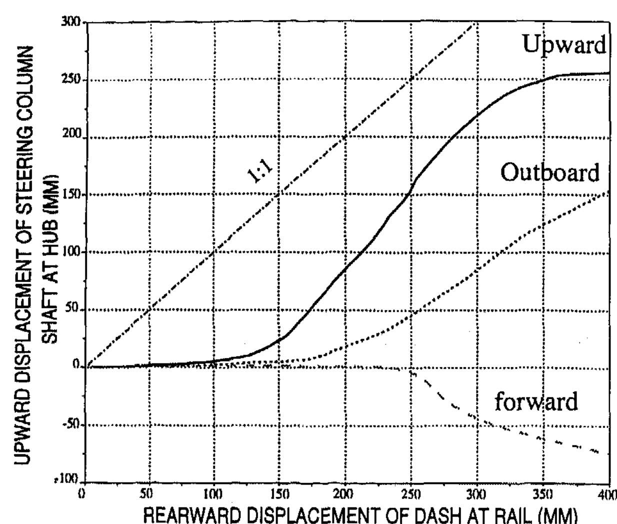


Figure 30: Steering column displacement at hub during aft-rail intrusion test of a partial passenger compartment equipped with cross-car beam support

Although the emphasis in this study is on the upward displacement of the steering column, the aft-rail intrusion testing proved to be valuable in showing the effectiveness of the alternative support designs in containing the movement of the column in other direction. Comparing the results (shown on Figures 29 and 30) of testing a cross-car beam support with the cowl mounted support column is clear indication of significant drop in the column movement in the longitudinal and lateral directions as well as the vertical direction mentioned earlier.

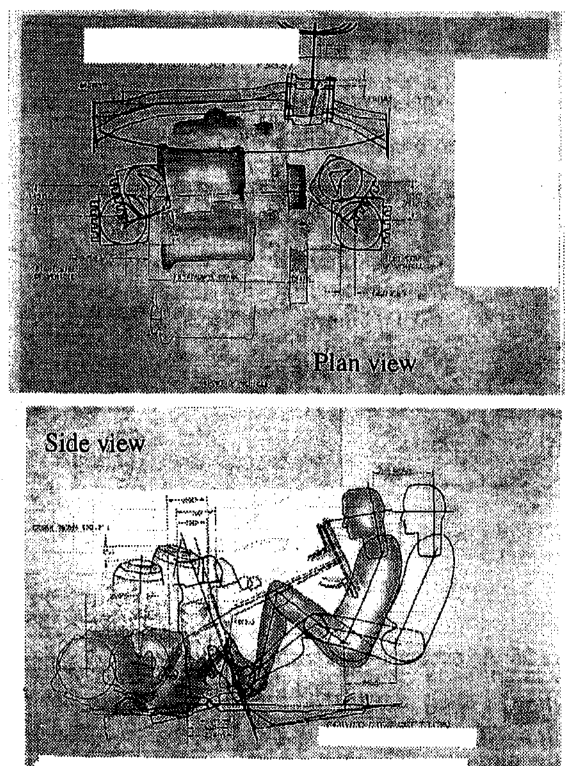


Figure 31: Intrusion profile and steering column displacement after frontal crash test of a midsize vehicle equipped with experimental cross-car beam column support

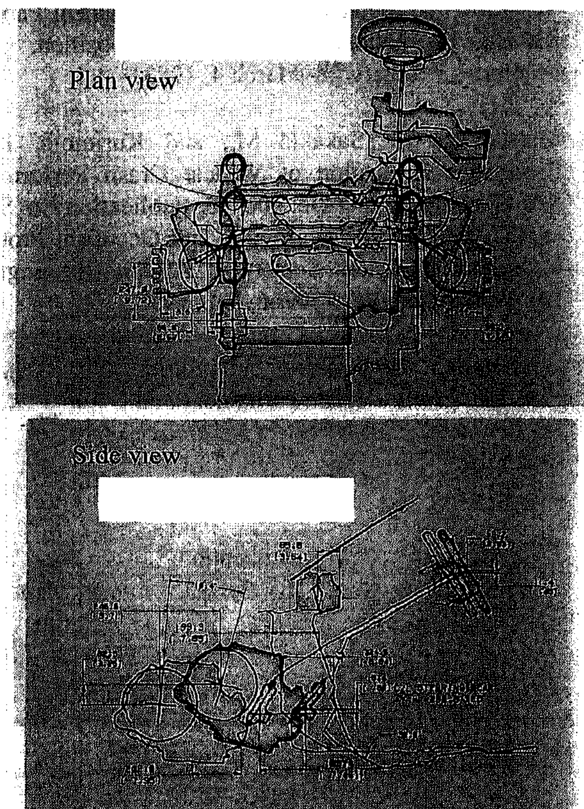


Figure 32: Intrusion profile and steering column displacement after frontal crash test of a midsize vehicle equipped with experimental half-car beam column support system

To validate the effectiveness of these proposed designs, the results of crash tests of vehicles that were equipped with a cross-car beam and a half-car beam were analyzed. The intrusion profiles and steering column displacements were plotted and shown in Figures 31 and 32. As expected from the results of the crusher testing methodology, the column was very stable, with approximately 25 mm of upward movement at a dashpanel intrusion of approximately 246 mm for the cross car beam design. Similarly, the behavior of the column with half-car beam was validated (Figure 28).

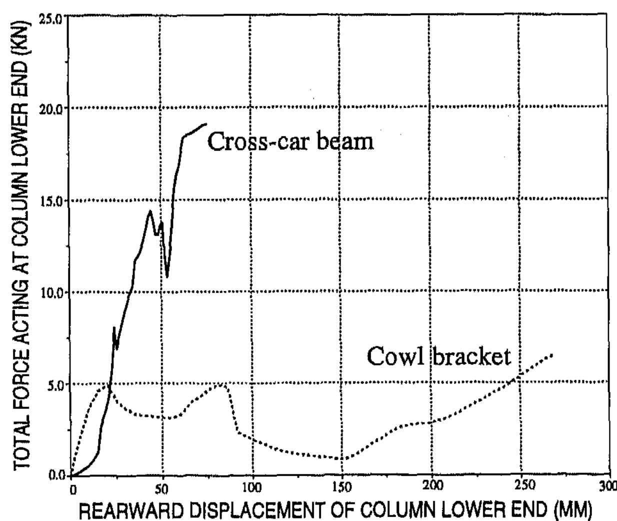


Figure 33: Steering column loads vs. deflection measured at lower end of column for two different support systems

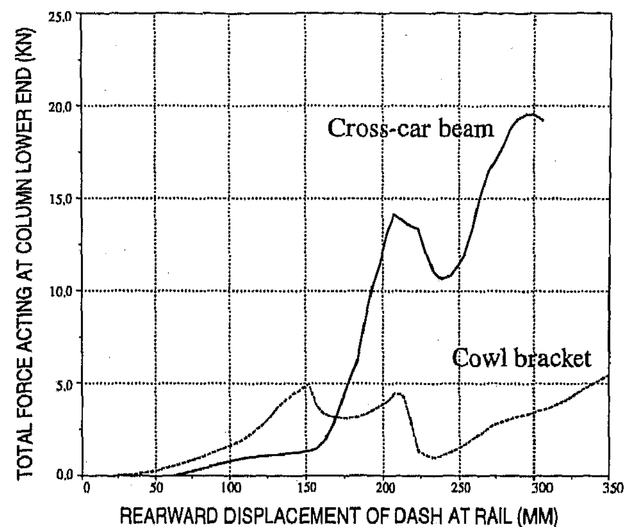


Figure 34: Steering column loads for two support systems

In addition to proving a design or helping to select a design, the aft-rail intrusion testing methodology provides information that can serve as a target to size and optimize the column support design. This unique tool as a design aid can be explained by reviewing the results of the loads produced at the lower end of the column given in Figures 33 and 34. The curves in Figure 33 present the load- deflection at the lower end of the column during the testing of cross-car beam and cowl bracket column support systems. Obviously, the higher load required to collapse the cross-car beam system (14.23 KN relative to 4.89 KN for the cowl bracket system), is the characteristic for the salient improvements in the column displacement. However, to reach the collapse load in each system, the amount of dashpanel intrusion is approximately 216 mm for the cross-car beam system, compared with only 152 mm for the cowl bracket system. That is a significant gain of 64 mm in intrusion potential that has no effect on the column displacement. The comparison in the performance characteristics can be shown in the form of the four different stages of column loading as shown in Table 3.

Table 3: Dashpanel intrusion at aft-rail and corresponding loads at different stages of column loading for two different support systems

Stage[a]	Cowl bracket		Cross-car beam	
	Intrusion mm	Column load-KN	Intrusion mm	Column load-KN
1	0-38		76	
2	38-89	1.3	76-165	1.3
3	89-152	4.9	165-206	14.2
4	>152		>206	

a. see Figure 23

CONCLUSIONS

The primary intrusion modes and how they combine to form the final intrusion mode or profile (which occurs in frontal

crashes of vehicles) has been explained. Also the role of each of these modes on causing the movement of the steering column has been clearly illustrated. A simple crush test configuration, aft-rail crush into dashpanel, was shown to be a powerful simple methodology in guiding vehicle engineers in making decisions on intrusion limits and management of the steering column movement as well as energy absorption of the vehicle.

This simplified test methodology could actually eliminate the need to conduct a vehicle crash test regarding amount of intrusion and selection of steering system design concepts. The methodology was shown to evaluate alternative concept designs of steering columns and its support structure. Many meaningful as well as quantitative performance characteristics of these alternative design proposals became available by conducting this test. Important examples of these evaluative performances include the capacity of the design to resist or tolerate intrusion, and the amount of intrusion at the dashpanel that can occur with only negligible steering column movement. These evaluations can serve as input to vehicle engineering when making design decisions during early phases of vehicle development.

The simplified test does not have to be a physical hardware test, because the case of aft-rail crush into dashpanel can be easily simulated using CAE crash codes. It is certainly simple and more economical to run these tests on the computer than simulating the complete vehicle.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. P. Prasad for sponsoring and supporting the research presented in this paper. Also, we acknowledged the contribution of the late Mr. Jerry Fritzler in planing of the experiments and in providing most of the hardware used in the testing. In addition, the authors are grateful to the crusher testing staff at Ford Motor Company for being so patient in accommodating our requests while conducting the experiments discussed in this report. In particular, we thank Mr. Steve Upmeyer, who helped in acquiring very valuable data by providing new and extensive instrumentation. Moreover, we wish to thank Mr. K. Presulmalli and Dr. Y. Lin for supervising several of these tests and conducting extensive analysis of the test data. Finally, we acknowledge Dr. E. Hara and Dr. C Chou for their editorial comments and advice.

REFERENCES

[1] Ragland, C. and Klemer, G.,"Intrusion Effects on Steering Assembly Performance in Frontal Crash Testing," Proceedings of the Twelfth International Technical Conference on Experimental Safety Vehicles Vol. 1, P.358-387, Sweden, 1989.

[2] Hackney, R.J. et al, "Analysis of Frontal Crash Safety Performance of Passenger Cars, Light Trucks and Vans and Outline of Future Research Requirements," Proceedings of the Twelfth International Technical Conference on Experimental Safety Vehicles Vol. 1, P.233-241, Sweden, 1989.

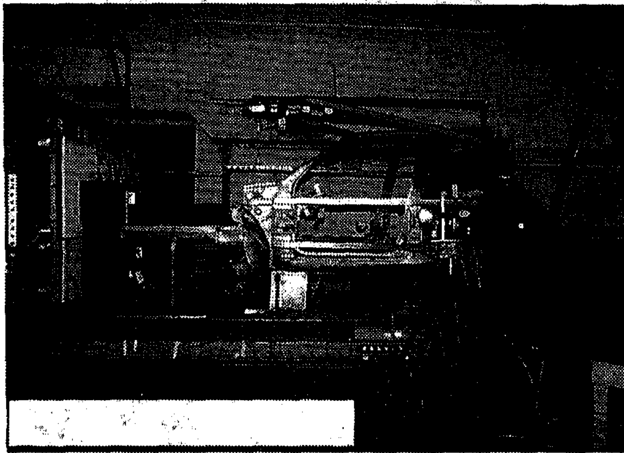
[3] Hensen, S.E., Dueweke, J.J., and Huang, M., "Computer Modeling of Intrusion effects on Occupant Dynamics in Very Severe Frontal Crashes," SAE Technical Paper No. 830613, International Congress & Exposition, Detroit, Michigan, February 28- March 4, 1983.

[4] Matsumoto, H., Sakakid M., and Kurimoto, K., "A Parametric Evaluation of Vehicle Crash Performance," SAE Technical Paper No. 900465, Publication No. SP-807 on the vehicle Crashworthiness and Occupant Protection in Frontal Collision, Society of Automotive Engineers, Inc. Warrendale, PA. 150966, Feb. 1990.

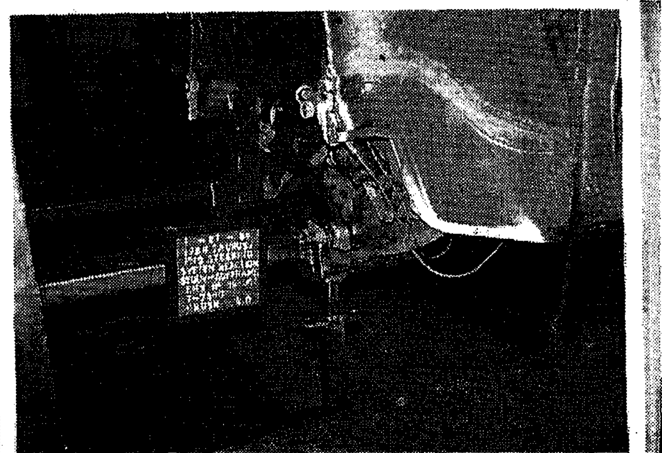
[5] Commission of European Communities, 1995, "Amended Proposal for a European Parliament and Council Directive Relating to the Front-Impact Resistance of Motor Vehicle Amending Directive 70/156/EEC," COM(95) final, Brussels.

[6] R.R. Clements, "A description of Ford Energy Management System," Safety Laboratory Department Internal Report, June, 1976.

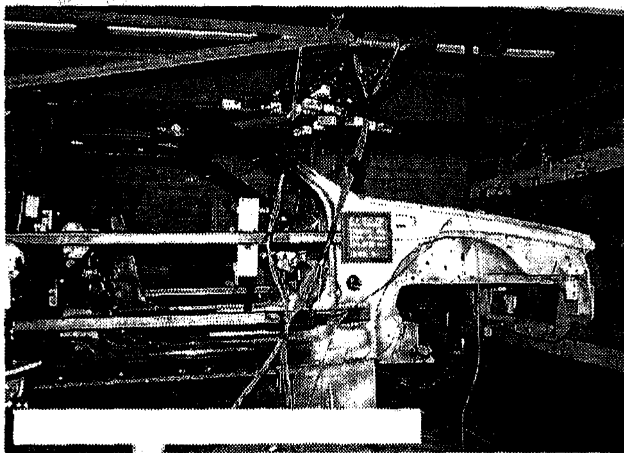
APPENDIX A - SET UP AND RESULTS OF SOME INTRUSION TESTS



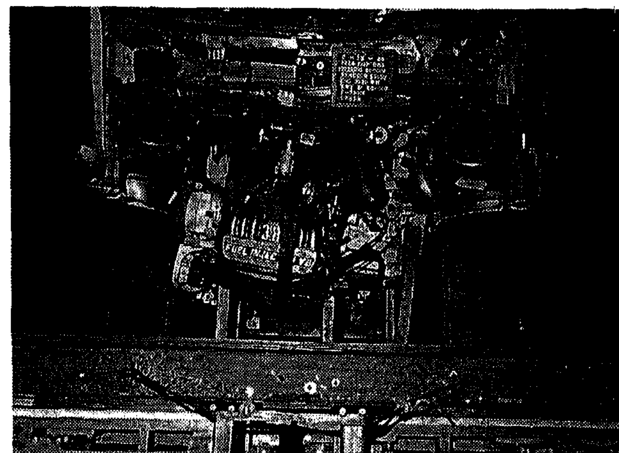
a. Aft-rail into dashpanel



b. Subframe into toe-board

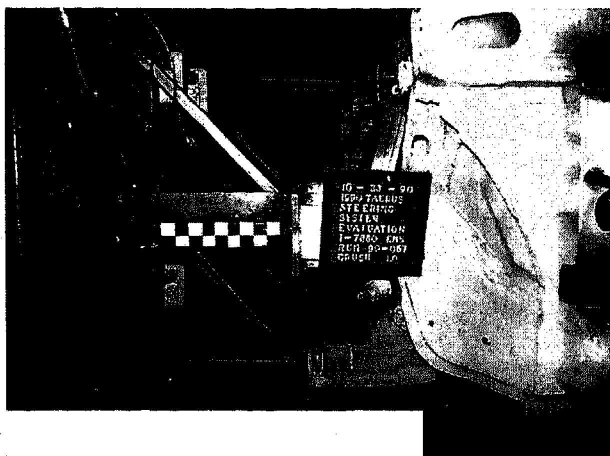


c. Shotgun/shocktower into cowl

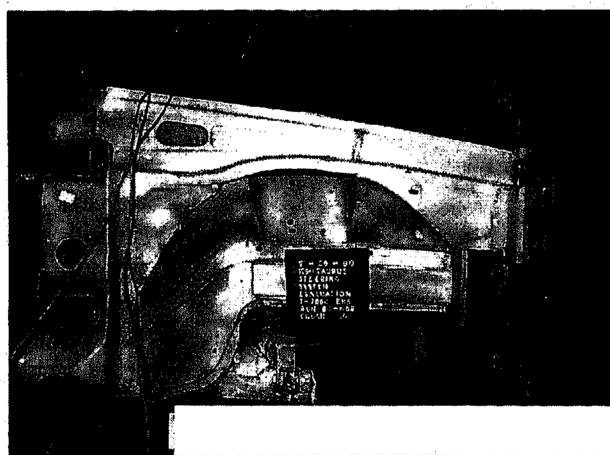


d. Powertrain into dashpanel

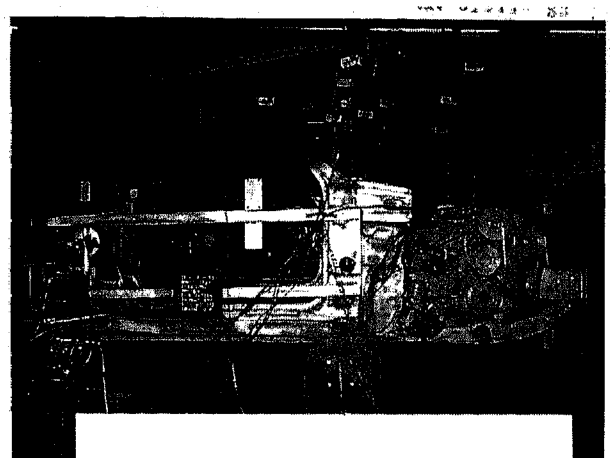
Figure A1: Set up of Primary (single) factors' intrusion tests



a. Aft-rail and subframe



b. aft-rail and shotgun/shocktower

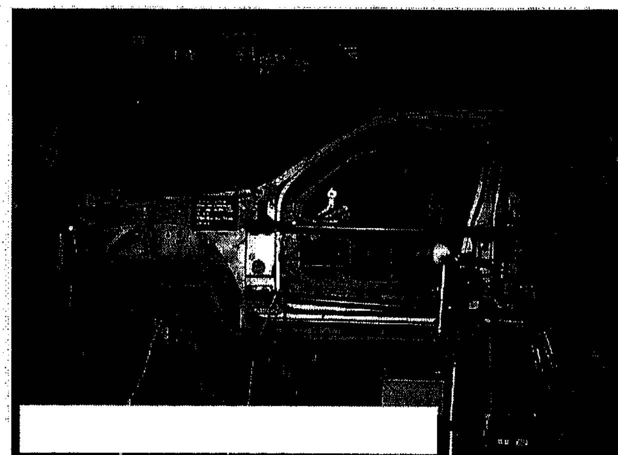


c. Powertrain on subframe

Figure A2: Set up of combined two factors' intrusion tests

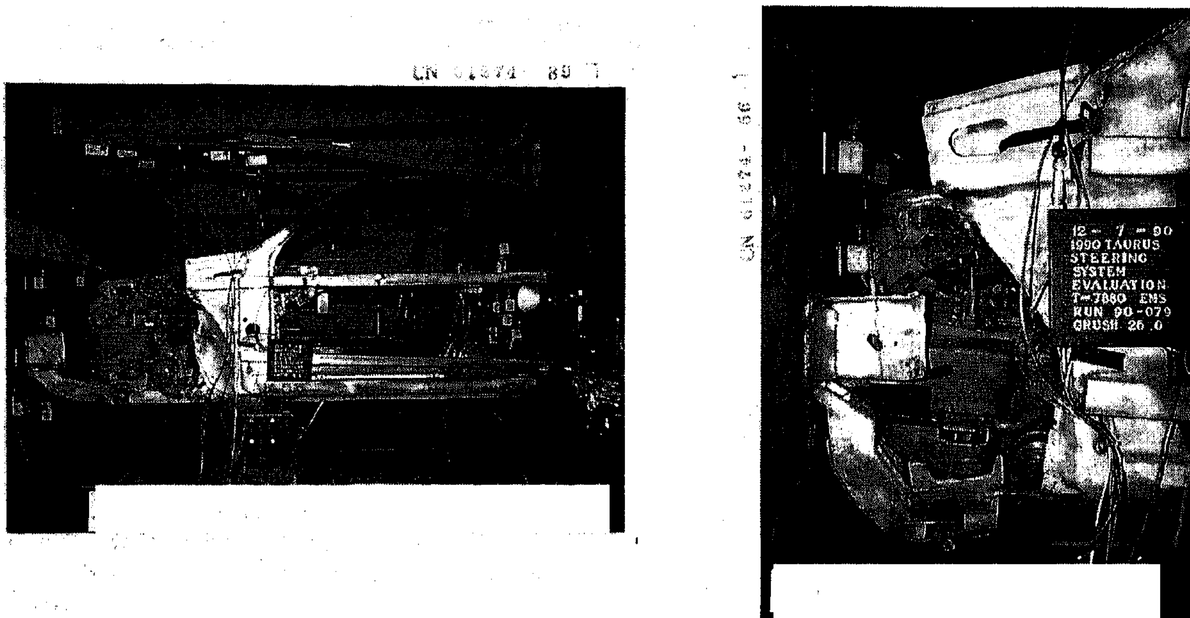


Top view

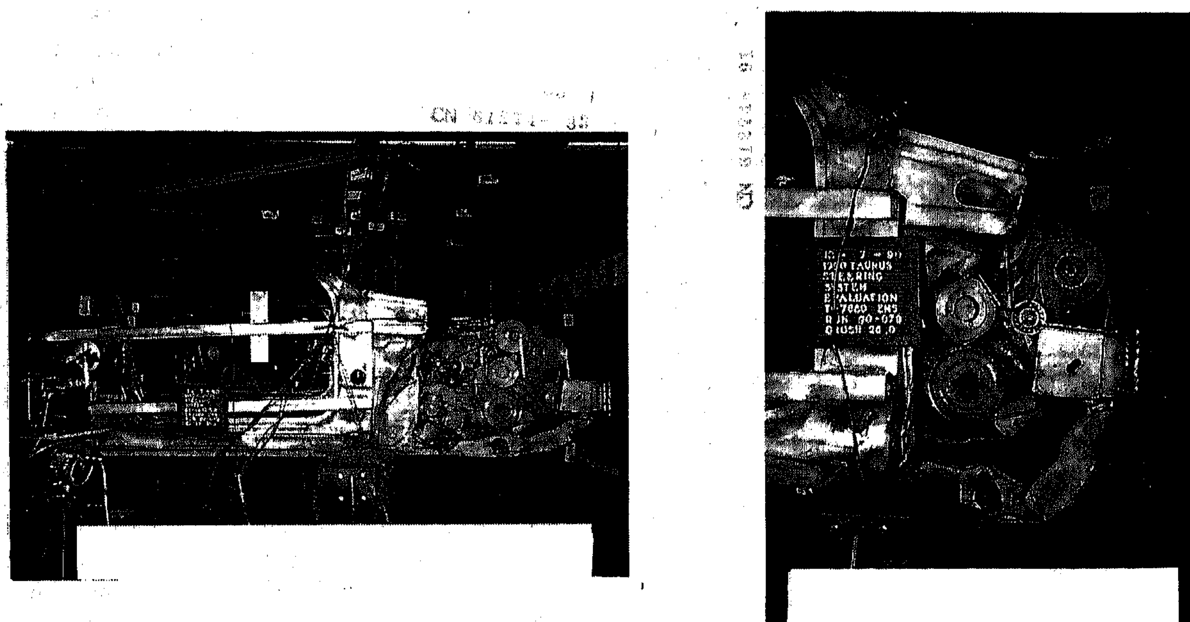


Side view

Figure A3: Set up of combined three factors' intrusion tests



Transmission side view



Engine side view

Figure A4: Powertrain before and after crush

970574

Experimental Devices to Simulate Toepan and Floorpan Intrusion

Cameron R. Bass
University of Virginia

Ehud Dekel
Rafael, MOD.

Jeff R. Crandall, Mike Lange, and Walter D. Pilkey
University of Virginia

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Two sled systems capable of producing structural intrusion in the footwell region of an automobile have been developed. The first, System A, provides translational toepan intrusion using actuator pistons to drive the footwell structure of the test buck. These actuator pistons are coupled to the hydraulic decelerator of the test sled and are powered by hydraulic energy from the impact event. Resulting footwell intrusion is characterized using a toepan pulse analogous to the acceleration pulse used to characterize sled and vehicle decelerations.

Sled tests with System A indicate that it is capable of accurately and repeatably simulating toepan/floorpan intrusion into the occupant footwell. Test results, including a comparison of lower extremity response between intrusion sled tests and no intrusion sled tests, indicate that this system is capable of repeatable, controlled structural intrusion during a sled test impact. Test results also suggest that lower extremity forcing during footwell intrusion is significantly larger than such forcing with no intrusion and that response and risk of injury for occupants in frontal or frontal-offset crashes are more severe. To expand the available range of simulated footwell intrusion, the design of a second intrusion test system, System B, is presented. This system includes toepan translation, rotation and combined translation/rotation motions as well as an expanded range of toepan accelerations. System B also allows the independent motion of the left and right toepan thus enabling the simulation of yawing motions in the occupant footwell.

INTRODUCTION

Structural deformation of a vehicle into the occupant compartment, termed intrusion, occurs in very severe frontal crashes for which all of the energy absorbing capability of the front structure is exhausted. Such intrusion, however, may be especially significant in frontal-offset crashes in which less than the full frontal area of the vehicle is engaged in energy absorption. Reviews of accident database cases (Morgan *et. al.*, 1991) and computer simulations of crashes (Henson *et. al.*, 1983, *Pilkey et. al.*, 1994) have shown that this structural deformation greatly increases the severity of the impact response and the risk of injury for the occupant. In addition, case histories of lower extremity trauma in automobile collisions indicate frequent costly rehabilitation and significant rates of lifetime impairment (Dishinger *et. al.*, 1994). Although current vehicle safety standards in the United States require only testing of full-frontal vehicle collisions against rigid barriers, offset frontal testing is currently being evaluated by the U.S. National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation (DOT), automobile manufacturers and foreign vehicle testing organizations (Grosch *et. al.*, 1989, Panath-Skogsmo and Nilsson, 1994).

Full-scale vehicle tests are expensive and time-consuming methods by which to conduct parametric studies of impact conditions, to estimate the likelihood of injury using anthropometric test dummies and to evaluate design changes intended to minimize the risk of injury resulting from intrusion. Ideally, a test procedure capable of reproducing the frontal-offset impact environment of the vehicle could be simulated in a laboratory setting where tighter control of test conditions and configurations is possible. At the University of Virginia, two systems for duplicating the structural intrusion of a vehicle have been developed using a modified sled system. The goal of this effort is to characterize the effects of structural intrusion on the occupant in a simplified and repeatable manner. This paper begins with the basic theory necessary to simulate the essential characteristics of vehicular intrusion in the footwell region and follows with the physical and procedural design of a first-generation translational intrusion simulation system, System A. The succeeding section presents intrusion experiments performed using System A; included is a description of structural and lower extremity response of the translational intrusion system at intrusion levels of 0 cm, 7 cm and 22 cm. The design of a second-generation intrusion system, System B, follows with a discussion of design criteria and the development of a system capable of toepan rotation, translation, and combined translation/rotation motions. We

finish with conclusions regarding System A test results and the System B design.

CHARACTERIZATION OF FOOTWELL INTRUSION

Few published studies exist that describe the displacement, collapse and buckling of a vehicle's thin metal structures during a crash. This lack of information is a result of both the complexity of the intrusion event and the mechanical response limitations of the instrumentation used to record the event. For frontal-offset vehicle tests, intrusion data of the footwell region has historically been obtained from a single uniaxial accelerometer mounted to the toepan along the longitudinal axis of the automobile. As the crash progresses, however, the toepan locally deforms in translation and rotation so that the actual orientation of the accelerometer axis becomes poorly defined (c.f. Prasad and Smorgonsky, 1995). To obtain more accurate intrusion data, many alternative methods of measurement have been attempted, but they also fail to fully characterize the toepan and floorboard deformation. Movement of the engine or structural members (e.g., the rails or subframe) relative to the occupant compartment has been recorded by both accelerometers and film analysis. Whereas this motion can provide estimates of the initiation and completion of the intrusion event, it does not provide a three-dimensional characterization of the collapse and buckling of the thin metal structures. Direct measurements of the displacement of the toepan have been attempted by affixing string potentiometers to the toepan. This instrumentation package has met with little success owing to the limited response characteristics of the spring-loaded potentiometers. Analysis of the structural intrusion during the crash event using high speed films often provides incomplete information owing to partial occlusion of views and to difficulty in obtaining orthogonal frames of reference. Post-crash measurements of the vehicle's occupant compartment provide cross-sections of the footwell region at only a single instant and cannot provide complete time histories or indicate the sequence of translations and rotations. In addition, no elastic effects occurring after the crash in the floorpan and toepan structure can be evaluated. Although the ability to measure the intrusion event in the footwell is currently limited, increased frontal-offset testing and advances in finite element models should provide more data and detailed descriptions of the structural deformation. Meanwhile, approximations based on available data must be made to characterize intrusion for reproduction in the laboratory sled test environment (Crandall, *et. al.*, 1995).

Sled systems have demonstrated their usefulness in reproducing a given vehicle deceleration pulse in a controlled, reproducible manner. In a similar way, a toepan or floorpan pulse can be used to define the intrusion event. Analogous to the crash pulse used to characterize vehicle and sled tests, Kuppa and Morgan (1993) have suggested the use of a toepan and floorpan acceleration time history to characterize the structural deformation of the footwell region. Kuppa and Morgan addressed the problem of limited experimental data by making several simplifying assumptions based on a 58 km/h (36 mph) frontal-offset test performed with a 1991 Ford Taurus 4-door sedan and a 1991 Honda Accord DX 4-door sedan. Toepan accelerations during the initial phases of vehicle deceleration were matched with those of the vehicle using the assumption that little rotation had taken place to that point in the event. They also used static post-test profiles to evaluate the overall level of intrusion and matched the second integral of the acceleration curve with the post-test measurements (c.f. post crash-test footwell profiles from Prasad and Smorgonsky). Using these approximations, Kuppa and Morgan produced an equivalent toepan pulse and brake pedal pulse in terms of the longitudinal and vertical motion. These pulses maintained the characteristics of the original data and were readily reproducible in the laboratory.

While these approximations are an excellent start at defining the structural intrusion issue, several issues still need to be addressed to fully characterize the complex motion of the toepan and floorpan intrusion. Examination of the static post-crash profiles for several automobile toepans indicates that both translational and rotational motion about the intersection of the floorpan and the toepan can occur during the crash events. Depending on the vehicle, these intrusion profiles may be described as predominantly translational, predominately rotational or a combination of translation and rotation. Thus, characterization of the toepan and floorpan intrusion as a sequence of finite longitudinal and vertical displacements is a rough approximation of the actual footwell deformation. As more data from computer models and vehicle crash tests becomes available, general motions about three axes will be able to be incorporated into the characteristic pulse description. Additionally, since the intrusion event is deformation of a structure, rigid body translations and rotations about the heel location may not sufficiently describe the intrusion event. Indeed, motion at the midpoint of the toepan may have little similarity with the intrusion at other regions of the footwell owing to the compliance and complexity of the footwell structure (e.g., wheel well asymmetry). In addition to improved instrumentation measuring more degrees of freedom, sensors must be mounted at multiple locations on the toepan and floorpan structures.

For the purposes of designing an intrusion system using a laboratory sled, however, the intrusion simulator sled system begins with a simple, easily characterized intrusion profile. This profile is translational intrusion of the toepan in the occupant footwell parallel to the ground. In the next section, the System A intrusion simulator is described that incorporates translational motions of 0 cm, 7 cm, or 22 cm total toepan displacement into the occupant compartment.

INTRUSION SYSTEM A - DESIGN CONSIDERATIONS

The System A intrusion simulator is the result of the application of several design criteria to enable accurate and reproducible investigation of occupant response, particularly lower extremity response, during a simulated crash event. These criteria are:

- **The system is designed for translational intrusion.** The intrusion simulator will generally be capable of rigid body translations into the footwell region.
- **The toepan produces approximately 20,000 N maximum force applied to both legs.** As the average tibial breaking strength in axial compression is approximately 10,000 N (Yamada and Evans, 1970), the maximum forcing of the intrusion simulator should be able to simulate conditions including the fracture of both tibias under axial load. This extreme condition should enable the simulation of other lower extremity injuries (for example, calcaneous or tibial pylon fractures) under lower loads.
- **The toepan is capable of approximately 25 cm maximum toepan displacement during a simulated crash event.** A recent study (Crandall *et. al.*, 1995) found that 85.6% of below-knee injuries sustained by the driver occur with less than 14 cm of intrusion. In addition, typical intrusions of ~20 cm are seen in full-frontal automobile crash testing.
- **Toepan translation is relatively independent of dynamic lower extremity load.** As the gross structural intrusion in an automobile is likely insensitive to the loading provided by occupant lower extremities, toepan motion should occur independently of this applied loading. Also, it is desirable to repeat intrusion conditions for occupants of various masses and lower extremity geometries.
- **The intrusion simulator is capable of toepan pulse tailoring and stroke selection.** To investigate various intrusion conditions and responses, flexibility is desired in the programming of the intrusion simulator.
- **The toepan mechanism is driven by the sled decelerator.** A substantial amount of fluid power is dissipated in the sled decelerator during an impact event. This energy is used to drive the intrusion simulator as discussed below.

Investigating crash-induced structural deformation of the vehicle footwell with a laboratory sled requires the capability to move the toepan and floorpan components into the footwell space in a programmed displacement versus time profile. One way to fulfill this requirement is to integrate the actuation of the floorboard components with the function of the hydraulic decelerator of the sled. The University of Virginia test sled (VIA Systems 713) functions by initially accelerating the test buck, carriage and payload to a predetermined velocity and then decelerating the payload along a programmed crash pulse. The deceleration profile is produced by a hydraulic decelerator (i.e., water brake) mounted on the sled carriage. The hydraulic decelerator is configured as a cylinder with an array of discrete orifices over its length. During the deceleration pulse, fluid flows through this array, dissipating energy from the test sled in a controlled manner. The individual orifices are sized to vary the amount and velocity of the fluid being forced through them and to regulate the energy absorption rate of the device. Since the decelerator is coupled to the sled carriage, it is possible to utilize the flows and pressures generated through the orifices to actuate the floor board components of the vehicle compartment mounted on the sled carriage.

Figure 1 shows a projected (flat pattern) view of the decelerator cylinder. The standard orifice rows for deceleration of the sled, referred to as program rows, have been supplemented by an output row used to provide hydraulic power to the intrusion simulator. The output ports are currently spaced at 7.6 cm intervals, whereas the two staggered rows of programming orifices provided for pulse tailoring are 1.9 cm apart. For finer control of the output pulse it is possible to add additional output ports to the decelerator between the current ports. When the decelerator cylinder is installed in the containment vessel, each port of the output row is adjacent to an opening provided in the containment vessel. This allows output pipes to connect directly to the ports and to be routed outside the containment vessel. Outside the containment vessel, flexible tubing connects the decelerator with pistons mounted on the buck. These actuator pistons, also referred to as slave cylinders, are used to displace the toepan and floorpan structure. During deceleration of the sled, the ports in the output row are used to transfer fluid from the hydraulic decelerator to the slave cylinders. Varying the size and location of the orifices in the output row permits detailed programming of the toepan/floorpan displacement time histories. **Figure 2** shows a basic circuit schematic investigated for the System A integrated hydraulic decelerator/floor board actuator system.

System A is comprised of a set of drive side hydraulic lines, a control side hydraulic line, an accumulator and a double-acting hydraulic slave cylinder with equal piston areas on both sides. The slave cylinder is driven from the output ports of the hydraulic decelerator discussed above. The system is initially fully filled with hydraulic fluid. Pressure develops in the sled hydraulic decelerator which results in pressure in the drive lines and control line and fluid flow into the hydraulic accumulator to be stored for the toepan actuation. As the impact event begins, the pressure in the hydraulic decelerator is transmitted to both sides of the slave cylinder. So, the toepan remains stationary with respect to the test buck.

At a certain point, the sled decelerator piston passes the control side output port on the sled decelerator. As this occurs, the control side pressure drops and the toepan begins moving under the force of the drive lines remaining in front of the sled piston and the pressure in the accumulator. The maximum attainable acceleration pulse of the actuator is a function of the instantaneous pressures and flows in the hydraulic decelerator and the hydraulic accumulator. Check valves prevent backflow down the drive lines when drive ports are passed by the sled decelerator piston. A constant force internal stop inside the slave cylinder at maximum mechanical stroke is used to terminate toepan travel at a location determined by the position of the hydraulic slave cylinder. The intrusion pulse is controlled by varying the orifice sizes on the drive or control side of the of the circuit. Further tailoring of the toepan force and displacement values is available using one or more 'tuning' circuits with control

orifices and check valves. Since the toepan actuator piston area is substantially smaller than the hydraulic decelerator piston area (typical values are 5 cm^2 for the former versus 20 cm^2 for the latter) minimal fluid drive volume is required from the decelerator.

The detailed physical arrangement of the System A intrusion simulator at the University of Virginia is shown in side view in **Figure 3** and in oblique view in **Figure 4.** The simulator consists of a toepan carriage with separate footrests for the left and right feet. This carriage travels in translation with respect to the ground into the occupant compartment on two rods mounted to the test buck. The limit of travel of the toepan carriage is 24.9 cm from the reference footwell location. The toepan carriage is driven by a 6.4 cm diameter hydraulic slave cylinder connected to the hydraulic decelerator on the test sled. The cylinder is a double rod/double acting hydraulic cylinder with equal piston areas on either side of the cylinder. The rear mounting position of the hydraulic cylinder may be adjusted to limit total toepan travel in increments of 2.5 cm.

The toepan carriage is instrumented with footplates mounted on the toepan carriage at 55 degrees from the horizontal, consistent with a typical midsize vehicle driver-side footrest as shown in **Figure 4.** The footplates are 30 cm in height by 17 cm in width and are instrumented with load cells at each corner oriented normal to the local direction of the footplate. In addition, accelerometers are mounted on the rear center of the plate to assist in mass compensation of the measured forces.

SYSTEM A - INTRUSION PULSE SELECTION AND PULSE PROGRAMMING

The pulses in both the hydraulic decelerator and the toepan actuator circuit are selected using hydraulic assumptions based on Bernoulli hydrodynamics. The volumetric flow rate, Q, from any orifice in the system is assumed to obey the relation:

$$Q = kA\sqrt{\frac{2}{\rho}(P_1 - P_2)}\,.$$

The quantity ρ is the density of the fluid, A is the orifice area, and P_1 and P_2 are pressures on the upstream and downstream sides of the orifice. Semi-empirical flow coefficients k are determined from experiments and from standard values of orifice coefficients (ASME, 1981).

To select the sled acceleration profile and the toepan acceleration profile simultaneously, the orifice array in the decelerator and the orifices leading from the decelerator to the toepan actuator must either be coordinated or the effect of the toepan actuation on the decelerator pulse must be negligible. In the first instance, toepan actuator flow and pressure requirements may contribute to the effect of the decelerator orifice array on the sled acceleration pulse. Thus, at any given instant, the volume of fluid for the operation of the actuator has to be a part of the volume of fluid being forced through the orifices of the decelerator to govern the test pulse. However, since considerable power is dissipated by the decelerator under a wide range of experimental conditions, the effect of toepan actuation on the deceleration pulse generated by the hydraulic decelerator may be negligible. For example, a 55 km/h, 1000 kg sled run dissipates over 1000 kW of power for a pulse length of 80 milliseconds. In contrast, the power required to move a floor/toeboard complex 15 cm at an average force of 10 kN over a 40 millisecond interval is approximately 37 kW, giving a power ratio from decelerator to toeboard actuator of over 25:1. However, for more extensive use of decelerator hydraulic power, the toepan intrusion actuation must be coordinated with the sled deceleration.

To model structural intrusion events, the toepan pulse should have certain acceleration, velocity and displacement characteristics. For instance, the toepan should experience a displacement relative to the vehicle center of mass in a direction opposite to the direction of sled travel. Ideally, this displacement is caused by a toepan actuator acceleration pulse. After some time, the toepan remains fixed in space with zero velocity relative to the ground while the sled continues to decelerate and move with respect to the toepan. The selected maximum toepan intrusion/deformation is the total distance of sled travel after the toepan velocity becomes zero relative to the ground. So, if a given toepan displacement, D, is selected, the toepan actuation should begin within a length of D from the end of the total sled travel.

Input requirements for programming this process are the desired acceleration on the sled, the force on the toepan and the degree of toepan intrusion. Each of these parameters is a function of the local pressure at various locations in the decelerator/toepan actuator hydraulic system. These parameters are adjusted specifying input/output orifice diameters and the port locations. An essential design characteristic for the intrusion system is the ability to adjust the dynamic intrusion parameters over a large range for a given sled pulse. In actual crashes, the level, duration and nature of the intrusion is highly dependent on vehicle characteristics (e.g., size, weight, engine mounting orientation and drive-train configuration) and crash configuration (e.g., vehicle offset distance and velocity). Thus, the degree of toepan intrusion for vehicles involved in a 55 km/h crash under various conditions may range from 0 cm to over 20 cm.

For System A, several toepan intrusion parameters can be controlled in the intrusion event:

- **Intrusion initiation.** Location of a control line in the hydraulic decelerator can be used to define intrusion initiation. The intrusion initiation of this intrusion drive circuit is accurate owing to the ability to control precisely a delayed initiation of the toepan actuator using the physical position of decelerator output ports.
- **Toepan stroke.** Stroke can be controlled by the cylinder mounting location and cylinder stop.
- **Toepan acceleration profile.** System A includes a combination of passive orifices in the drive lines and a

hydraulic accumulator so that the dynamic effects of the sled pulse can be used to tailor the toepan pulse. The simulator is capable of toepan forcing late in the sled deceleration stroke.

Spreadsheet-based computer software has been developed to identify the orifice array for a given pulse with the hydraulic decelerator while incorporating the essential pressure and flow information necessary to select the hardware configurations and operational conditions for the toepan actuator. The procedure requires the user to specify the toepan acceleration, velocity and position time histories. Conservation of energy and momentum are used to derive the flow rates and pressures required from the decelerator for the actuator pistons within limits. As the flow in the intrusion hydraulics has a significant effect on the sled decelerator, an iterative solution for the sled deceleration profile must be sought, constrained by the desired intrusion. An algorithm for performing this iteration is presented in **Figure 5**. For a given circuit configuration, the program outputs the location and size of orifices for the output rows (i.e., toepan actuator piston) and the program rows (i.e., sled pulse) of the hydraulic decelerator.

SYSTEM A - TEST RESULTS

To demonstrate the use of the intrusion simulator, sled impact tests were performed with zero intrusion, 7 cm intrusion and 22 cm intrusion using dummy and cadaveric subjects restrained by a knee bolster and a three-point seat belt. Sled tests were performed using the University of Virginia Automobile Safety Laboratory acceleration sled test system. The test buck was configured as a 1994 Ford Taurus vehicle and the deceleration pulse was selected to reproduce the crash pulse of a mid-size automobile measured using damped accelerometers. As discussed above, hydraulic power from the sled decelerator drives the intruding toepan. Instrumentation on the toepan includes force-sensing footplates, a toepan carriage accelerometer and a spring-loaded string potentiometer to measure toepan displacement relative to the sled.

Sled velocity at the initiation of the crash pulse is selected to be 58.2 km/h. The intrusion stroke is selected to be 7 cm or 22 cm relative to the sled. This intrusion is selected to coincide with the last 7 cm or 22 cm of sled travel respectively, late in the deceleration stroke. Intrusion acceleration is selected as approximately sinusoidal in the sled frame.

The dummy tests utilized the Advanced Lower Extremities (ALEX) developed by the NHTSA. The goals in developing the ALEX legs include increased biofidelity through such means as improved inertial distribution of the lower extremities and reduced stiffness using advanced composite materials. The ALEX lower extremities also offer an increased range of motion for ankles compared with the standard Hybrid III lower extremities. This is significant in intrusion tests where the standard Hybrid III lower extremities likely reach hard joint stops, disrupting testing. In addition, the ALEX legs incorporate improved instrumentation compared with the standard Hybrid III lower extremities. For the intrusion runs, a Hybrid III dummy with ALEX lower extremities and cadaveric subjects were used.

In **Figure 6**, a representative deceleration sled pulse and toepan intrusion pulse are plotted for a representative intrusion sled test of 7 cm intrusion. The toepan intrusion deceleration relative to the test sled is approximately sinusoidal with an approximately 80 g peak in the intrusion pulse (100 g relative to the ground). The oscillation at the end of the sled deceleration results from a single control orifice remaining in the hydraulic decelerator at that sled travel position. To eliminate this oscillation would require active control of decelerator orifices or an addition of passive control orifices in the decelerator. **Figure 6** illustrates that the toepan acceleration follows the sled acceleration until the actuation of toepan intrusion. The toepan then quickly stops relative to the ground while the sled continues to decelerate (i.e., the hydraulic decelerator continues to stroke). The net effect of the sled and toepan decelerations is movement of the toepan into the occupant compartment of the buck. As the initial peak and impulse most relevant for lower extremity injury, this toepan intrusion provides a potentially realistic translation intrusion pulse.

In **Figure 7**, the sled displacement relative to the ground and the toepan displacement relative to the sled are compared. The sled travel is derived as the double integration of the unfiltered sled deceleration signal in **Figure 6**. The toepan intrusion relative to the test buck, measured using a string pototentiometer attached to the toepan, is selected to approximate sled travel in the laboratory reference frame for the last 7 cm. This process imitates a crushing process wherein a rigid body engine stops in the laboratory reference frame and intrudes through the occupant footwell at the very end of the vehicle crash pulse. For exact replication of this process, the toepan displacement relative to the sled and the sled displacement relative to the ground would be the same. From the graph in **Figure 7**, it is seen that the variation between toepan displacement and sled displacement is less than 5% of full scale during the intrusion event.

Repeatability of the toepan acceleration relative to the test sled is shown in **Figure 8**. ATD-35 is a dummy run with ALEX lower extremities. Runs CAD-35-CAD-36 are tests with cadaveric subjects. The inertial loading varies widely between these runs with occupants ranging from 50 kg to 106 kg. However, the magnitude of the initial peaks and troughs varies less than 2% and the timings of the peaks are closer than 1% as a percentage of the 110 ms deceleration event. It is important to note that the initial sled velocity of the three events varied only from 58.2 km/h to 58.4 km/h.

Typical lower extremity response of a 7 cm translation intrusion sled test using System A is presented in **Figure 9**. Initially the lower extremities experience an inertial slap in tibia axial load, identified in high speed films taken of the sled test. This slap results from heel contact with the toepan following initiation of the deceleration crash pulse. It occurs in both intrusion and non-intrusion tests with peak axial force values comparable when intrusion initiation is late enough.

Computational simulations of this slap suggest that the magnitude of the initial force is relatively insensitive to the size of the gap but very sensitive to the existence of a gap. After this inertial loading, translational intrusion onset is approximately 65 ms. The effect of the translational intrusion on the tibia axial load is somewhat obscured by the seat belt/knee bolster interaction which occurs at the same time. However, the peak value of axial load in the tibia is nearly 70% larger than the peak axial load in sled tests without intrusion. For the 7 cm intrusion sled test, the tibia A/P moment clearly shows the impact of knees into the knee bolster at approximately 51 ms. During the intrusion event after 65 ms, the maximum bending moment is more than twice the value of a similar test with no intrusion.

A 22 cm intrusion test is shown in **Figure 10.** Again, the initial slap in axial force is seen in tibia axial load and A/P moment. As the intrusion occurs after this period, the peak load values and slap duration is comparable to that seen in the 7 cm intrusion sled test. In contrast with the 7 cm intrusion test, a clear intrusion peak is seen in the tibia axial load after intrusion initiation at approximately 47 ms; the intrusion is early enough to produce a force peak distinct from the knee bolster/belt interaction. In addition, the tibia A/P moment peaks at nearly 120 N-m during the intrusion event. Also, we again clearly see knee bolster contact in the femur axial load.

A comparison of the lower extremity response using a Hybrid III dummy between a 22 cm intrusion sled test using System A and a NHTSA full frontal crash test of a 1992 Ford Taurus is plotted in **Figure 11.** Qualitatively, we see the characteristic triple humped features in both outboard distal tibia axial forces. We identify the first peak as inertial slap, the second as onset of intrusion (can also see this in as an acceleration peak in Taurus vehicle acceleration) and the third peak as the knee bolster/ belt interaction. The magnitudes of these peaks are quite comparable though the Taurus crash test deceleration pulse is nearly 20 ms shorter than that of the sled test.

The behavior of the maximum forward location of the foot center of pressure, for tests with both dummy and cadaveric subjects, is shown in **Table 1** for various levels of intrusion. The foot center of pressure has been normalized by foot length and is an average of multiple cadaveric and dummy runs for each condition. The location of the furthest forward center of pressure increases with intrusion level for both cadaver and dummy since increasing dorsiflexion of a cadaver ankle tends to shift the center of pressure towards the toes. The timing of the maximum occurs earlier with increasing levels of intrusion for cadaveric lower extremities since earlier intrusion results in earlier dorsiflexion; the bending moment that can be applied about a cadaver ankle during a crash event increases as the cadaver ankle dorsiflexes. In contrast, Alex dummy ankles see relatively constant and low moment values until the joint stops are reached. The Alex joint stops are not reached in either 0 cm or 7 cm intrusion, only in 22 cm intrusion. So, the maximum center of pressure is obtained very early for 22 cm.

A comparison of dummy center of pressure maxima vs. cadaver maxima again shows the differences between moment/angle properties of dummy ankles vs. cadaveric ankles. Center of pressure forward extent is always greater for cadavers for a given intrusion condition since cadaveric ankles can offer significant moment resistance across the ankle before reaching natural joint stops. This tends to shift the center of pressure forward, towards the toes. Other intrusion subject sled test data from these tests, including cadaveric ankle injuries, are reported elsewhere (Crandall *et. al.*, 1996).

Level of Intrusion	Subject	Center of Pressure Maximum	@Time (ms)
0 cm	Dummy	45.5%	58.8
	Cadaver	52.4%	81.4
7 cm	Dummy	49.9%	73.1
	Cadaver	61.1%	66.6
22 cm	Dummy	55.5%	46.3
	Cadaver	66.6%	62.8

Table 1: Longitudinal Center of Pressure Maxima for Cadaveric and Dummy Left Feet vs. Level of Intrusion, Normalized by Foot Length, Multiple Run Average.

INTRUSION SYSTEM B

Intrusion System B was developed to address limitations of System A. These include a limitation on the maximum acceleration developed in the toepan and certain 'quantization' limits on travel. Assuming reasonable intrusion behavior, toepan travel was limited in System A to discrete 7.6 cm segments owing to decelerator output port locations. In addition, since sled hydraulic cylinder orifice spacing is constant, the toepan and the sled deceleration pulses were sensitive to the last 7.6 cm of sled travel. Late in the sled deceleration event, the sled hydraulic decelerator piston travels relatively slowly over the last few orifices at the end of sled travel. The lack of control orifices significantly reduces control of toepan and sled pulses at end of travel. Also, a finite element analysis of System A footplates/toepan indicated substantial dynamic deflections during the impact event. The existence of these deformations was confirmed by observing the permanent deformations of several structural components over the series of tests using System A.

A second-generation design, System B, was developed incorporating the experience gained operating System A. An augmented set of design criteria successfully addresses significant limitations of System A. These modified design criteria are:

- **The system is designed for translational, rotational and combined translational/rotational intrusion.** The intrusion simulator will generally describe rigid body motions into the footwell region. Each footplate is

capable of independent motions and is thus capable of 'yawing' motions between the left and right feet.

- **The toepan produces approximately 20,000 N maximum force applied to both legs.**
- **The toepan acceleration may be prescribed in excess of 150 g's.** Maximum available acceleration with System A using a typical Ford Taurus crash pulse at 58 km/hr is approximately 100 g's.
- **The toepan is capable of approximately 25 cm maximum toepan displacement during a simulated crash event and approximately 55^0 of rotation.**
- **Toepan translation is relatively independent of dynamic lower extremity load.**
- **The intrusion simulator is capable of toepan pulse tailoring and stroke selection.**
- **The occupant footplate is more rigid and includes enhanced instrumentation.** The System A toepan instrumentation was capable of measuring forces normal to the plane of the footplate. Additional measurement capability was desired in System B.
- **The toepan mechanism is driven by independent contact with stationary reaction mass.** A disadvantage of System A is the dependence of the sled pulse on hydraulic power derived from the sled decelerator very late in the impact event. With the existing hydraulic cylinder design, this reduces the ability to control either the sled or the intrusion deceleration pulses near the spatial end of the impact event.

System B operation is based on the impact of three separate hydraulic decelerators into a fixed reaction mass: a sled decelerator, a left toepan decelerator and a right toepan decelerator. These decelerators must be coordinated to produce appropriate sled crash and toepan intrusion pulses. For pure rotational motion the base of the carriage is locked and the intrusion decelerator actuates the toepan only. For pure translation, the toepan is locked with respect to the footplate carriage which is allowed to slide along a carriage guide rod. Combined rotational/translational motions can be achieved using the cam follower on the Rotational Guide Bar shown in **Figure 12.** The Rotational Guide Bar is adjustable to allow different combinations of translations and rotations. Forces on each footplate are measured using a six-axis load cell measuring local shears, moments and normal force.

For System B, several toepan intrusion parameters can be controlled in the intrusion event:

- **Intrusion initiation.** Intrusion initiation occurs when the intrusion hydraulic decelerator contacts the intrusion impact cushion. This intrusion impact cushion is generally located in a different spatial location from the sled impact cushion.
- **Toepan stroke.** Stroke can be controlled by the cylinder mounting location and cylinder stop. The left and the right toepans can be controlled independently.
- **Toepan rotation.** May be controlled independently or in a combined motion with translation.
- **Toepan acceleration profile.** Acceleration profile tailoring is available using orifice arrays in the intrusion hydraulic decelerators. Again, can be independently adjusted for each footplate.

System B is diagrammed in **Figure 13.** Individual toepan carriages allow independent toepan motions, and cylindrical hydraulic toepan decelerators provide toepan actuation. There are two attachment points behind each toepan for the toepan hydraulic decelerators. The first is aligned with toepan carriage translational motion to actuate translation and combined translation/rotation. The second is located at the top of the toepan to actuate purely rotational motions.

For intrusion pulse programming, the desired sled crash pulse is selected initially. Then, an intrusion profile is selected. This profile is subject to certain admissibility constraints. For example, no acceleration and displacement may be selected that results in the toepan intrusion cylinders reaching limits of travel. The rest of a typical iterative algorithm is shown in **Figure 14.** One primary difference between the procedure followed using System A and that for System B is in the use of simulation to select a toepan orifice profile. Applied force on the footplates affects the pulse derived from the toepan hydraulic decelerators. So, the test must be simulated prior to the run to allow appropriate selection of orifice parameters for the toepan decelerators. A lumped mass model of the test system will be used to perform these simulations.

An ideal triangular toepan intrusion pulse, comparable to translational pulses produced using System A, is shown for System B in **Figure 15.** This triangular pulse has equal rise and drop times after which it drops rapidly to zero in the laboratory reference frame. This pulse produces zero toepan translation thereafter in the laboratory frame while the sled continues to translate resulting in motion of the footplate into the occupant compartment. The calculation procedure is adaptable to complicated toepan motions.

There are several advantages and disadvantages of System B as compared to System A. The primary advantage of System B is that it can describe more complicated motions including combined translational/rotational motions. In addtion, System B allows increased toepan acceleration as compared with System A. The primary disadvantage of System B is a potentially complicated pulse programming calculation. This may be eased somewhat if the energy associated with the toepan motion is a small percentage of the sled energy over the entire range of motion.

CONCLUSIONS

Structural intrusion into automobiles is a complex process involving deformation, translation and rotation of automobile structural components. However, two test systems, one capable of simulating translational toepan intrusion, the other capable of simulating translational, rotational and a combined translational/rotational motion, have been developed to simulate the structural deformation

seen in car crashes. Although current data on intrusion is limited, System A is capable of reproducing available toepan and floorpan intrusion time histories in translational motion. This intrusion simulator adapts an extensively tested and documented technology developed to tailor deceleration pulses for sleds using hydraulic decelerators. Similar to the sled pulse tailoring, the intrusion profiles can be programmed in a precise manner with relative ease. Parameters affecting the intrusion pulse that can be adjusted include orifice sizes, orifice locations and hydraulic circuit layout. A computer program has been developed to iteratively determine these parameters for both the sled hydraulic decelerator and the intrusion actuator pistons simultaneously.

Using this intrusion simulator, the significant effects of a 7 cm and a 22 cm structural intrusion on dummy and cadaver lower extremity force and moment values have been demonstrated. The displacement of this structural intrusion is selectable to within 5% of the selected displacement condition. The peak forcing of the demonstrated intrusion pulse is repeatable in both timing and magnitude to within 5% under widely differing inertial lower extremity loads. Also, using rigid body toepan and floorpan structures, the system requires minimal energy from the hydraulic decelerator and does not significantly affect the sled pulse.

Building on experience gained operating System A, an enhanced second-generation system, System B, was developed. This intrusion system incorporates pure translation, pure rotation and combined translational/ rotational motion. The system may be actuated independently for each foot. It includes increased translational acceleration limits as compared with System A. With more options for system control, however, comes a more complicated pulse programming procedure. Tests with System B will proceed in the future to explore various aspects of occupant footwell intrusion.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support and guidance of Richard M. Morgan and Rolf H. Eppinger of NHTSA in the conduct of this research. This study was supported in part by DOT NHTSA Grant DTNH22-93Y-07028. All findings and views reported in this manuscript are based on the opinions of the authors and do not necessarily represent the consensus or views of the funding organization.

REFERENCES

Anonymous, The ISO-ASME Orifice Coefficient Equation, *Mechanical Engineering*, July 1981, p. 44-45.

Crandall, J.R., Bass, C.R.,Klopp, G.S. and Pilkey, W.D., *Sled Tests with Toepan Intrusion Using Post-Mortem Human Subjects and the Hybrid III Dummy*, IRCOBI, Dublin, Ireland, October 1996.

Crandall, J.R., Jordan, A., Bass, C.R., Klopp, G.S., Pilkey, W.D. and Sieveka, E.M. Reproducing the Structural Intrusion of Frontal Offset Crashes in the Laboratory Sled Test Environment, SAE Paper 950643, 1995, Society of Automotive Engineers, Warrendale, PA.

Crandall, J.R., Martin, P.G., Sieveka, E.M., Klopp, G.S., Pilkey, W.D., Dischinger, P.C., Burgess, A.R., O'Quinn, T.D. and Schmidhauser, C.B. The influence of footwell intrusion on lower extremity response and injury in frontal crashes. 39th Annual Proceedings of the AAAM, Chicago, October 16-18, 1995.

Dischinger, P., Burgess, A.P., Cushing, B.M., Pilkey, W.D., Crandall, J.R., Sieveka, E.M. and Klopp, G.S. Lower Extremity Trauma in Vehicular Front Seat Occupants. 4th Annual CDC *Injury Prevention through Biomechanics* Symposium, Wayne State University, Detroit, Michigan, May 5-6, 1994, pp. 29-42.

Grosch, L., Baumann, K. H., Holtze, H. and Schwede, W. Safety Performance of Passenger Cars Designed to Accommodate Frontal Impacts with Partial Barrier Overlap, SAE Paper 890748, 1989, Society of Automotive Engineers, Warrendale, PA.

Henson, S.E., Dueweke, J. J. and Huang, M. Computer Modeling of Intrusion Effects on Occupant Dynamics in Very Severe Frontal Crashes, SAE Paper 830613, 1983, Society of Automotive Engineers, Warrendale, PA.

Kuppa, S. M. and Morgan, R. M. Recommendations for a Sled and Intrusion Pulse in the Study of Injuries to the Lower Extremities, Proceedings of 21st Annual Workshop on Human Subjects for Biomechanical Research, San Antonio, TX, 1993, pp. 71-80.

Morgan, R. M., Eppinger, R. H. and Hennessey, B. C. Ankle Joint Injury Mechanism for Adults in Frontal Automotive Impact, Proceedings of the 35th Stapp Car Crash Conference, San Diego, CA, 1991, pp. 189-198.

Panath-Skogsmo, I. and Nilsson, R. Frontal Crash Tests - A Comparison of Methods, SAE Paper 942228, 1994, Society of Automotive Engineers, Warrendale, PA.

Pilkey, W. D., Sieveka, E., Crandall, J. R. and Klopp, G. S. The Influence of Vehicular Intrusion and Rotation on Occupant Protection in Full Frontal and Frontal Offset Crashes, Proceedings of the 14th International Technical Conference on the *Enhanced safety of vehicles*, Munich, Germany, May 1994, pp. 734-741.

Prasad, P. and Smorgonsky, L. Comparative Evaluation of Various Frontal Impact Procedures, SAE Paper 950646, 1995, Society of Automotive Engineers, Warrendale, PA.

Yamada, H. and Evans, F.G. *Strength of Biological Materials*, Williams and Wilkins Company, Baltimore, MD, 1970.

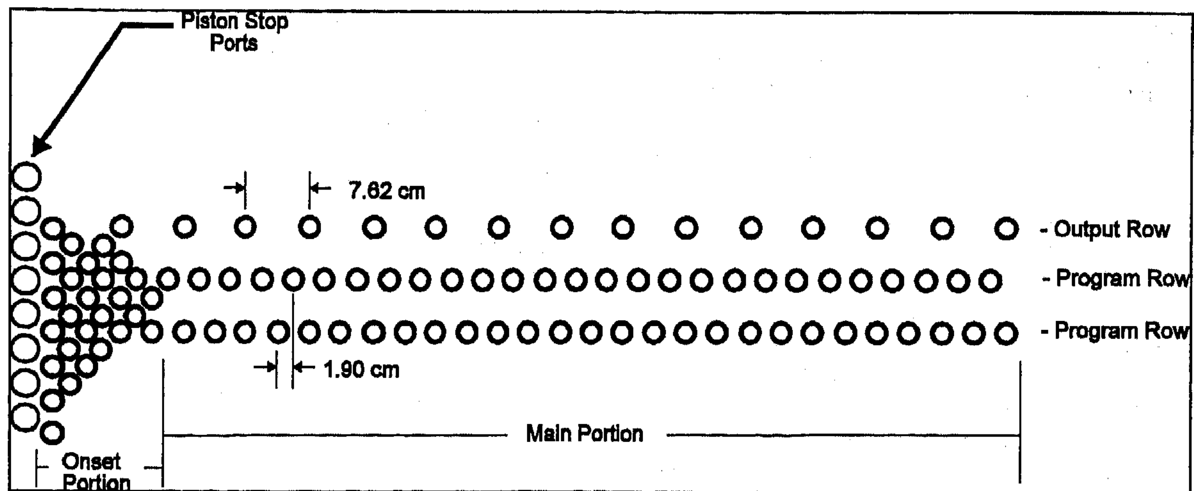


Figure 1: Projected View of Modified Hydraulic Decelerator Cylinder with Program Row

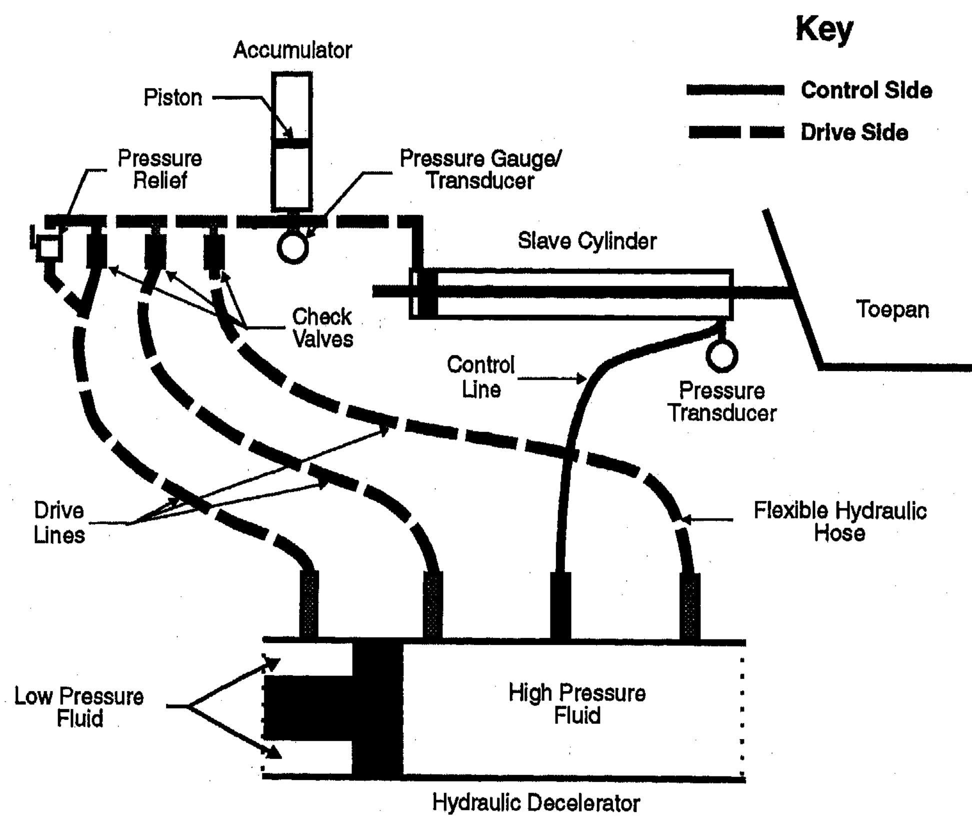


Figure 2: System A - Hydraulic System Schematic

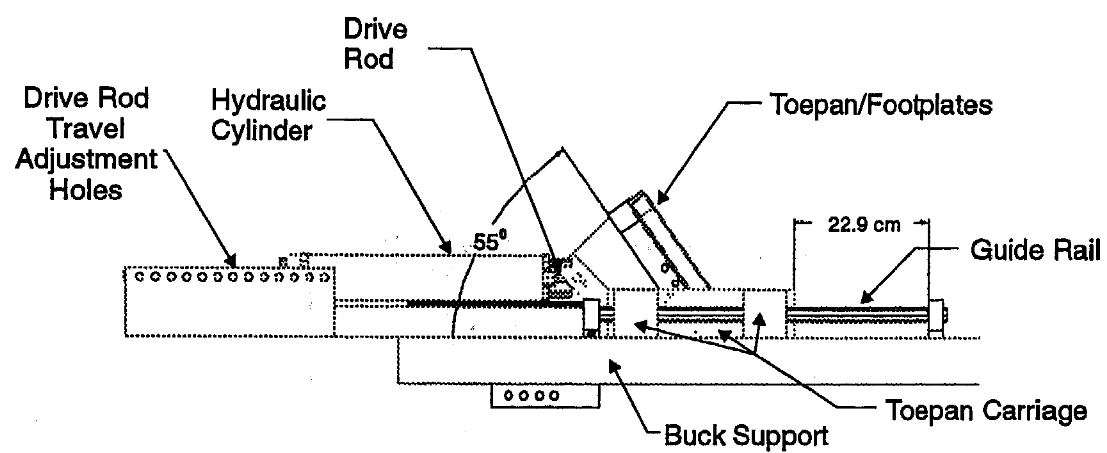


Figure 3: System A - Intrusion Simulation Mechanism - Side View

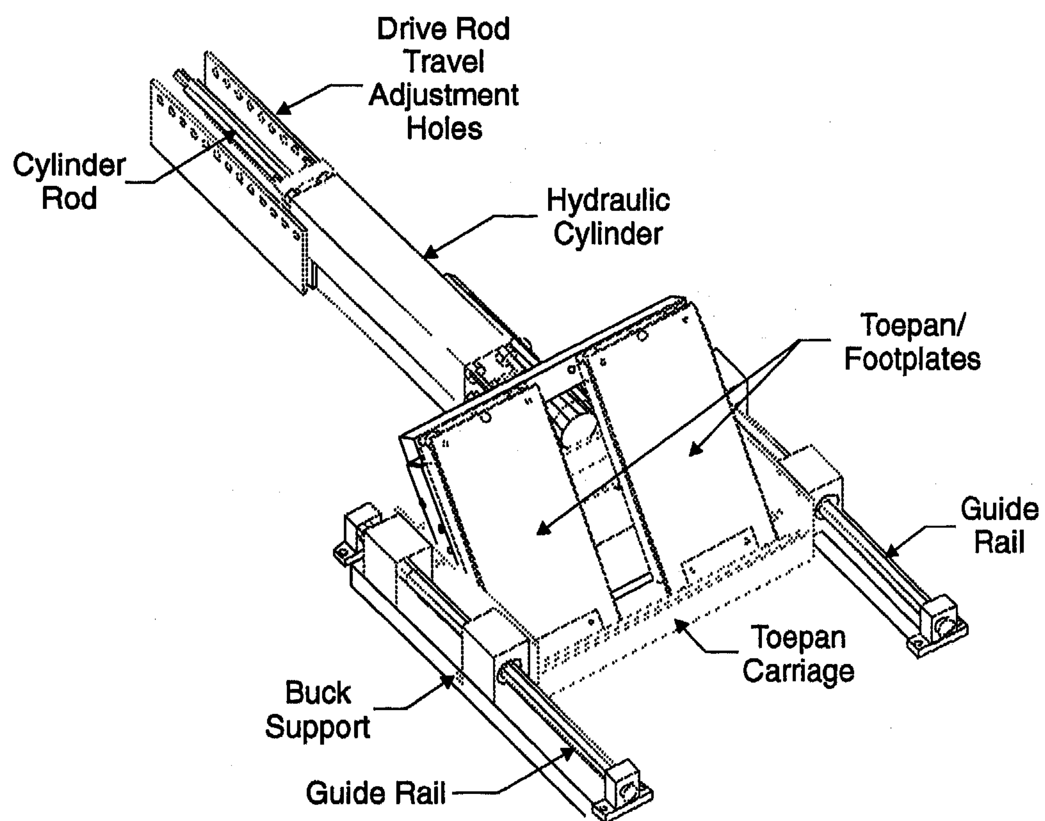


Figure 4: System A - Intrusion Simulation Mechanism - Oblique View

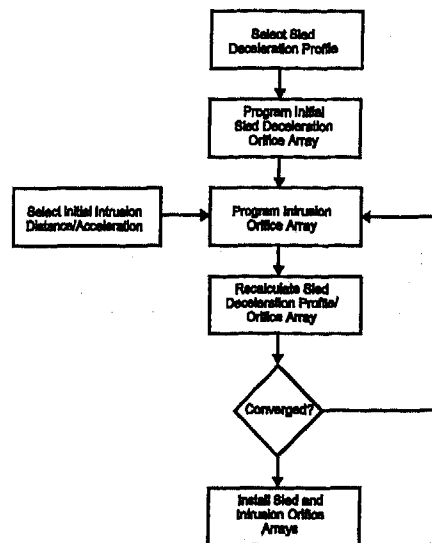


Figure 5: System A - Intrusion Deceleration Pulse/Sled Deceleration Pulse Calculation Procedure

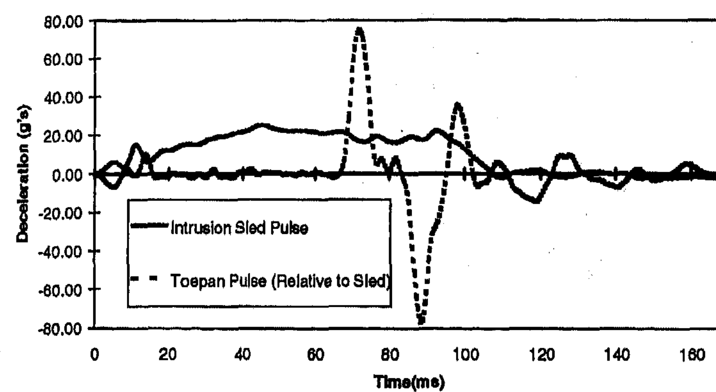


Figure 6: System A - Sled Deceleration vs. Toepan Acceleration in Sled Frame: 7 cm Intrusion (Signals Filtered to SAE J-211 Channel Class 60)

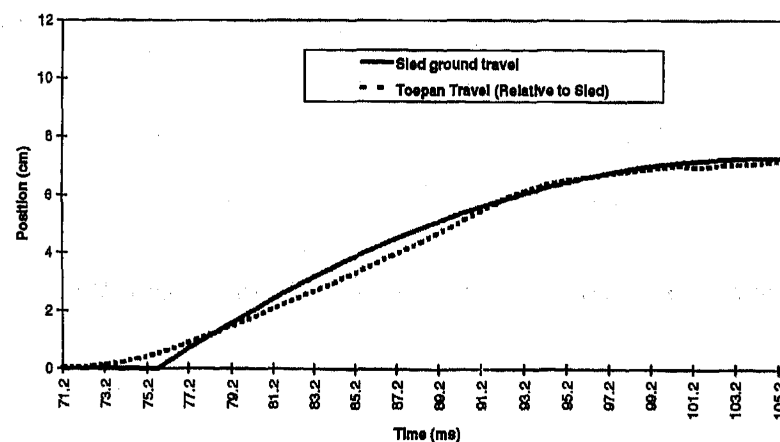


Figure 7: System A - Sled Travel Relative To The Ground and Toepan Intrusion Relative to the Test Sled (Signals Filtered to SAE J-211 Channel Class 180)

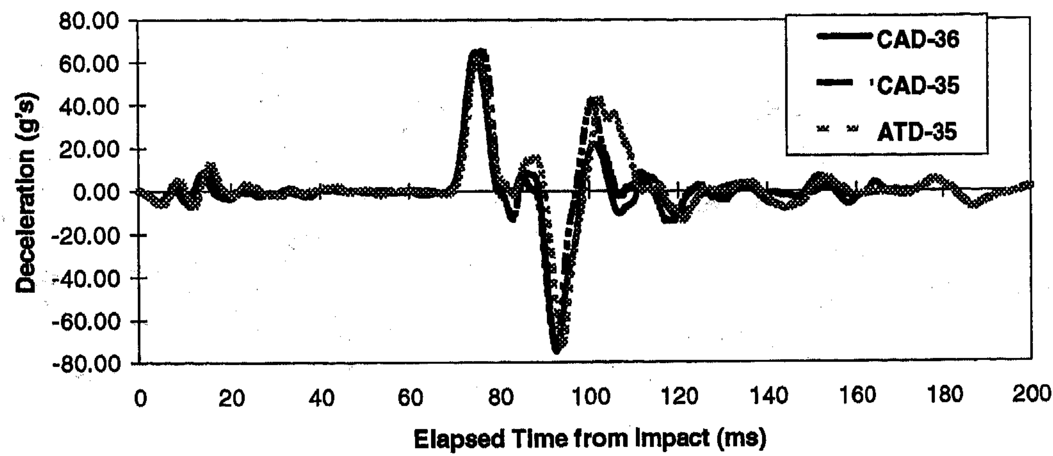


Figure 8: System A - Intrusion System Repeatability, Toepan Carriage Acceleration Relative to Test Sled, 7 cm Intrusion

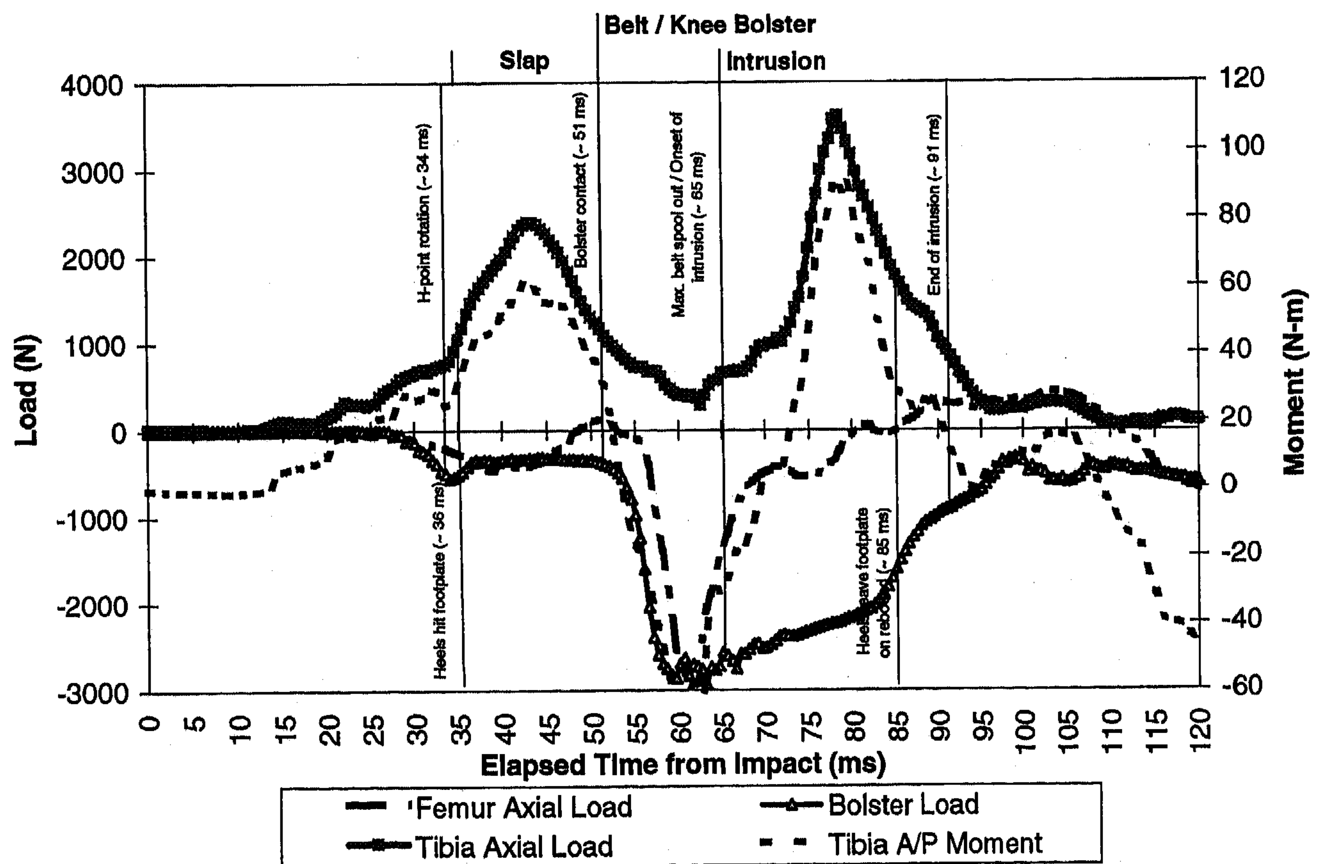


Figure 9: System A - Outboard Driver Side Lower Extremity Response to 7 cm Intrusion

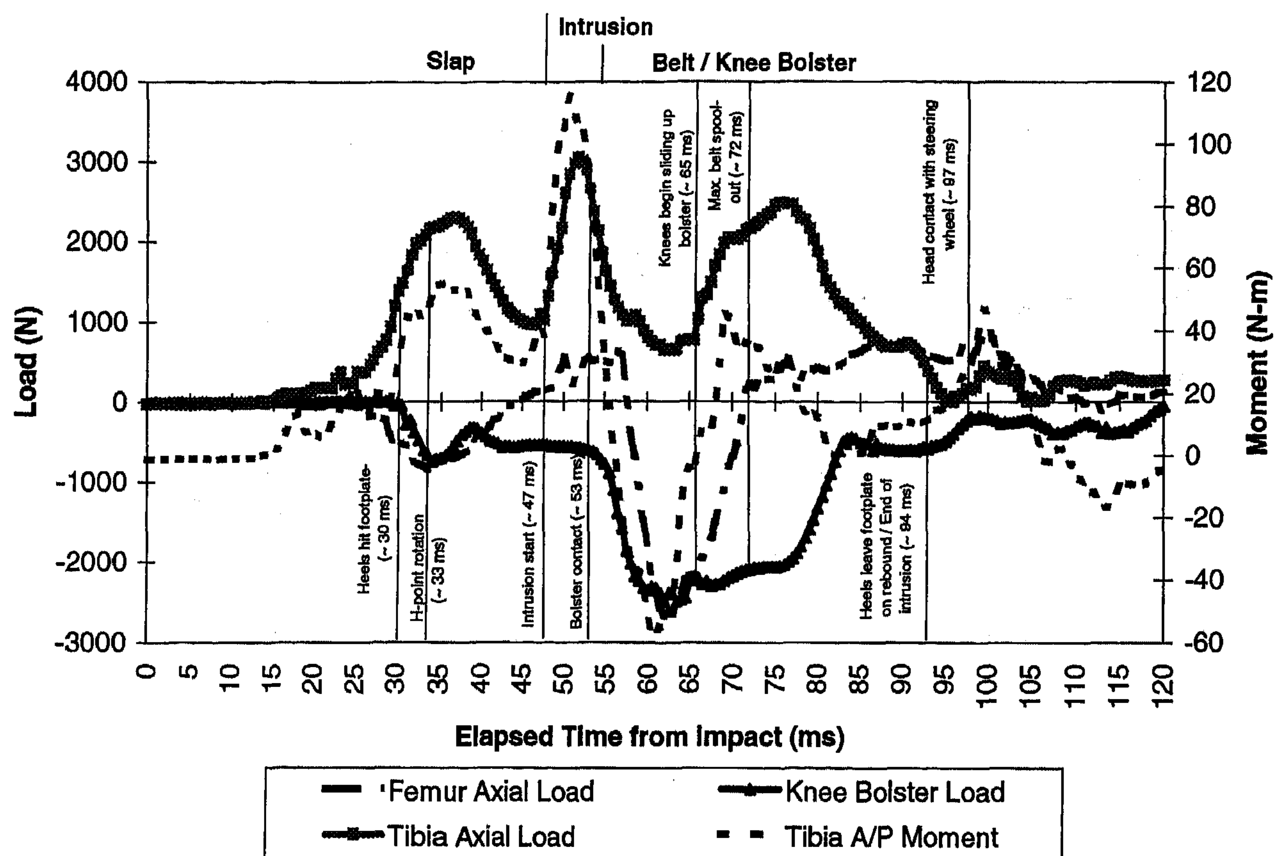


Figure 10: System A - Outboard Driver Side Lower Extremity Response to 22 cm Intrusion

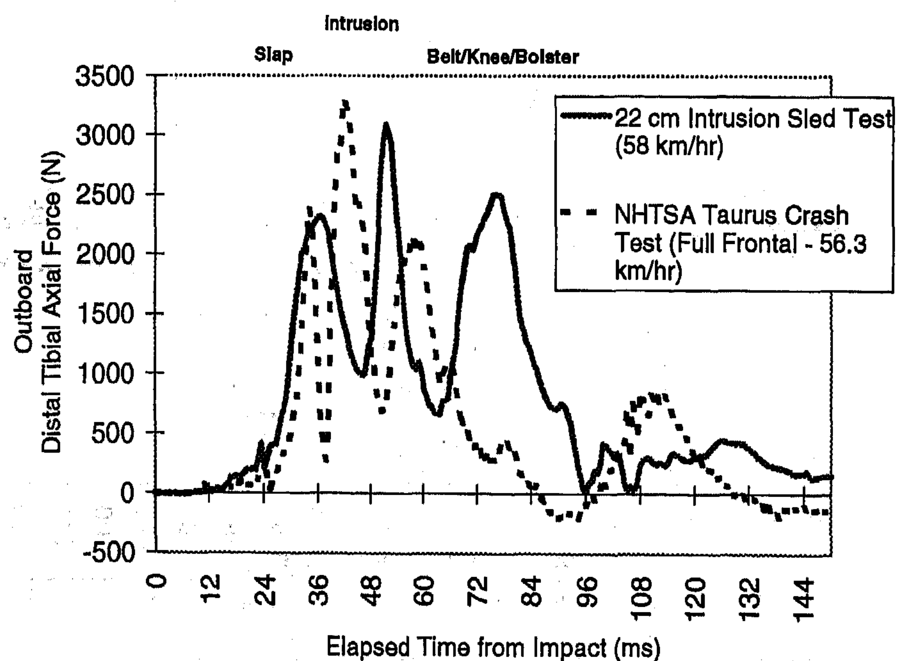


Figure 11: System A - Comparison of NHTSA Taurus Full Frontal Crash with 22 cm Intrusion Sled Test, Outboard Lower Extremity

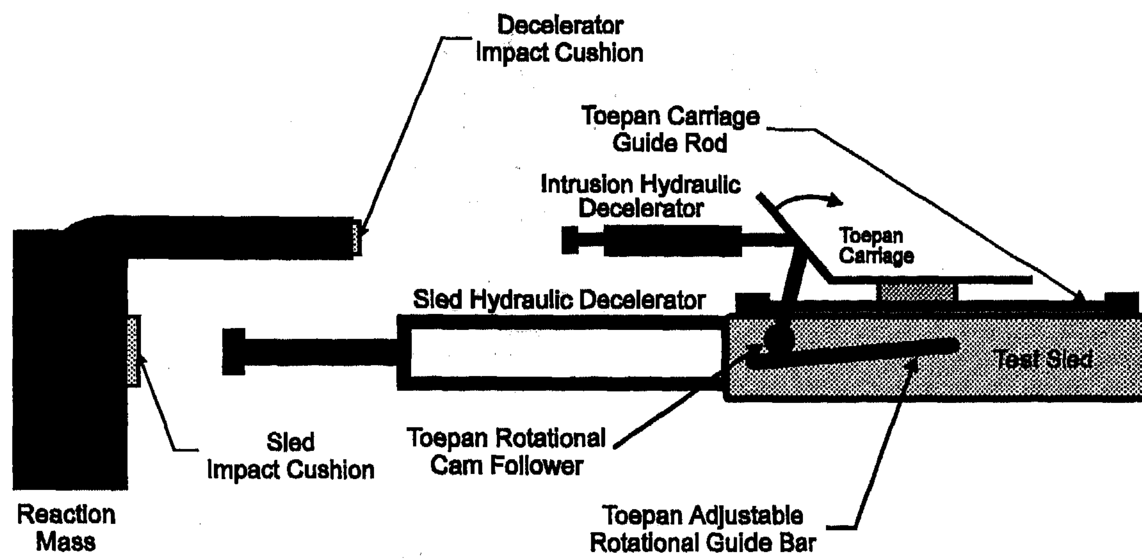


Figure 12: System B - Intrusion System Schematic

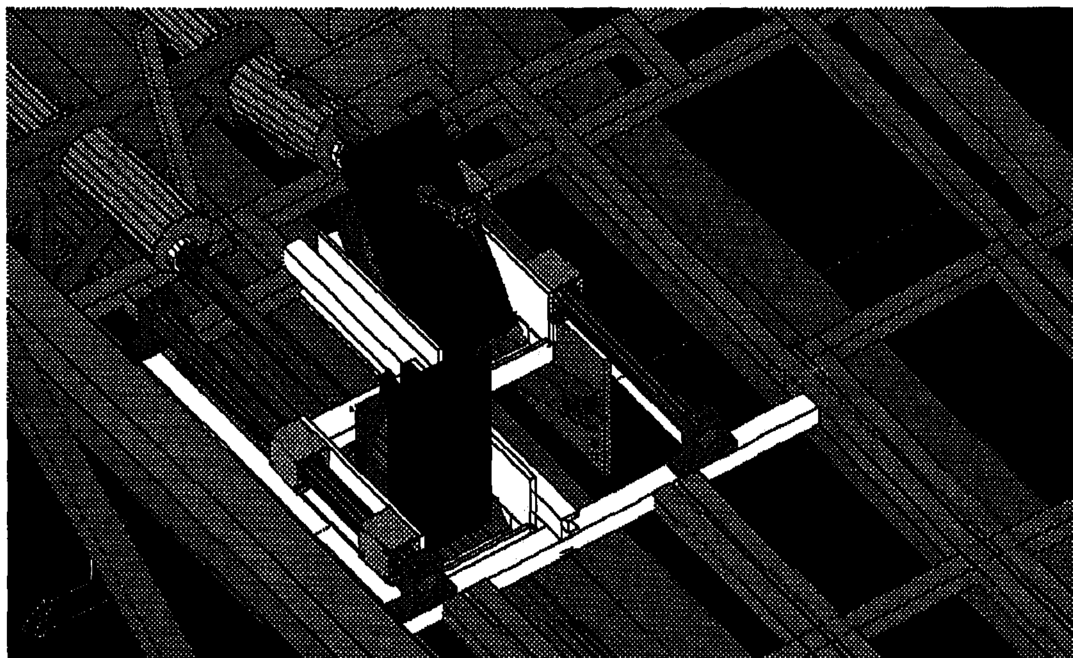


Figure 13: System B - Intrusion System Drawing

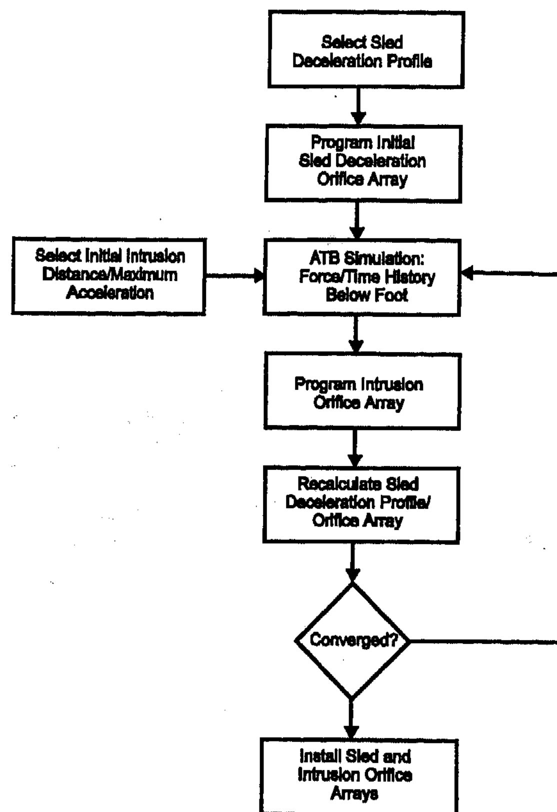


Figure 14: System B - Intrusion Deceleration Pulse/Sled Deceleration Pulse Calculation Procedure

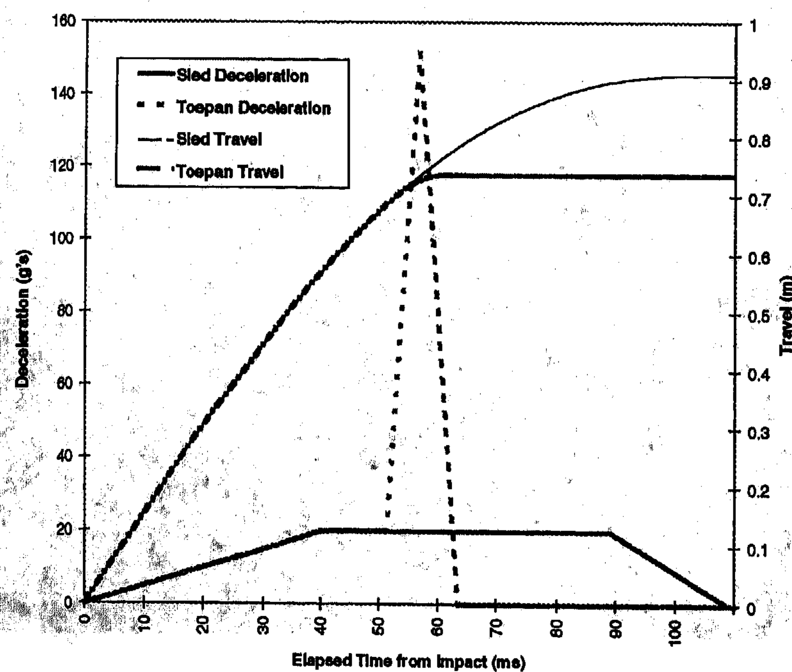


Figure 15: System B - Idealized Intrusion and Sled Acceleration and Post-Impact Travel

970575

Measurement Techniques for Angular Velocity and Acceleration in an Impact Environment

P. G. Martin, J. R. Crandall, and W. D. Pilkey
University of Virginia

C. C. Chou and B. B. Fileta
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The University of Virginia is investigating the use of a magnetohydrodynamic (MHD) angular rate sensor to measure head angular acceleration in impact testing. Output from the sensor, which measures angular velocity, must be differentiated to produce angular acceleration. As a precursor to their use in actual testing, a torsional pendulum was developed to analyze an MHD sensor's effectiveness in operating under impact conditions. Differentiated and digitally filtered sensor data provided a good match with the vibratory response of the pendulum for various magnitudes of angular acceleration. Subsequent head drop tests verified that MHD sensors are suitable for measuring head angular acceleration in impact testing.

INTRODUCTION

Head Injuries Due To Angular Acceleration

Brain injuries occurring in automobile accidents fall generally into two categories: focal brain injuries and diffuse brain injuries. Focal brain injuries occur as a result of head contact, or impact loading. Diffuse brain injuries occur as a result of sudden head motion, or inertial loading. Translation and rotation, independently or in combination, convey inertial loading. Diffuse Axonal Injury (DAI), in particular, has been linked to the presence of non-impact rotational acceleration during impact collisions (Gennarelli, 1993). Margulies, et al. (1990) subjected models of the skull-brain tissue structure to rotational accelerations of varying magnitudes. Results indicated that DAI, which manifests itself immediately after injury as a prolonged state of coma, may be experienced when the head encounters rotational accelerations on the order of 20,000 rad/s^2.

Although brain injuries usually result from a combination of contact loads and inertial loads, they can be produced by a single event such as pure head rotation. As such, measuring angular head acceleration is an important means of identifying injury potential. However, the current standard for measuring head injury potential in automobile collisions is the Head Injury Criterion (HIC), a value based upon the vector sum of translational accelerations in anthropomorphic dummy headparts. Angular accelerations alone have no affect on the HIC value.

Determining HIC in Cadaver Tests

Aside from its importance in assessing brain injury risk, angular head acceleration can be used in cadaver tests to find the HIC value. This is accomplished by attaching sensors to the outer surface of the cadaver's skull. The sensors measure the skull's angular velocity and acceleration (assuming the skull acts as a rigid body) and the linear acceleration of a point on the skull. This information is required to determine the acceleration of the cadaver head's center-of-gravity, which is then used to find the HIC value. In this manner, it is possible to compare HIC values derived in cadaver tests with those obtained in dummy tests.

MEASURING ANGULAR RATE

The angular velocity and acceleration of a rigid body can be measured using accelerometers arranged in an array, or with angular rate sensors. Several methods have been reported, such as a 3-2-2-2 accelerometer array described by Padgaonkar et al. (1975), a 3-3-3 accelerometer array described by Nusholtz (1993), and the use of three angular rate sensors described by Laughlin (1989).

The 3-2-2-2 Array

Padgoankar describes a method of measuring angular acceleration of a rigid body using nine accelerometers positioned in a 3-2-2-2 pattern. This configuration consists of a tri-axial accelerometer group at the body origin, and three pairs of accelerometers positioned at some distance on the three mutually orthogonal body axes. The appeal of this scheme lies in the parallelism of the three matched pairs of accelerometers which allows cancellation of the Coriolis components of angular acceleration. This leaves only the differences between the various pairs of accelerometer signals and their separations along the axes as the direct determinators of the angular acceleration components at each point in time.

The 3-3-3 Array

Nusholtz describes a measurement system of nine accelerometers arranged in a 3-3-3 array. This scheme applies spherical geometric analysis which exploits the topology of the sphere in its approach. Three tri-axial accelerometer groups are positioned in a circular pattern in manner that allows both angular acceleration and angular velocity to be obtained by manipulating the output from the linear accelerometers.

Nine accelerometer array schemes can be used to measure HIC and angular head acceleration. DiMasi (1995) illustrates how Padgoankar's 3-2-2-2 accelerometer array can be mounted within a Hybrid III head, and Nusholtz et al. (1985) has demonstrated how a 3-3-3 accelerometer array can be mounted to a cadaver head.

MHD Angular Rate Sensors

The recent development of a magnetohydrodynamic (MHD) angular rate sensor has generated new possibilities for measuring angular acceleration. MHD sensors, which measure angular velocity, have four components: the permanent magnet, the electrode areas, the conductive fluid, and the case. The conductive fluid lies in a constrained annular region within the field of the permanent magnet, which is attached to the case. The sensitive axis of the unit runs through the center of the annular region, perpendicular to the plane of the annulus. As the instrument rotates about the sensitive axis, the case, permanent magnet, and magnetic field rotate producing a magnetic field, ***B***. The conductive fluid, however remains stationary with respect to its own reference frame producing a relative motion, ***U***, between a magnetic field and a conductor. This motion produces a voltage, ***E***, governed by the relationship:

$$\boldsymbol{E} = \boldsymbol{U} \times \boldsymbol{B} \qquad [1]$$

Laughlin demonstrated that MHD sensor output (angular velocity) may be numerically differentiated to obtain accurate angular acceleration data. In his tests, data was collected from impacts of a Hybrid III dummy head fitted with both an MHD angular rate sensor and a 3-2-2-2 accelerometer array. Angular acceleration obtained by differentiating MHD sensor output was a close match to that obtained from the 3-2-2-2 array.

The MHD sensors used in Laughlin's tests were developed by Applied Technology Associates, Inc. (ATA) of Albuquerque, NM and are marketed as model number ARS-01. ATA has since developed an orthogonal set of three MHD sensors housed in a single 3.18 cm x 3.18 cm x 4.06 cm aluminum cube. This cube, referred to as the Dynacube, weighs 86 grams and furnishes separate measurements of angular velocity about three orthogonal axes. Furthermore, accelerometers can be mounted on the aluminum case to provide a 6-degree-of-freedom system. However, the Dynacube has only about half the dynamic range and half the resolution of the ARS-01.

Using Dynacubes to find angular acceleration

At the University of Virginia (UVa), an effort is being made to incorporate head response measurements using Dynacubes into dummy and cadaver impact tests. UVa already uses Dynacubes to measure foot and ankle rotation of dummies and cadavers in dynamic sled tests (Hall et al., 1996). Dynacubes have also been used by Klopp et al. (1995) at UVa to record whole-body rotation in tests involving the Side Impact Dummy (SID).

Rather than using a nine accelerometer array, UVa is considering extending the use of the Dynacube to determine angular head acceleration. A Dynacube is easy to mount and handle, it is light-weight and compact, and it does not greatly influence the movement of the body part upon which it is attached. Moreover, the HIC value and angular head acceleration can be evaluated using three accelerometers mounted to one Dynacube for a total of just six data channels.

VALIDATION OF MEASUREMENT METHODOLOGY

Notwithstanding Laughlin's reported success using the ATA model ARS-01 MHD sensor to determine angular acceleration in an impact environment, the effectiveness of the Dynacube, given its lower resolution and lower dynamic range, is not as certain. Therefore, UVa has developed a test fixture that generates predictable angular velocities and accelerations of similar magnitude to those seen in head impact tests against which sensor output may be compared. The test fixture consists of a torsional pendulum that is capable of handling various angular rate measuring instruments, including a Dynacube.

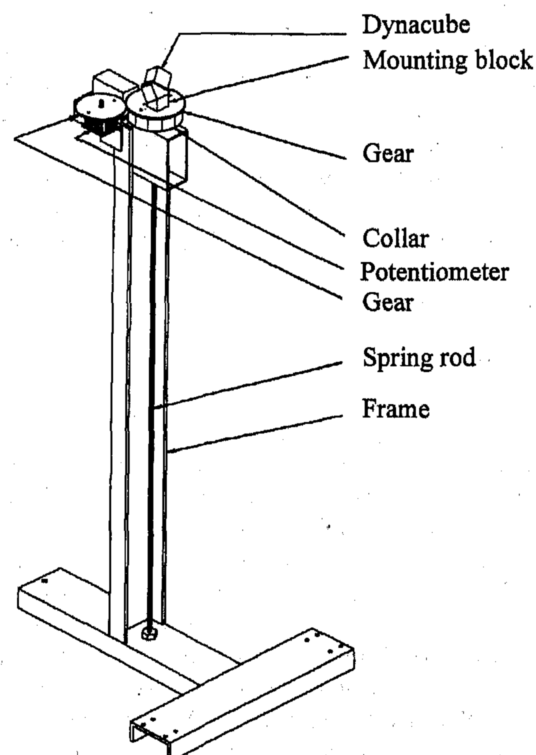


Figure 1. Torsional pendulum.

Torsional Pendulum

The torsional pendulum is a single-degree-of-freedom system constrained to rotate about its axis. When the pendulum disk is rotated and released suddenly, it undergoes free vibration that follows a simple damped harmonic motion. The device is used to compare the output of a three-axis Dynacube (or other rate sensors) with the theoretical response of the harmonic system. The pendulum also serves as a means to check sensor outputs for any signal irregularities.

Description of the pendulum

The pendulum is a very simple mechanical structure consisting of the following components:

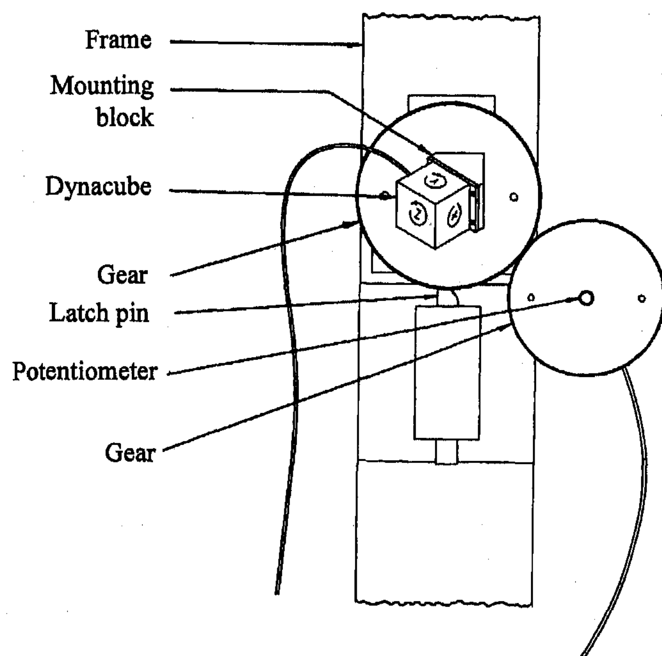


Figure 2. Top View of Pendulum.

Frame - Constructed of steel channels, the rigid frame houses the pendulum spring. Radial bearings built into the frame constrain motion of the spring to rotation about its axis only. The size and weight of the frame assure that it will not vibrate significantly as energy is released by the oscillating pendulum.

Spring Rod - A 3-meter long steel rod provides the system with enough stiffness to produce angular accelerations up to 20,000 rad/sec². Rods of various diameters may be interchanged to produce different frequencies.

Collar - The collar, or pendulum disk, is bolted directly to the top end of the spring rod. Markings on the collar allow the amplitudes of the pendulum to be easily tracked by a high speed video camera positioned above the apparatus. The collar also provides a surface onto which the Dynacube (or other instrumentation) may be mounted.

Potentiometer - The collar is connected by gears to a rotary potentiometer. Potentiometer measures the position of the collar and serves as a check against the theoretical predictions and Dynacube data.

Mounting block - The block enables the Dynacube to be mounted so that all three of its axes form an equal angle with the pendulum's axis of rotation. When mounted as such, each Dynacube axis should produce the same response. Thus, the pendulum serves as a means to check the Dynacube's sensitivity factors, and enables them to be adjusted if needed.

Latch pin - Upon setting the collar at an initial angle of twist, the collar is released suddenly by disengaging the latch pin. This sends the pendulum immediately into a state of free vibration.

Theoretical response

The movement of the pendulum is assumed to obey simple harmonic motion subjected to viscous damping. The equation for the angular position as a function of time for the free vibration of such a system is:

$$\theta(t) = \mathrm{Exp}(-\zeta\,\omega_n\,t)\cdot\{\theta_0\,\mathrm{Cos}[\omega_d\,t] + ((\zeta\,\theta_0\omega_n + \dot{\theta}_0)/\omega_d)\,\mathrm{Sin}[\omega_d\,t]\} \quad [2]$$

where:

ω_n = Natural frequency of the pendulum
ω_d = Damped natural frequency
θ_0 = Initial angular position (angle of twist)
$\dot{\theta}_0$ = Initial angular velocity
ζ = Damping factor

The damped natural frequency is related to the undamped natural frequency by:

$$\omega_d = \sqrt{1-\zeta}\;\omega_n \quad [3]$$

and the initial position, θ_0, is set by adjusting the position of the latch pin. The initial velocity, $\dot{\theta}_0$, is zero since the pendulum is released from rest. Moreover, it is possible to calculate the natural frequency of the pendulum from the spring constant, mass, and moment of inertial of its components. These values can be obtained either experimentally or from published material property lists.

Unfortunately, damping arises in a manner that does not allow the damping factor to be determined directly. Damping in the pendulum results from looseness of joints, internal damping in the spring rod, friction within the bearings, and does not result from some viscous dashpot. It is not possible to treat all of these forms of damping. However, since the amplitude of the vibration follows an approximately exponential decay, viscous damping is used in the mathematical model of the system.

The angular velocity and angular acceleration is determined by taking the time derivative of the position:

angular velocity - $$\dot{\theta}(t) = \frac{d}{dt}\theta(t) \quad [4]$$

angular acceleration - $$\ddot{\theta}(t) = \frac{d^2}{dt^2}\theta(t) \quad [5]$$

Pendulum Operation

Operation of the pendulum is simple. A Dynacube is affixed to the mounting block and the block is attached to the pendulum collar. The collar is rotated to a given angle and held in place by the latch pin. The collar may be rotated to any angle up to 45°, and variations in the angle of rotation may be used to control the amplitude. The latch pin is then quickly disengaged, sending the pendulum into a state of free vibration. As the pendulum oscillates, the rotary potentiometer tracks its position. Meantime, a high-speed camera (Kodak Ektapro model 1012) operating at 10000 frames per second captures the event on video tape.

Data Analysis

Data analysis is performed on three levels: displacement, velocity, and acceleration. Displacement analysis is performed first, with the validity of the potentiometer data verified against the high-speed video data. Since the pendulum displacement is measured directly by two independent sources, it is on this level that the parameter constants for the equation of motion of the pendulum are determined. Parameter constants are initially estimated from the material properties of the pendulum, and then adjusted so that the theoretical response fits the potentiometer/camera data. These same constants are subsequently used in analyses at the velocity and acceleration levels.

At the velocity level, Dynacube data is compared with potentiometer data and the theoretical response. First, the potentiometer data is differentiated using a 3-point differentiation scheme, and then digitally filtered using a 2-pass second order Butterworth algorithm evaluated at 200 Hz. Then, the Dynacube data is analyzed for consistency. All three Dynacube channels should have identical output. If they do not, it indicates one or more calibration factors need to be adjusted. In this fashion, the pendulum serves as a means to check angular rate sensors for any signal irregularities and determine correction factors.

After the calibration factors are adjusted, all three Dynacube outputs should be nearly identical to each other and to the differentiated potentiometer data scaled by $\sqrt{3}$. Finally, the corrected Dynacube data must be transformed from the local Dynacube reference frame to the collar reference frame using the following transformation:

$$\begin{bmatrix} X_{COL} \\ Y_{COL} \\ Z_{COL} \end{bmatrix} = \begin{bmatrix} X_{MHD} & Y_{MHD} & Z_{MHD} \end{bmatrix} \cdot \begin{bmatrix} Cos(\alpha) & Sin(\alpha)\cdot Sin(\beta) & Sin(\alpha)\cdot Cos(\beta) \\ 0 & -Cos(\beta) & Sin(\beta) \\ -Sin(\alpha) & Cos(\alpha)\cdot Sin(\beta) & Cos(\alpha)\cdot Cos(\beta) \end{bmatrix} \quad [6]$$

where α and β define the orientation of the Dynacube with respect to the collar. The rotation angles formed by attaching the Dynacube to the mounting block are:

α = rotation about the y-axis = 45°
β = rotation about the x-axis = 35.26°

When using the mounting block, X_{COLLAR} and Y_{COLLAR} should approach zero after the transformation is complete.

At the acceleration level Dynacube data is again compared with the theoretical response and potentiometer data. The potentiometer output is differentiated and filtered a second time. The three corrected Dynacube outputs are first differentiated using a 3-point differentiation scheme, then digitally filtered using a 2-pass second order Butterworth algorithm, and finally transformed to the collar reference frame.

RESULTS

Results for a typical test are shown in Figures 3 - 6. Figure 3 shows the pendulum displacement time-history. The potentiometer data and the fitted theoretical data are close, but not a perfect match due mainly to the fact that the damping is not purely viscous.

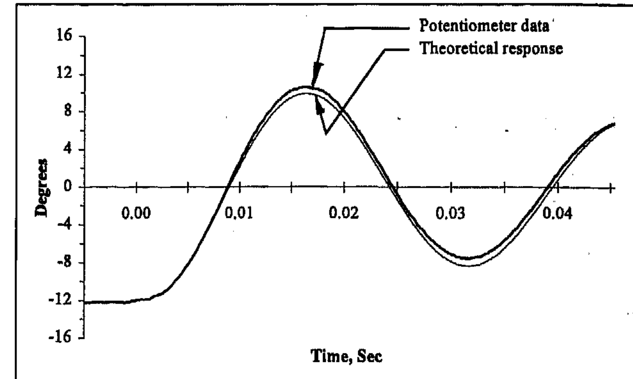


Figure 3. Angular displacement.

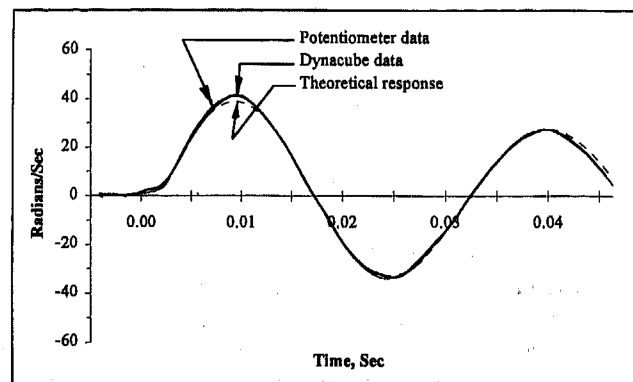


Figure 4. Angular Velocity

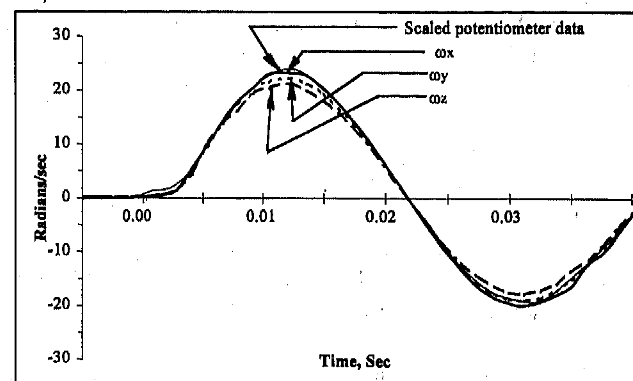


Figure 5. Dynacube output.

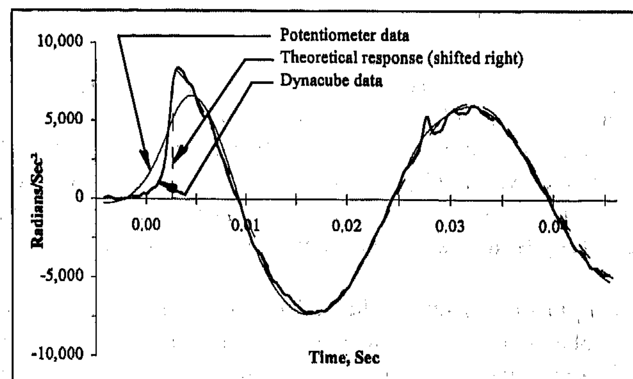


Figure 6. Angular acceleration.

Figure 4 shows the Dynacube output for the test and illustrates a slight variance in the amplitudes of each channel. Changing the position of the Dynacube on the mount and re-running the test at different speeds and frequencies verified that the inconsistent output was caused by inaccurate Dynacube sensitivity factors, and not a consequence of out-of-plane movement in the pendulum due to looseness in bearings and joints. The output was adjusted so that all three outputs most closely matched each other and the potentiometer data.

Figure 5 shows the pendulum velocity time-histories obtained from three sources: Dynacube data corrected and transformed to the collar reference frame; potentiometer data differentiated and filtered at 200 Hz; and the theoretical response. There is almost a complete overlap of Dynacube and potentiometer data indicating that the Dynacube works very well in measuring angular velocity.

Figure 6 shows the pendulum acceleration time-history of the three sources: Dynacube data corrected, differentiated, filtered at 600 Hz, and transformed to the collar reference frame; potentiometer data differentiated twice and filtered at 100 Hz; and the theoretical response. The theoretical response has been shifted manually to the right to be in phase with the Dynacube and potentiometer data. The out-of-phase condition occurs during the brief moment when the latch pin disengages from the collar.

The data match is not as precise as that seen on the velocity level, but it is still very close, particulary between the Dynacube and the theoretical response at the initial peak. The initial acceleration peak is oftentimes the highest in actual head impacts. This peak has been lost in the potentiometer data due to the heavy filtering.

Other tests were performed using the same Dynacube at higher and lower pendulum frequencies and amplitudes. In all cases, Dynacube data differentiated and filtered at 600 Hz produced a good correlation with the theoretical angular acceleration, like that seen in Figure 6.

Dynacube Application

A dummy head drop test was conducted to determine if the pendulum response is representative of an actual impact. The behavior of a Dynacube in an actual head impact was also examined. The test was configured so that angular motion was constrained to the head's y-axis. A Dynacube was placed inside the dummy head, and three linear accelerometers were affixed to the Dynacube to produce a 6-degree-of-freedom measurement system.

In Figure 7, the unfiltered vs. filtered (600 Hz) Dynacube output is shown. The frequency and amplitude are similar to those that can be produced by the pendulum, confirming that the pendulum can provide an impulse comparable to an actual head impact. As in the pendulum tests, the 600 Hz filter is seen to remove only the very high-frequency components. The actual angular acceleration of the head is unknown, but an analogy with the pendulum test would indicate that it is very close to the filtered Dynacube data. As a final note, this head impact was considered moderate by the current head injury standard: the calculated HIC value was 570.

CONCLUSIONS

The measurement of angular acceleration in head impact studies has gained recognition as an important indicator of head injury potential. Although there are several methods of measuring angular acceleration in a rigid body, no standard measurement protocol has yet been recognized by all testing agencies.

Differentiated output from a Dynacube, an angular rate sensor marketed by ATA, Inc., simplifies the measurement of angular head acceleration in impact tests because it is compact, easy to handle, and requires only three data channels. However, limited information was available on the reliability of differentiated Dynacube output. UVa therefore constructed a torsional pendulum to study Dynacube output and to develop a measurement protocol to determine angular acceleration. Two important findings came out of this study:

1. Dynacube output filtered at 600 Hz provides a good measure of angular acceleration for head impacts in automobile crashes.

2. Although the three orthogonal Dynacube outputs are not always consistent with one another, the pendulum can be used to determine appropriate correction factors.

Future analyses of head, limb, and torso kinematics in crash tests may be expanded by using angular acceleration measurements determined with Dynacubes.

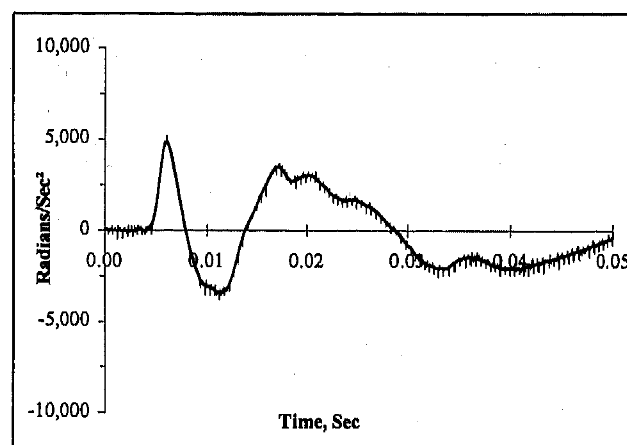


Figure 7. Dynacube output from head drop test. Filtered vs. unfiltered data.

ACKNOWLEDGMENTS

The angular rate sensor analysis and measurement methodology was supported in part by the Ford Motor Company and the University of Virginia School of Engineering and Applied Science. Angular rate sensors were provided by ATA, Inc.

REFERENCES

1. DiMasi, F.P., "Transformation of Nine-Accelerometer-Package (NAP) Data for Replicating Headpart Kinematics and

Dynamic Loading", NHTSA Technical Report DOT-HS-808-282, National Technical Information Service, Springfield, VA; 1995.

2. Gennarelli, T.A., "Mechanisms of Brain Injury", Journal of Emergency Medicine, Vol. 11, pp 5-11; 1993.

3. Hall, G.W., Crandall J.R., Klisch S.M., Klopp G.S., Pilkey W.D., "Measurement of Dynamic Joint Motion using Magnetohydrodynamic Angular Rate Sensors", Shock and Vibration Digest, Vol. 28, No. 5, pp 12-17; 1996.

4. Klopp, G.S., Hall, G.W., Crandall, J.R., Pilkey, W.D., "Measurement of Rotational Kinematics using Magnetohydrodynamic Angular Rate Sensors", Proc. 23rd International Workshop on Human Subjects for Biomechanical Research, San Diego, CA; November, 1995.

5. Laughlin, Darren R., "A Magnetohydrodynamic Angular Motion Sensor for Anthropomorphic Test Device Instrumentation", SAE Paper No. 892418, Society of Automotive Engineers, Warrendale, PA; 1989.

6. Margulies S.S., Thibault E.T., Gennarelli T.A., "Physical Model Simulations of Brain Injury in the Primate", Journal of Biomechanics Vol. 23, No. 8, pp 823 836; 1990.

7. Nusholtz G.S., "Geometric Methods in Determining Rigid-body Dynamics", Experimental Mechanics, Vol. 33, pp 153-158; June 1993.

8. Nusholtz G.S., Kaiker P.S., Muscott G.J., Suggitt B.R., "UMTRI Experimental Techniques in Head Injury Research", SAE Paper No. 851244, Society of Automotive Engineers, Warrendale, PA; 1985.

9. Padgaonkar, A.J., Krieger, K.W., King A.I., "Measurement of Angular Acceleration of a Rigid Body using Linear Accelerometers." Journal of Applied Mechanics, ASME Preprint 75-APMB-3, pp 552-556; September, 1975.

970583

A Study on the Improvement of the Structural Joint Stiffness for Aluminum BIW

Young Woong Lee, Yong Woo Kwon, Soon Yong Kwon, and Won Suk Cho
Kia Motors Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The application of aluminum Body-In-White has several technical barriers in press forming, joining, and chemical conversion treatment processes. Among them, the optimization of joining processes with which structural stiffness and durability will be assured might have the key role for the success of aluminum applications to BIW.

In this study, stiffness, strength and fatigue strength of BIW joints with aluminum sheets were evaluated as a function of joining methods, such as resistance spot welding and weld bonding, via both experimental and analytical routes(FEM). For the evaluation, single-lap joint and T-shaped joint were made, with each joining method, as variation of pitch, sheet thickness and even materials - steel sheets.

Based on the experimental and FEM analysis results, the optimum joining method for the aluminum BIW which is suitable for weight saving and has equivalent stiffness and strength to steel is suggested. In addition, for obtaining optimum processes with improved stiffness and strength of the joining method, design parameter, such as the influence of the spot pitch on the structural joint stiffness and the minimum wall thickness of aluminum which is able to maintain the equivalent stiffness and strength to steel are discussed.

INTRODUCTION

The automotive body which maintains the shape of a vehicle and has to endure the interior and exterior load is comprised of many joints. Since the structural joint stiffness, especially between pillar and rocker, pillar and roof rail affects not only the body stiffness(such as bending or torsional stiffness) but also body strength, vibration, and crash characteristics, it is an important factor to be considered carefully in designing of a vehicle.[1,2]

In the case of aluminum body, the structural joint stiffness should be looked over more carefully, because of its difficulties in joining.

To get the equivalent structural joint stiffness of aluminum body with conventional steel body, it is the purpose of this study to investigate the relationship in the simple specimens between joint properties and process, such as spot pitch, sheet thickness, even adding adhesives(weld bonding) replacing conventional spot welding, via experimental and FEM. Weld bonding offers increased joint stiffness, more even load distribution, and improved fatigue performance. [3,4,5] In addition, because of insufficient open data for analyzing whole structural stiffness of real automotive body, FEM simulation for BIW with above mentioned process is performed.

SINGLE-LAP JOINT

First of all, parameter was evaluated with very simple model or specimens, such as single-lap joint as shown in Fig. 1. Aluminum 5xxx alloy for automotive body and mild steel sheets were used. (See Table 1) According to the plate theory, the stiffness of sheets are proportional to the cube of the thickness multiplied by the modulus of elasticity. Therefore, in order to achieve the same stiffness, the aluminum sheets would have to be about 1.4 times as thick as the steel one, making

it about half the weight.[2] The thickness of steel and aluminum sheets used for the evaluation of single lap joint were 0.7 mm, 1.0 mm respectively. The number of spot weld in aluminum sheets were two(spot pitch =80 mm), three (spot pitch = 40 mm) and five (spot pitch = 20 mm), and the number of spot weld in steel sheets were two and three.

Table 1 : Mechanical Properties of Aluminum and Steel

Property / Material	Elastic modulus (MPa)	Poissons ratio	density (kg/m³)	Yield strength (MPa)	U.T.S (MPa)	elongation (%)
Aluminum	7×10^4	0.33	2700	130	280	35
Steel	20.5×10^4	0.33	7800	180	320	45

SHEAR STIFFNESS - Fig. 2 shows analytical and experimental results of shear stiffness as a function of spot pitch in case of aluminum single-lap joint. As expected, for each spot pitch, weld bonded joint shows higher stiffness, while the difference in stiffness between weld bonded and spot welded joint increase with increasing spot pitch. This reason can be explained as : the local stress concentration in spot welded joint is able to be released by distributing the stress uniformly in weld bonded joint with adhesive layer. Because spot welded part is unbonded area in weld bonded joint, the part of local stress concentration decrease with the increase of the spot pitch, therefore, stiffness and strength will become higher.[3,4]

To confirm the experimental results, FEM analysis was developed, as shown in Fig. 3. A linear static analysis was performed using MSC/NASTRAN software. Various FE modeling have been used for the spot welded and adhesive bonded parts.[3,6,7] Among them, the adopted modeling method was as follows : In the spot welded structural model, 3-D linear quadrilateral shell element was applied for adherend sheet, and spot welded part was connected with rigid element, respectively. Then, bar element was inserted at the ends of the rigid element, In the weld bonded structural model, spring element that has three directional stiffness was adopted for adhesive layer.

As shown in Fig. 2, FEM analysis agrees well with experimental result. Therefore, this FE modeling method was also applied to T-shaped joint and structural joint of real automotive body, mentioned in next section.

STATIC STRENGTH - Table 2 gives the optimized welding conditions. An adhesive used in weld bonding was single component, heat cured epoxy type. Its curing was completed at 170℃ for

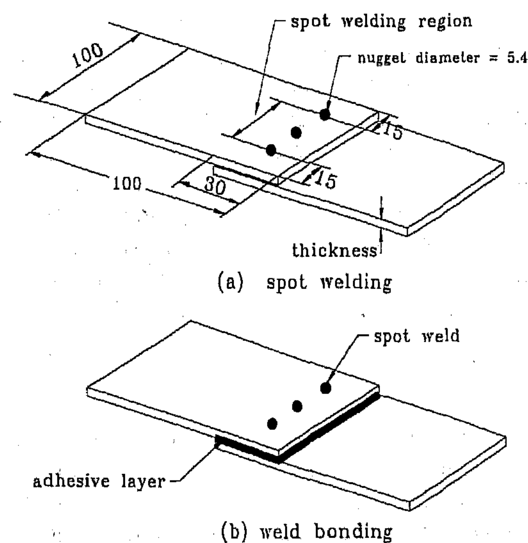


Fig. 1 : Schematic Illustrations of Single-Lap Joint (unit : mm)

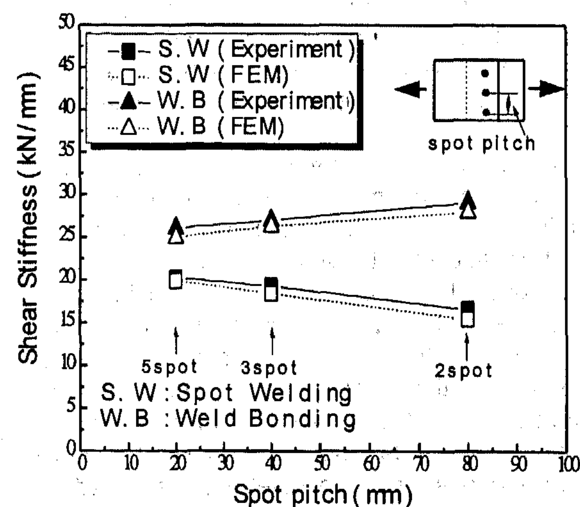
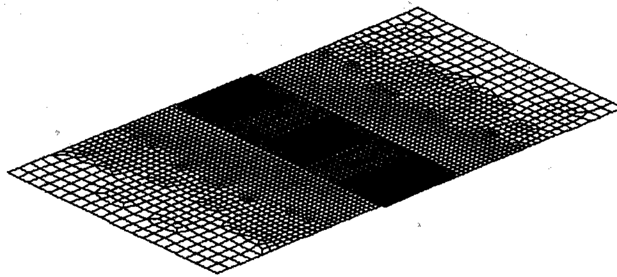
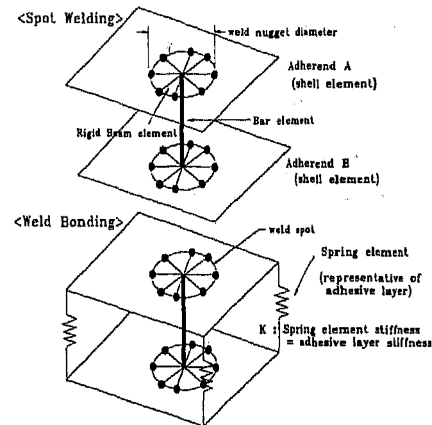


Fig. 2 : Shear-Stiffness of Aluminum Single-Lap Joint





(a) FE mesh for single-lap joint (b) modeling of spot welded and weld bonded part

Fig. 3 : FE Model of Single Lap Joint

30 minutes. Table 3 shows the general properties of the adhesive.

Fig. 4 exhibits lap-shear strengths of single lap joint vs. spot pitch, in a similar manner as showed in Fig. 2. For comparison, data for steel sheets were also plotted.

Table 2 : Welding Parameters

parameter / material	Welding current(kA)	Welding time(cycle)	Electrode force(kN)
Aluminum	28.0	9.0	3.43
Steel	8.5	15.0	2.45

Table 3 : Properties of Adhesive

Viscosity (cps/20 °C)	Specific gravity (20 °C)	Shear stress (MPa)	Young's modulus (MPa)	Poissons ratio
10^6	1.62	27.0	4.83×10^3	0.33

FATIGUE - Results of fatigue strength of single-lap joint performed under the condition of R = 0(where, R = min/max load) and 70Hz frequency were shown in Fig. 5. Increasing number of spot weld caused improved fatigue performance, while a little moderate decreased fatigue strength was obtained in weld bonded aluminum sheets. This phenomenon might be quite consistent with the results of shear stiffness and strength.

The superior fatigue strength of weld bonding was proved quite well even in aluminum sheets when compared with the results of conventional spot welded steel sheets. (even 3 times higher)

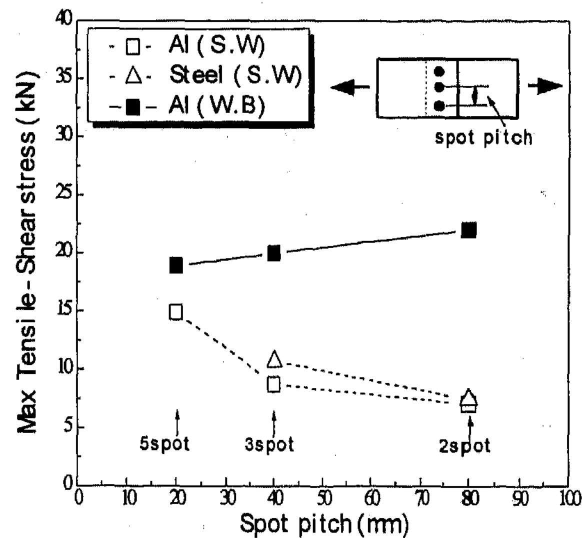


Fig. 4 : Static Lap - Shear Strength of Single Lap Joint

T-SHAPED JOINT

Automotive body structures are comprised of structural joints of various shapes. The configuration of the structural joints connecting structural members are classified into three typical shapes, such as L-shaped joint as Rocker Panel to

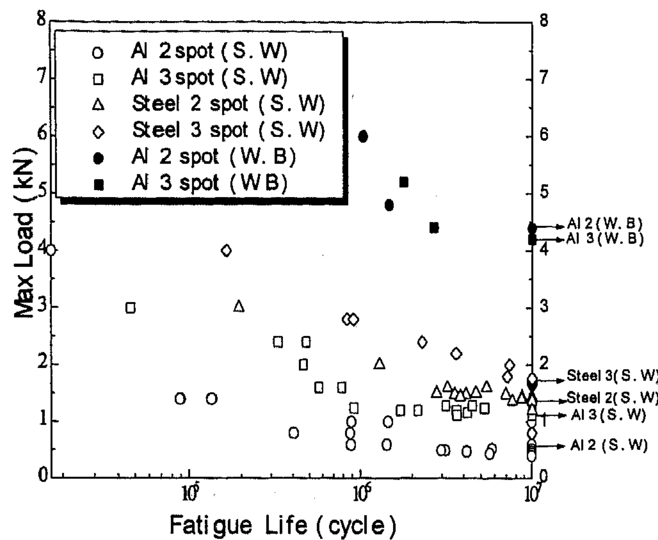


Fig. 5 : Fatigue Test Results of Single-Lap Joint

Front Pillar, T-shaped joint as Rocker Panel to Center Pillars, and Y-shaped joint as Front Pillar to Roof Rail. Among them, T-shaped joint was selected, for the joint requires higher stiffness and strength than the other two shapes. The structural joint stiffness was obtained analytically (FEM) and experimentally as variation of joining parameters : spot pitch, sheet thickness and joining method as summarized in Table 4.

Table 4 : Joining Parameters of T-shaped Joint

Thickness(mm) & Material / Spot pitch (mm)	Al & Steel 1.0 & 0.7	Al 1.5	Al 2.0
17.5 ● ● ● ● ●	S. W		
35 ● ● ●	S. W W. B	S. W	S. W
70 ● ● ●	S. W		

THE METHOD OF FEM ANALYSIS AND STRUCTURAL JOINT STIFFNESS TEST - The shape of T-shaped joint and boundary condition are shown in Fig. 6 and FE model of T-shaped joint is shown in Fig. 7. The shape of T-shaped joint was simplified to be handled easily. FE model shown in Fig. 7 describes T-shaped joint of spot welded aluminum. FE model of spot welded part and weld bonded part was equal to the model applied to single lap joint described already.

Structural joint stiffness is classified into three values : In-plane bending stiffness, Out-plane bending stiffness and Torsional stiffness. In this paper, however, only In-plane bending stiffness was measured in order to compare analytical results with experimental ones easily. According to the previous research[8], structural joint stiffness(K_j) was defined as follows.

$$\frac{1}{K_t} = \frac{1}{K_j} + \frac{1}{K_1} + \frac{1}{K_2} \qquad (1)$$

$$K_t = \frac{F \cdot L_1}{\Phi} \quad \text{at tip : Global stiffness}$$

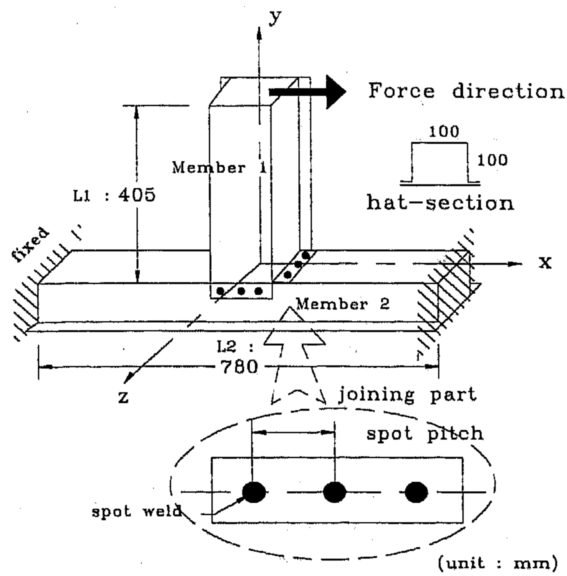
$$K_1 = \frac{EI}{L_1} \quad \text{: rotational stiffness of member1}$$

$$K_2 = \frac{8EI}{L_2} \quad \text{: rotational stiffness of member2}$$

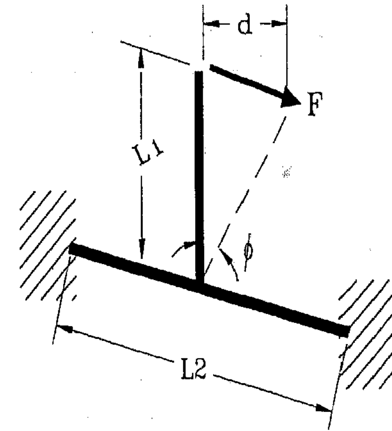
In Eq.(1), 1, 2, and j denotes member 1, 2 and joining part respectively. Eq.(1) represents that the work done by external force is equal to the sum of energy stored in member 1, 2 and joining part. Therefore, the structural joint stiffness of T-shaped joint, K_j, in Eq.(1) was calculated by FEM analysis as a function of joining parameters in Table 4.

To check the validity of the FEM analysis result, structural joint stiffness test was performed under the same joining parameters as with analytical method. The schematic drawing of the stiffness test apparatus are shown in Fig. 8. Spot welded and weld bonded portions of T-shaped joint specimens were made under the same condition as described for single lap joint.

RESULTS AND DISCUSSION - Results of structural joint stiffness from analytical and experimental method in T-shaped joint are shown in Fig. 9 and 10 with various spot number and sheet thickness. Fig. 11(a) shows global deformation of T-shaped joint as a result of analysis and Fig. 11(b) shows local buckling that happens in joining part as a result of experimental test. It is seen from Fig. 9 and Fig. 10 that structural stiffness of T-shaped joint is proportional to sheet thickness and spot number, and joint



(a) geometry and dimension
(in case of 3 spot welding)



(b) boundary condition

Fig. 6 : The Geometry and Dimension of T-shaped Joint

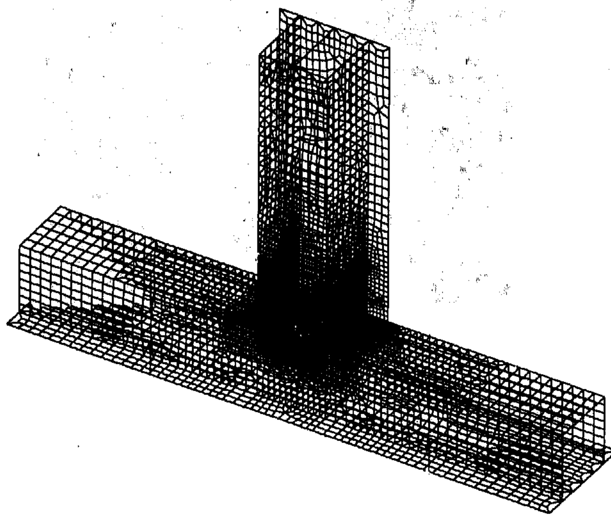


Fig. 7 : FE model of T-shaped Joint

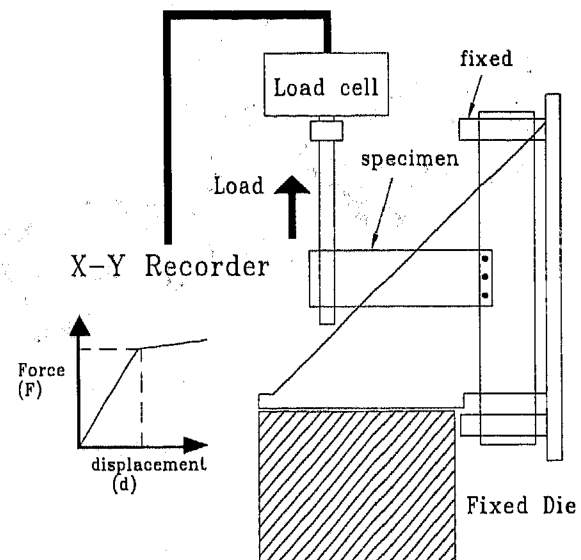


Fig. 8 : Equipment of Structural Joint Stiffness Test

stiffness was more affected by sheet thickness, which is contributed to strengthening base material, than spot number in spot welding. In other words, joint stiffness is more affected by strengthening of base material than increase of spot number. Fig. 11(b) shows base material fractured near spot welded part, and make the phenomenon clear. Therefore, the increase of sheet thickness is more effective to enhance joint stiffness than the increase of spot number in spot welding.

Comparing the joint stiffness of 1.0 mm thick aluminum with that of 0.7 mm thick steel under three welded spot from Fig. 9, it can be noticed that the stiffness of spot welded aluminum and

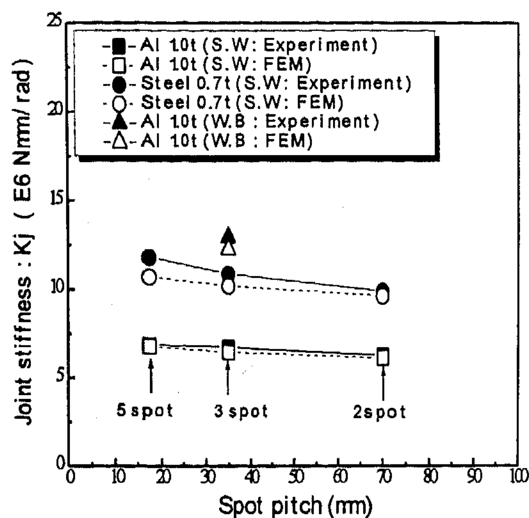


Fig. 9 : Effect of Spot Pitch on the T-shaped Joint Stiffness

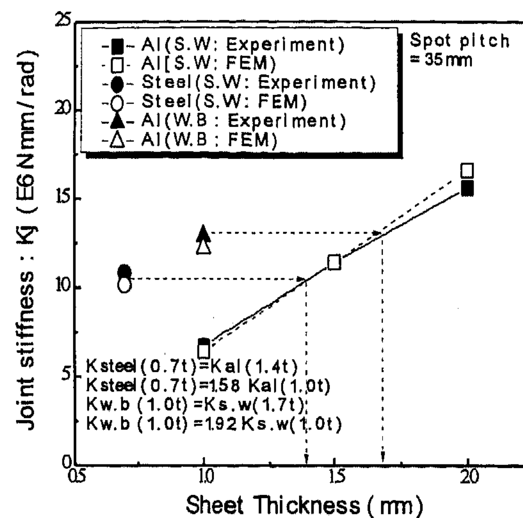
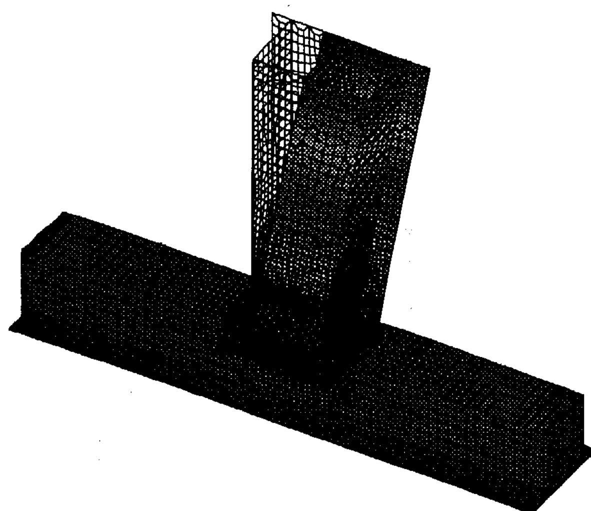
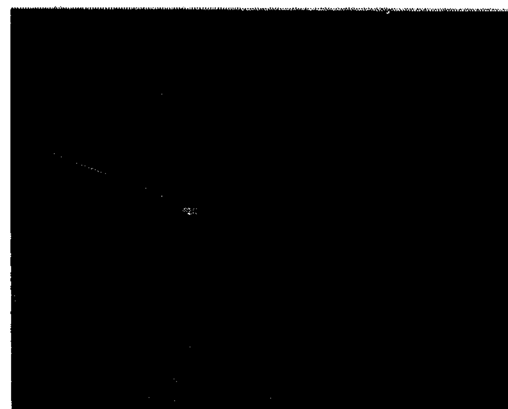


Fig. 10 : Effect of Sheet Thickness on the T-shaped Joint Stiffness



(a) overall deformed shape (FEM)



(b) local deformed shape (Experiment)

Fig. 11 : Deformed Configuration of T-shaped Joint

weld bonded aluminum amount to 62% and 119% of steel, respectively. This indicates that, stiffness of weld bonding is about two times higher than that of spot welding, as in the case of single lap joint. This result is in accordance with FEM analysis.

In spot welded aluminum, the thickness of aluminum with equivalent stiffness to steel(at 0.7 mm gauge) should be 1.4 mm, not 1.0 mm predicted in plate theory as shown in Fig. 10. These results are compatible with the results obtained by analytical and experimental methods.

Table 5 shows the results of analysis and experiment with T-shaped joint. It can be seen that the difference between the two results is within 10%, thus we can confirm the compatibility of FEM analysis.

APPLICATION TO BIW STRUCTURAL JOINTS

The study for simple specimens such as single lap joint and T-shaped joint is unpractical for real BIW, so we need to study real body joint of BIW. In this section, structural stiffness of real body joint

Table 5 : Comparison between FEM and Experiment (in case of spot pitch = 35 mm)

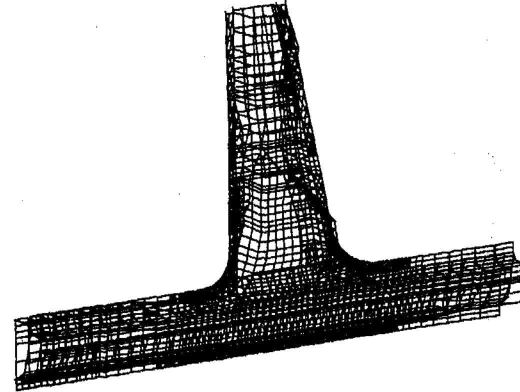
Joining Parameters \ Stiffness			Structural Joint Stiffness ($\times 10^6$ N·mm/rad) FEM (A)	Experiment (B)	difference(%) (A-B)
Al	1.0t	W.B	12.2465	12.9224	-5.23
		S.W	6.4455	6.7380	-4.34
	1.5t	S.W	11.4705	10.8519	+5.70
	2.0t	S.W	16.6530	15.6509	+6.40
Steel	0.7	S.W	10.1887	10.8599	-6.18

and BIW was obtained by FEM analysis as functions of sheet thickness, spot number and joining method. Based on the results, the optimized joint design for aluminum BIW with minimum weight to obtain the equivalent value of stiffness to steel is suggested.

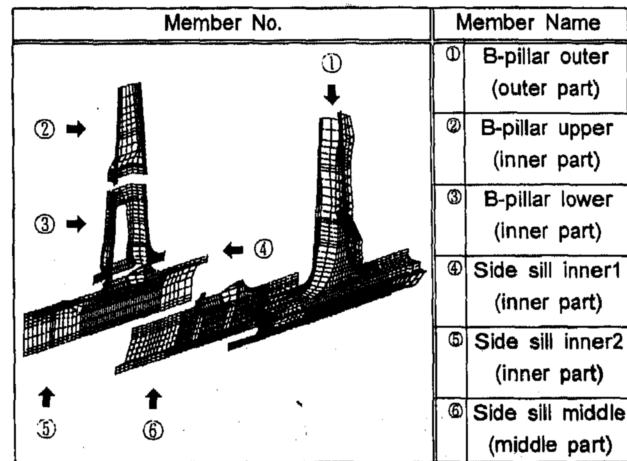
STRUCTURAL JOINT STIFFNESS OF B-PILLAR - As a structural joining part of real automotive body, Center pillar to Rocker Joint(B-pillar Joint) was considered in the present investigation for a real model of T-shaped joint. FE model and the construction of B-pillar joint were shown in Fig.12. As shown in Fig 12(b), B-pillar joint is made of six part members.

For designing sheet thickness, four types of B-pillar joint with various part thickness were analyzed. The first and second type were designed to increase uniformly the thickness of inner part and outer part, respectively. For the third type, the thickness of both parts increased uniformly and for the fourth type, based on theresult of analysis, the thickness of parts which have lower stiffness than other parts increased locally.(Fig. 12 : member no. ①, ③)

Fig. 13 and Fig. 14 show the results of FEM analysis. Joint stiffness of B-pillar joint with various spot pitch is plotted in Fig. 13, in which stiffness equivalent sheet thickness is considered for aluminum and steel sheets.(t_{Al} = 1.4 t_{steel}) It can be seen that the decrease in joint stiffness of spot welded B-pillar and the increase in weld bonded B-pillar are found as increasing spot pitch. The stiffness of spot welded aluminum amounts to 57% of steel at 25 mm spot pitch even though aluminum is 1.4 times thicker than steel. On the contracy, weld bonded aluminum shows higher stiffness than steel as shown in Fig. 13, which is also consistent with the results of T-shaped joint.



(a) FE model



(b) construction of B-pillar joint

Fig. 12 : FE model of B-pillar Joint

The plate theory assumes that the section is constant throughout the whole region and that joining part does not exist. The local deformation and the effects of joining would, therefore, be neglected. A real automotive body, however, has various section and joining part, so the local deformation in the area of low stiffness and the effects of joining are not negligible and sometimes substantial.

In Fig. 14, relative joint stiffness of aluminum to steel sheets is plotted as a function of relative mass to investigate the stiffness equivalent sheet thickness of various aluminum sheets. In this figure, weight reduction can be notified quickly depending on joining method and thickness variation of inner and outer parts.

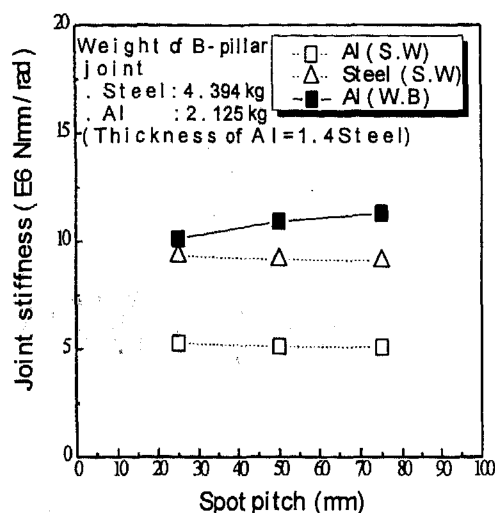


Fig. 13 : Effect of Spot Pitch on the B-pillar Joint Stiffness

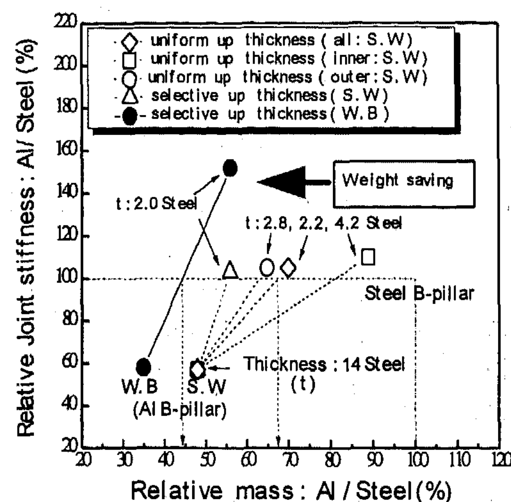


Fig. 14 : Effect of Sheet Thickness on the B-pillar Joint Stiffness (spot pitch = 25 mm)

In spot welding, thickness variation of inner parts(—□—) is the least effective, and thickness variation of local parts(—△—) is the most effective in terms of weight reduction and performance.

In weld bonding, increasing thickness of sheets improves the structural joint stiffness of aluminum B-pillar remarkably, thus, stiffness equivalent sheet thickness of aluminum would be 1.2 times of steel one.

Consequently, local reinforcement of sheets thickness is desirable for both conditions of joining method. And weld bonding would be strongly recommended, if optimum process for the increasing productivity is developed.

STRUCTURAL STIFFNESS OF BIW - Overall stiffness of BIW as a whole, depends on whether it is mainly a frame assembly or a monocoque construction, but even so, in all types of structure, the contribution of the individual joint stiffness to the overall stiffness varies widely. Individual joint stiffness is affected again by process, such as sheet thickness, joining method. It is, therefore, beneficial to investigate the relationship between overall stiffness of BIW and process in such a way as to attain maximum stiffness, via FEM analysis.

For the analysis of BIW stiffness, joint scheme is assured as shown in Fig. 15 to simplify the model.

First of all, FEM analysis was performed in which thickness of aluminum parts increased uniformly up to 1.4 times of thickness of steel. The result of the FE analysis indicated weak joining parts, such as rocker panel to center pillar joint, front pillar to roof rail joint and rocker panel to front pillar joint. After correction of sheet thickness of these weak joining parts, subsequent FE analysis was performed. It is, however, our restriction of this study that thickness should be maintained within 1 mm to 3 mm.

For the summary of the present investigation, Fig. 16 shows the relative bending stiffness of aluminum to steel BIW as a function of relative mass. Fig. 17 shows FE model and simulated deformed shape of aluminum BIW. In Fig. 16, weight reduction can be notified quickly depending on joining method and thickness variation as in the case of B-pillar joint. It is noticeable that for spot welded aluminum with uniform increase of thickness(—△—), the stiffness equivalent to steel is obtained with 70% of the weight of the steel, while for the weld bonded aluminum with selective increase of thickness(—■—), the stiffness equivalent to steel is obtained with 52% of the weight of the steel. Additional weight reduction of 18% is achieved when the spot welded BIW joint is replaced by weld bonding process. For the intensive weight reduction, extension of weld bonded region which mentioned in this investigation will be quite strongly recommended.

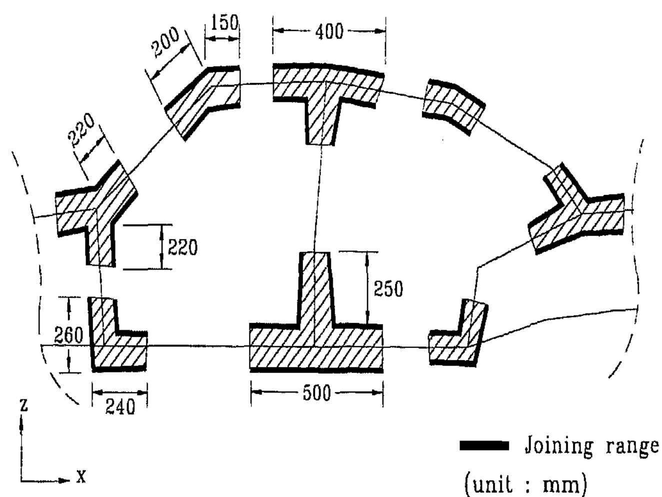


Fig. 15 : Representative of Joining Parts Applied in Full BIW Model

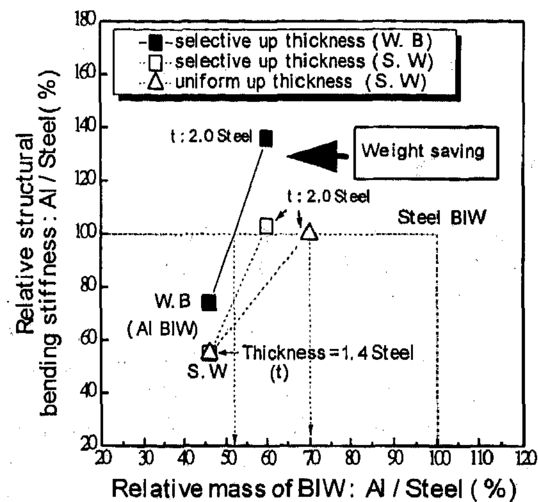
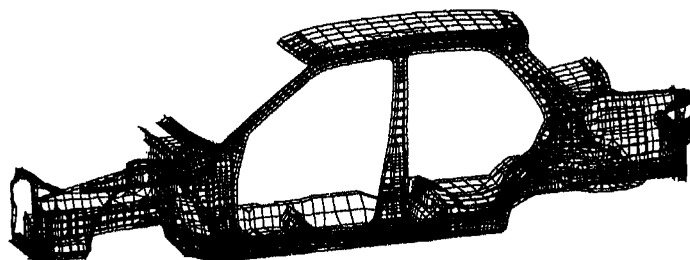
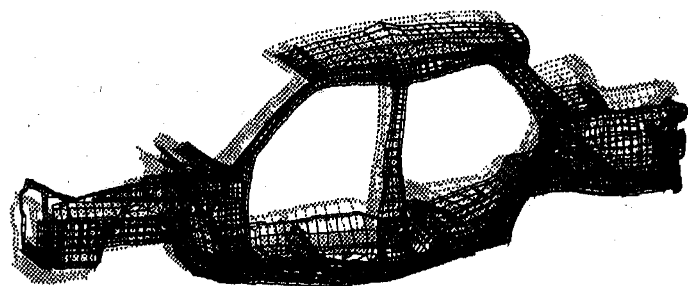


Fig. 16 : Effect of Sheet Thickness on the Structural Bending Stiffness of Aluminum BIW



(a) FE mesh for aluminum BIW



(b) deformed shape of BIW in case of bending

Fig. 17 : FE Model of Aluminum BIW

SUMMARY

1) Structural joint stiffness of aluminum sheet is more affected by sheet thickness than spot pitch, and it is more effectively increased by reinforcing the thickness of an individual sheet which was turn out to be a bit weak, via the analysis.

2) For the joining method, compared to the spot welding, weld bonding has more excellent property in stiffness and strength, especially in fatigue strength, because it enables the uniform stress distribution. The fatigue strength of weld bonded specimens is about seven times higher than that of spot welded ones.

3) Our FEM analysis showed that by optimizing sheet thickness of the part which has lower stiffness and applying weld bonding, the weight of aluminum with equivalent stiffness to steel is reduced to 44% of the weight of steel for B-pillar joint and 52% for BIW.

It is our concluding remarks that local reinforcement of an individual sheet thickness and application of weld bonding would be the most

effective way to improve stiffness, strength and durability of aluminum BIW if optimum process for the increasing productivity is developed.

REFERENCES

1. Toshiaki Sakurai and Yoshinobu Kamada, "Structural Joint Stiffness of Automotive Body", SAE Paper 880550, 1988.
2. D. Nardini and A. Seeds, "Structural Design Considerations for Bonded Aluminum Structured Vehicles", SAE Paper 890716, 1989.
3. Hideo Kitagawa, Yukio Yoshida and Hiroyuki Kanagawa, "A Study of Bending and Torsional Rigidities of Weld-Bonded Structures", JSAE Review Vol. 13, No. 4, pp. 72-76, 1992.
4. P. C. Wang, S. K. Chisholm and G. Banas, "The Role of Failure Mode, Resistance Spot Welded and Adhesive on the Fatigue Behavior of Weld-Bonded Aluminum", WELDING JOURNAL, Feb., pp. 41s-47s, 1995.
5. D. A. Wagner, C. M. Cunningham and M. A. Debolt, "Weld-Bonded Joint Design : Pickup Box Case Study", IBEC '93 Automotive Body Design & Engineering, pp. 123-128, 1993.
6. I. J. McGregor, D. Nardini and A. D. Seeds, "Analysis and Testing of Adhesively Bonded Aluminum Structural Components", SAE Paper. 900795, 1990.
7. D. Nardini, I. J. McGregor and D. J. Meadow, "The Developments of a Joint Design Approach for Aluminum Automotive Structures", SAE Paper 922112, 1992.
8. Y. Y. Kim, H. J. Yim and J. H. Kim, "Reconsideration of the Joint Modelling Techniques : In a Box-Beam T-Joint", SAE Paper 951108, 1995.

970584

Finite Element Prediction of Backlite Molding Squeak Noise

Yuzhao Song, Fu S. Chang, Paul Lipinski, and Mike Paiva
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The backlite molding squeak noise is caused by the stick-slip type of friction between the window molding and the body panel. To predict if the molding would squeak, a finite element analysis technique which uses the nonlinear explicit code LS-DYNA3D has been developed. The three dimensional finite element simulation technique is based on the threshold displacement, velocity spectrum and the relative movement of the window glass with respect to the body panel. Comparisons between FEA analysis and tests are also presented in this paper.

INTRODUCTION

The backlite molding noise is generated by the friction between the window molding and the body shell. The nonstationary and transient nature of the squeak noise makes it very difficult to be predicted.

To reduce and eventually eliminate the squeak noise, a simplest squeak generation mechanism model (plate-molding model) has been studied. In this model (figure 1), a plate oscillates in a specified frequency and amplitude to generate squeak noise. The model is used to calculate the threshold displacement which is defined as the critical plate displacement beyond which the relative movement at the contact point will occur. Secondly, the vibration frequencies of the molding which are excited by the stick-slip type of friction are calculated and compared with the critical radiation frequency of the backlite. Finally, the full vehicle analysis is employed to provide the excitation information for the squeak noise. The analysis procedure is outlined below:

1. Build a plate-molding model for the backlite molding and calculate the spectrum of the noise and the threshold displacement.
2. Calculate the critical radiation frequency of the backlite.
3. If the peaks in the frequency spectrum of the molding velocity are sufficiently offset from the critical frequency of the glass, the design will be accepted.
4. If sufficient offset does not exist, build a full vehicle model and calculate the relative displacement between the glass and the body shell.
5. If the relative displacement is less than the threshold, the probability that the molding will squeak is minimized. The design will be accepted. Otherwise, modify the design and calculate again.

PLATE-MOLDING MODEL ANALYSIS

A typical plate-molding model is shown in figure 1. It includes a 25 to 50 mm long molding strip and a plate which is used to push the molding and to generate the squeak noise. The molding is modeled using solid elements. The material properties of the PVCs, which include the Young's modulus, shear modulus, density and coefficient of friction are from material tests. The aluminum foil between the PVC and the rigid PVC is modeled with shell elements. The molding is fixed at the channel. The total number of elements is approximately 7000-8000. The size of the elements depends on the purpose of the analysis. For response analysis, the smallest element size is around 0.4 mm and for tip pressure analysis the element size around the tip of the molding strip is around 0.06 mm.

In the simulation process, the plate (either rigid or deformable) is first moved up toward the molding strip, so that the contact between the plate and the molding can be created. The plate is then vibrated horizontally to generate the squeak noise. The input of the analysis is the interference distance, amplitude and frequency of the vibration. The output of the analysis is the contact forces and the response of the molding.

1). CONTACT FORCE CALCULATION - Figure 2 and figure 3 show the contact forces in x (tangential) and z (normal) directions. At the first 10 milliseconds, the plate pushes the molding up and then vibrates horizontally with frequency of 15 Hz. After 10 milliseconds, the curve has peaks which represent the evidence that the molding and the plate stick together and plateaus which represent the friction slip. The

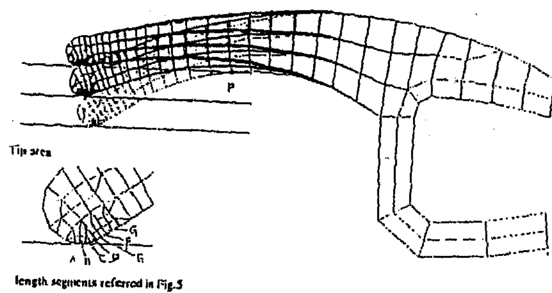


Figure 1: Plate-molding model

comparison of the average forces between FEA and a plate-molding test are shown in table 1. Although there are still differences in extension stroke, the contact forces generally agrees with the tests, which validate the FEA model.

Table 1: Comparison of FEA analysis and bench test

	Freq. (Hz)	Compression Stroke			Extension Stroke		
		Anal	Test	Dif.	Anal	Test	Dif.
Normal force (N)	15	4.0	3.8	5%	1.85	1.25	48%
	10	4.05	4.1	1%	1.95	1.6	22%
Tangential force (N)	15	6.4	6.5	2%	3.28	3.0	9%
	10	6.4	6.5	2%	3.2	3.0	7%

2). LIP PRESSURE ANALYSIS - Since the coefficient of friction is the function of contact pressure, it is suggested to conduct the friction tests at the pressure which is as close as possible to the reality. The FEA analysis shows that for a specified molding design, the contact length on a profile of the molding is around

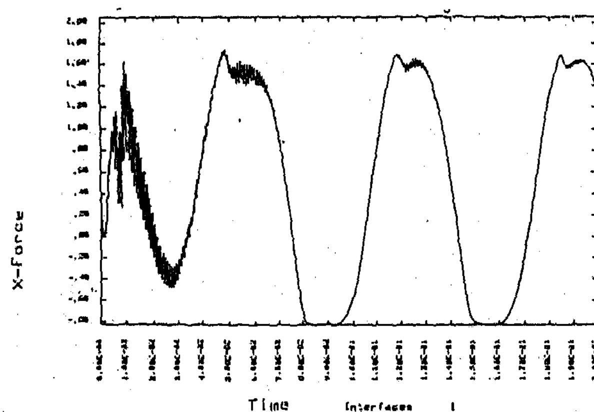


Figure 2: Contact force in x direction

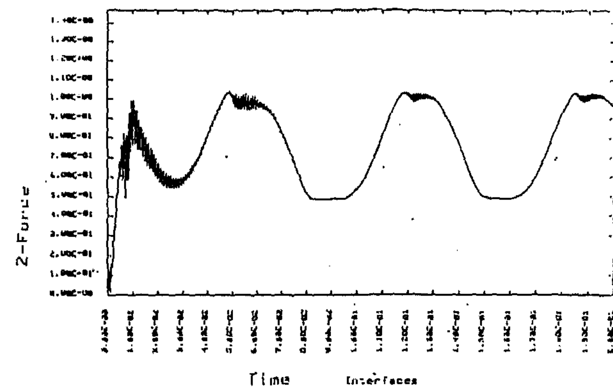


Figure 3: Contact force in z direction

0.2 mm. Figure 4 shows the pressure in the tip area vs. vertical displacement. A specific curve represents a specific length segment. The curve gives a full pressure time history for the segment. The maximum pressure can reach 0.413 MPa if the interference is 2 mm.

3). ANALYSIS OF THE RESPONSE OF THE MOLDING - Figure 5 shows the nodal displacement in x direction, where curve A represents the displacement of the plate which is actually the input and curve B is the nodal displacement at the tip point of the molding. The transition point from 'stick-state' to 'slip-state' are marked with 'T'. The displacement between the compression and extension transition point gives the analytical threshold value of the squeak noise for a particular frequency. If the full amplitude of the excitation, which in reality is the relative displacement of backlite and body shell, is less than the threshold value, there will be no transition from stick stage to slip stage or, in other words, there will be no squeak noise. In the design process, the configuration and the material employed should make the threshold value as large as possible, so that the excitation amplitude could be less

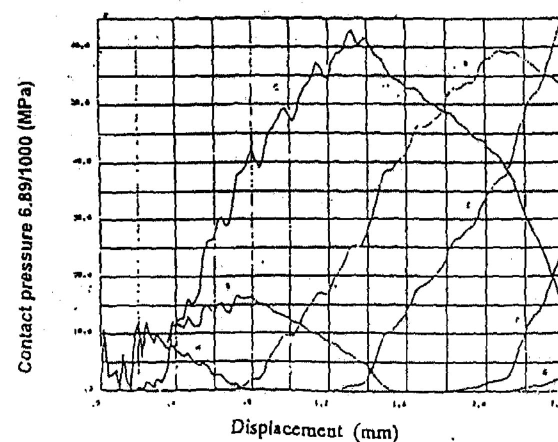


Figure 4: Pressure time history

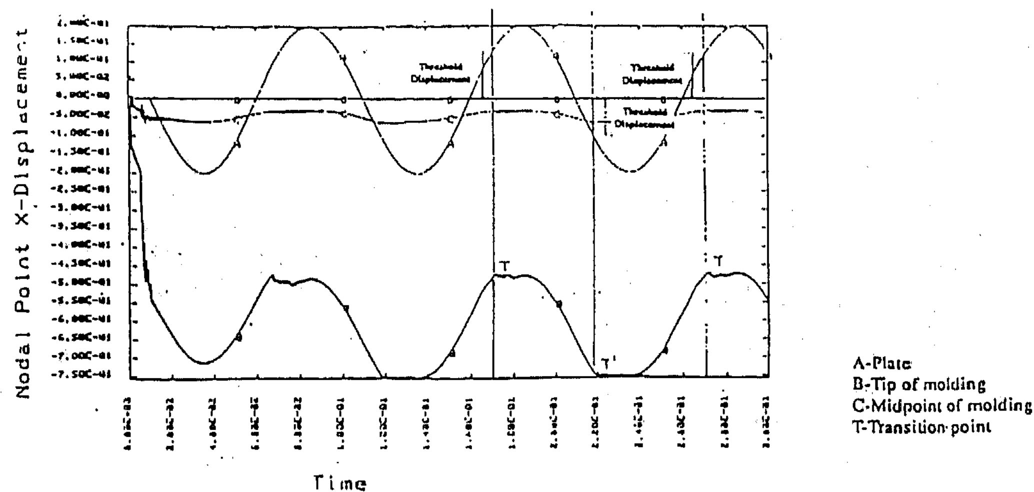


Figure 5: Nodal displacement in x direction

than the threshold. Theoretically, it is a sufficient condition by keeping the excitation amplitude under the threshold to avoid the backlite squeak noise. Table 2 gives a comparison of threshold between the analysis and test. In the table, the threshold of the test is defined as the current excitation amplitude of the plate when the molding begins to squeak. It shows that the analysis agrees with the test well.

Table 2: Comparison of the threshold between the analysis and test

Compression Stroke (mm)			Extension Stroke		
Anal.	Test	Dif.	Anal.	Test	Dif.
0.128	0.120	6.6%	0.11	0.106	3.7%

4). SPECTRUM ANALYSIS OF THE BACKLITE NOISE - In most of the cases, it may not be practical to control the relative displacement under the threshold value. Therefore, an alternative approach has to be found to control the noise. Since the squeak noise which a passenger could hear are mainly emitted from the backlite, reducing the noise intensity to the background noise level could be achieved either by eliminating the source of the noise or by controlling the transmission of the noise. For the backlite, or a curved panel, the noise radiation ratio varies according to the noise frequency. If

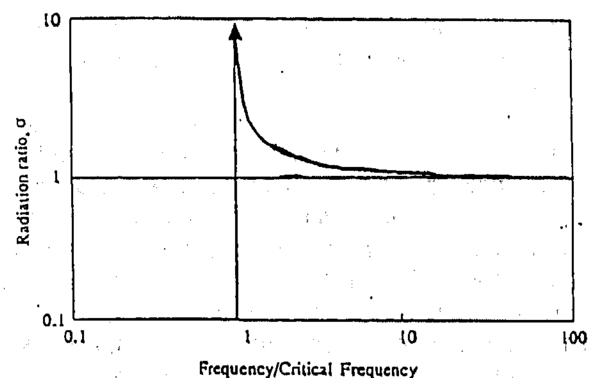


Figure 6: Radiation ratio vs. frequency for an infinite plate

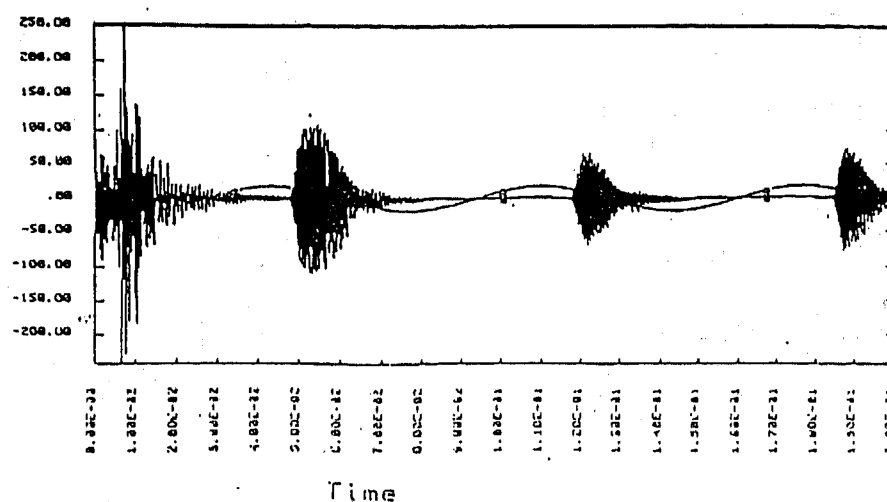


Figure 7: X-velocity of a point at the middle of the molding profile

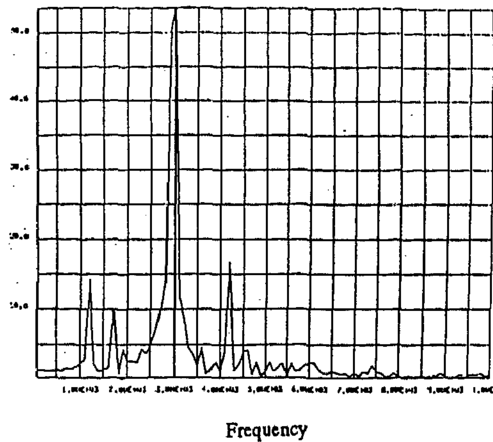


Figure 8: Spectrum of nodal velocity of design proposal 1

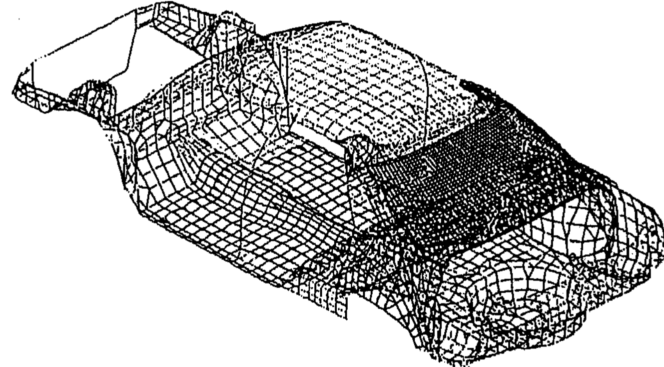


Figure 10: Simplified body in white model

it is an infinite plate, the critical frequency can be determined as:

$$f = \frac{c^2}{1.8t}\sqrt{\frac{\rho(1-\nu^2)}{E}} \qquad (1)$$

where c is the sound speed in air; ρ, E and t are the density, Young's modulus and the gage of the plate, respectively.

Figure 6 shows the relation of radiation ratio vs. frequency for an infinite plate. It is clearly depicted that if the frequency of noise is lower than the critical frequency, there is no sound radiation from the plate to the air. If the noise frequency is equal to the critical frequency, the radiation ratio becomes infinite. When the noise frequency is higher than the critical frequency, the radiation ratio becomes constant. Although the backlite, as a curved panel, does not behave that extremely, it still radiates sound very efficiently in a certain frequency range and relatively inefficiently out of certain frequency range. Considering the size of the backlite and the distance between a passenger and backlite, it is reasonable to first estimate the critical frequency by expression (1).

For a example, suppose the Young's modulus is taken as 61300 MPa, and the gage is 3.8 mm for a specific vehicle, the critical radiation frequency is 3065 Hz. Therefore, if the excited vibration of the molding has a central frequency which is close to 3065 Hz, the noise level will be very high. In other words, if the vibration frequency of the molding is far from 3065 Hz, the noise level could be low enough that it may not be able to be recognized from the background noise.

Figure 7 shows the velocity of the molding at point P (Fig.1). When there is a transition from stick-state to slip-state, there will be excited noises. Figure 8 and figure 9 show the spectrums of the noise for different molding design proposals. From Fig. 8, the spectrum curve of proposal 1 has a very high peak at 3000 Hz which is very close to the critical frequency of 3065 Hz. That means the noise could be radiated to air very

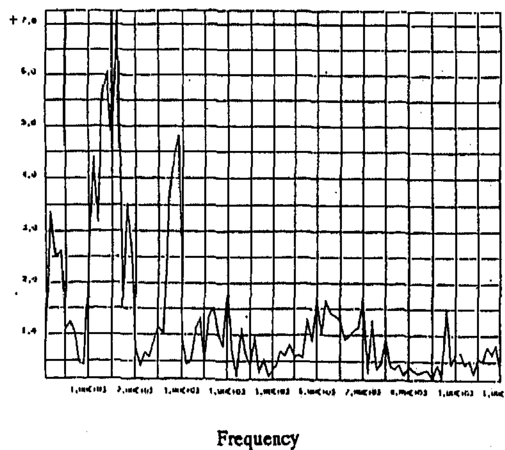


Figure 9: Spectrum of nodal velocity of design proposal 2

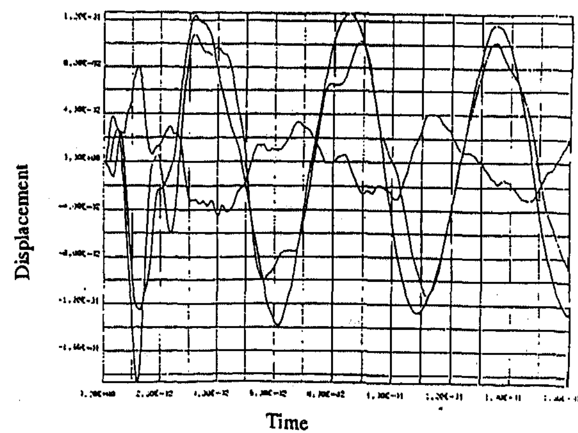


Figure 11: Relative disp. between body shell and backlite

efficiently. Another peak of the spectrum of proposal 1 is around 4200 Hz, which will also be radiated to air, although it is not as efficient as the 3000 Hz peak. Compared to proposal 1, the highest peak of proposal 2 is at 1600 Hz, which is far below the critical frequency. Therefore, the noise level will be greatly reduced since the radiation ratio is very low at that frequency range. There is another peak which is around 2900 Hz. But the amplitude is only 1/8 of proposal 1. Test results support this theory as there is no audible squeak noise for design proposal 2 while there is significant squeak noise for design proposal 1.

FULL VEHICLE ANALYSIS

The simplified Body-in-White Model has been modified to evaluate the body structure influence on the relative glass movement. The model is shown in figure 10. The model contains 10,000 elements. Among them, there are nearly 800 beam elements. To maker the analysis more accurate, the backlite is modeled with fine mesh and the urethane which connects the backlite and the body shell is modeled with solid elements. The input is a twisting displacement at the front shock tower. The amplitude of the input is 2 mm and the frequency is 20 Hz. Figure 11 shows the relative displacement between the body shell and the backlite at the left hand upper corner. From figure 11, the maximum relative displacement between the backlite and body shell is 0.12 mm. This data, together with the threshold displacement, can then be used to determine whether the molding will squeak or not.

CONCLUSION

1). A finite element analysis technique can be used to simulate and predict the backlite molding squeak. The method is based on the threshold displacement prediction, the velocity spectrum and the relative movement of the window glass to the body shell.

2). The explicit nonlinear dynamic code LS-DYNA3D and the software created in house are suitable for the study of the backlite squeak noise.

3). The analysis of the plate-molding model agrees with the test data for both the contact forces and the threshold displacements.

REFERENCES

1. E. Hughes and T. Judek, "An Integrated Experimental and Analytical Approach in Modeling Squeak", Ortech Report, 1993.

2. J. T. Oden and J. A. C. Martins, "Models and Computational Methods for Dynamic Friction Phenomena", Comp. Math. Appl. Mech. and Eng., pp 527-639,1985.

3. M. P. Norton, "Fundamentals of Noise and Vibration analysis for Engineers", Cambridge University Press, 1989.

4. A. D. Pierce, "Acoustics: An Introduction to Its Physical Principles and Applications", McGraw-Hill Co., 1981.

970585

Design and Control of Adaptive Structures Using the Finite Element Method

Baruch Pletner, Moshe Idan, Haim Abramovich, and Tanchum Weller
Technion - I.I.T.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Current work develops and presents a clear methodology for all stages of adaptive structure design using a commercially available FE program. A procedure for replacing the induced strain actuation of the piezo-elements by mechanical forces is developed and validated numerically using the ANSYS^R coupled field piezoelectric 3-D solid elements, as well as experimentally. The use of reduced FE analysis as a spatial discretization tool for controller design is addressed and shown to be advantageous in many cases. Balanced order reduction schemes are applied to the discrete structural model in state space in order to ensure controllability and observability, as well as to maintain a low number of states for the controller design and implementation. Linear quadratic (LQ) and $\mathcal{H}_\infty$ controller designs are presented and implemented using computer simulations of the controlled structure. The methodology is applied to a simply supported beam as a bench-mark case, and is shown to effectively control the beam vibrations.

INTRODUCTION

A common form of intelligent or adaptive structures is a thin shell equipped with piezoelectric laminae. These laminae are made of piezoelectric materials such as Lead Zirconate Titanate (PZT), and are capable of transducing electric fields into mechanical strains, and mechanical strains into electric charges. These 'active' laminae are used to actuate the shell by inducing strains in the non-piezoelectric 'passive' laminae, and to sense deflections in the shell by measuring the local strain fields. The active laminae can be continuous over the entire domain of the structure, or discontinuous as in the case of piezoceramic 'patches'. Sufficiently accurate models for induced strain actuation are essential for the design of intelligent structures and their use for vibration suppression and steering.

Finite Elements Analysis (FEA) has been the method of choice for the analysis of structures which do not yield to closed-form solutions for some time now. It is widely used in the automotive industry, among others. In the context of smart structures, however, its role must be reassessed. The applicability of FEA to the adaptive structure design process should be investigated with respect to its capability of modelling the piezoelectric effects, and through them, piezoelectric actuation and sensing. In addition, the method's possible use as a structural discretization scheme for controller design should be addressed. In this work the piezoelectric modelling capability of the ANSYS^R commercial FE program was investigated and shown to be in excellent agreement with both analytical and experimental results.

In recent years the developers of commercial Finite Element (FE) software packages have anticipated demand and developed coupled-field piezoelectric elements. In general, these elements include electrical parameters such as voltage, as an additional degree of freedom at each node. The coupled-field elements can be incorporated into the FE model of the structure, yielding the structural response to the application of electric fields as part of the general output. A summary of these elements and their capabilities was given by Lin *et al.* [1]. The ANSYS^R FE program [2], which was used in this work posesses two coupled-field piezoelectric elements, an eight-node three dimensional solid, and a four node two dimensional solid.

While these elements can be very useful for the analysis of intelligent structures, they make the model much more expensive numerically [3]. The increased numeric cost stems from the additional degrees of freedom in the model (volts), as well as the lack of piezoelectric beam and shell elements, which can be very useful for the modeling of planar structures. This can be avoided by replacing the piezoelectric elements with the appropriate mechanical loads found using the extended pin-force model as presented in this work.

Discrete models of continuous structures tend to be very large, making the implementation and simu-

lation of the controlled structure extremely difficult [4]. This difficulty can be circumvented by bringing the system to a balanced representation and truncating the lightly controllable/observable states [5]. As a result of balanced truncation the structure is presented in state-space as a stable controllable and observable linear system of a relatively low order. This permits the application of state-space controller design tools to the controller and state estimator design [6].

The advantage of the proposed methodology is in its generality. It is applicable to any structure that can be modeled with FE, and is linearly elastic. In addition, it takes advantage of the state-of-the-art capabilities of commercial FE packages, and permits the application of the latest developments in state-space MIMO controller design techniques to continuous flexible structures.

METHODOLOGY OF ADAPTIVE STRUCTURE DESIGN WITH FEM

The objective of this work is to develop a methodology for the design and implementation of an active controller for the attenuation of vibrations in thin shells of general geometric and material properties. Since in general the vibrational modes cannot be predicted analytically, modal analysis is not applicable to this problem.

The design methodology therefore uses finite elements programs with coupled-field capabilities to discretize the structure and to identify undesirable modes, as well as to model piezoelectric sensing and excitation. Once the undesirable modes are identified, actuators and sensors are placed in a way that will ensure that these modes are observable and controllable. When sensors/actuators mass and stiffness contributions become significant, they should be included in the FEM.

Master degrees of freedom are then chosen and included in the model. The reduced model is run to obtain the mass and stiffness matrices and load vectors, which are processed to construct the matrices of the state-space model. A cost function quantifying the energy content of the vibration is found as a quadratic form of the system states, and a LQ controller is designed.

PIEZOELECTRIC ACTUATION AND SENSING

In this section existing models for piezoelectric actuation and sensing are briefly presented. The general results are modified to provide the scaling factors for converting piezoelectric actuation to mechanical loads.

PIEZOELECTRIC ACTUATION — When piezoelectric actuators such as PZT patches are used to control flexible structures, we need to know the scaling factor between the voltages applied to the actuators and the forces and moments which are applied by the actuators to the structure. This factor can be found using the extended pin-force models as reported in the literature, as well as coupled-field piezoelectric finite elements and experimental techniques. Here we present results for piezoelectric actuation of shells equipped with pizoceramic patches bonded to either one or both of the shell's surfaces.

The induced-strain load vector for a shell equipped with a symmetrically bonded actuator pair assuming that the poling axes are aligned with the coordinate system (see Figure 1) is [7]

$$\begin{Bmatrix} N_{11} \\ N_{22} \\ N_{12} \\ M_{11} \\ M_{22} \\ M_{12} \end{Bmatrix} = \frac{E_a}{1-\nu_a} \begin{Bmatrix} d_{31}V^+ \\ d_{32}V^+ \\ 0 \\ \frac{h_s}{2}\left(\frac{1}{T}+1\right) d_{31}V^- \\ \frac{h_s}{2}\left(\frac{1}{T}+1\right) d_{32}V^- \\ 0 \end{Bmatrix} \qquad (1)$$

where V^+ and V^- are the sum and the difference of the voltages applied to the upper and lower actuators, respectively. In case of only one actuator we get $V^- = V^+ = V$, and in case of two $V^- = 2V$. The thickness ratio between the actuators and the structure is defined as $T \triangleq h_s/h_a$.

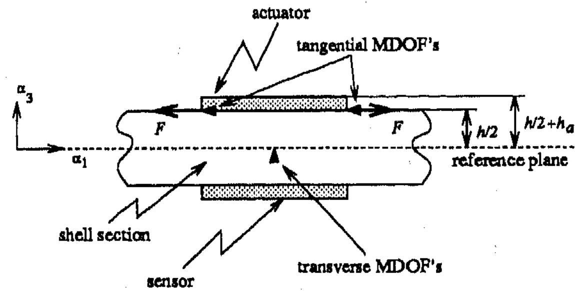


Figure 1: Section of actuated structure

In most cases we are interested only in the bending moments, since they can actuate transverese displacements. The moments are generated by the line-forces F, which act at the boundaries of the piezoelectric actuator. We can express these forces per unit width as

$$F = \frac{M_{ii}}{h_s} = \frac{E_a d_{3i}}{2(1-\nu_a)}\left(\frac{1}{T}+1\right) V \triangleq \xi_i V,\ i=1,2 \qquad (2)$$

where ξ_i is the volt-to-force scaling factor in the i-th direction. Alternatively, the scaling factors can be expressed through the induced curvatures as

$$\xi_i = \frac{M_{ii}}{h_s V} = \frac{EI\kappa_{ii}}{h_s V},\ \ i=1,2 \qquad (3)$$

where the curvatures are given by [3, 8]

$$\kappa_i = \frac{6\left(\frac{1}{T}+1\right)\left(\frac{1}{h_s}\right)}{\psi_\nu \frac{1+\nu_s}{1+\nu_a} + \tau}\left(\frac{d_{3i}}{h_a}\right) V \qquad (4)$$

The stiffness ratio ψ_ν between the actuators and the structure is defined by

$$\psi_\nu \triangleq T\frac{E_S}{E_a} \qquad (5)$$

and the non-dimensional coefficient τ as

$$\tau \triangleq 6 + \frac{12}{T} + \frac{8}{T^2} \quad (6)$$

The scaling factors can be easily incorporated into the formulation of the adaptive structure as a controlled system.

PIEZOELECTRIC SENSING — As shown by Pletner and Abramovich [3], the voltage measured across a piezoelectric sensor is proportional to the difference in the slope of the neutral axis deflection function $w(\alpha_1)$ evaluated at the sensor edges, as follows

$$\tilde{V}_l = \frac{h}{C_l}\int_{\alpha_1^l}^{\alpha_1^r} e_{31}^l \kappa(\alpha_1)dx = -\frac{he_{31}^l}{C_l}\int_{\alpha_1^l}^{\alpha_1^r} w''(\alpha_1)dx = -\eta_l\left[w'(\alpha_1^r) - w'(\alpha_1^l)\right] \quad (7)$$

where h is half the shell thickness, e_{31}^l is a piezoelectric constitutive parameter, and C_l is the l-th sensor capacitance. This model of piezoelectric sensing can be discretized and realized in state space, as is shown in the following sections.

CONTROL ORIENTED FINITE ELEMENT MODELING

A well established and validated technique for the discretization of continuous flexible structures is the Finite Element Method (FEM). In this method the structural continuum is divided (meshed) into a finite number of elements. Each element posesses a finite number of nodes, for each node a finite number of degrees of freedom (DOF's) are defined by the element type. These DOF's are most often mechanical displacements and rotations, but can also be thermal and electrical properties. In this manner a continuous structure is represented by a finite number of DOF's. The forces acting on the structure are applied at the nodes in the directions of the DOF's that are relevant for the node in question. For instance, at a node posessing the (U_x, U_y) translational DOF's, only forces in the (x, y) directions may be applied. The FEM uses energetic principles to assemble the global mass and stiffness matrices for the structure, and constructs the load vectors for each load case. The general form of the discretized equations of motion for the shell can be stated as

$$\boldsymbol{M}\ddot{\vec{d}} + \boldsymbol{C}_d\dot{\vec{d}} + \boldsymbol{K}\vec{d} = \vec{F}f(t) \quad (8)$$

where $\boldsymbol{M}, \boldsymbol{C}_d, \boldsymbol{K}$ are the global mass, damping and stiffness matrices, respectively, and $\vec{d}$ is the nodal displacement vector. $\vec{F}$ is the nodal load vector, and $f(t)$ is a time-dependent load.

REDUCED FINITE ELEMENT MODELS

In many cases a subset of all the DOF's in the FE mesh are of more interset to the structural analyst. For example, the vibrational analysis of thin shells and plates often requires the solution for the transverse DOF's located on the reference plane of the shell. In these cases the size of the model can be reduced by defining 'master' DOF's (MDOF's) along the reference plane, and solving the structure for these MDOF's only. This reduced analysis is a feature of many commercial FE programs, and is particularly useful if the FE representation of the structure is to be used in control applications. This is due to the fact that even 'modestly' sized FE models are far too large for many controller design tools. The number and location of the MDOF's should be chosen in such a way that the solutions obtained by the full and reduced models for a pre-specified number of vibrational modes coincide within a given tolerance.

Reduced FE Models of Piezolaminated Shells — A reduced FE model of a piezolaminated thin shell consists of three types of MDOF's. The transverse motion of the shell is described by MDOF's denoted by MDOF_3's. These MDOF's are located on the reference plane of the shell, and are aligned in the transverse direction α_3. Since the piezo-sensors are scaled to yield the structural deflection at their centers, the nodes located at the centers of all sensors should be defined as MDOF_3's.

The MDOF's needed for the application of piezoelectric loads are located on the upper and lower surfaces of the shell, and are aligned in the tangential directions (α_1, α_2). These MDOF's are denoted as MDOF_a's. Finally, the MDOF's used for the application of the external forces in the FE model are located on the reference plane of the shell, and aligned in the transverse direction (α_3). These MDOF's are denoted as MDOF_e's.

The shell is equipped with a number of sensors and actuators, and may be excited by a number of external forces. It should be noted that the MDOF_3's located at the sensors' centers are a subset of the total number of MDOF_3's, so that $N_s \leq N_3$. In addition, the excitation forces, which are usually normal to the shell, can be applied at the locations of the MDOF_3's, so that $N_e \leq N_3$.

STATE SPACE MODELOING OF ADAPTIVE STRUCTURES — The equations of motion for a piezolaminated shell excited by N_e forces, and controlled by N_a piezoceramic actuators are

$$\boldsymbol{M}\ddot{\vec{d}} + \boldsymbol{C}_d\dot{\vec{d}} + \boldsymbol{K}\vec{d} = \sum_{i=1}^{N_e}\vec{F}_e^i f^i(t) + \sum_{j=1}^{N_a}\vec{F}_a^j V^j(t) \quad (9)$$

where $\vec{F}_e^i$ is the load vector given by the FE model for the i-th excitation force, and $\vec{F}_a^j$ is the load vector for the j-th actuator. The excitation forces and controlling voltages are given by $f^i(t)$ and $V^j(t)$, respectively. To obtain a state-space representation of Equation (9) a state vector is defined as

$$\vec{x} \triangleq \left\{ \begin{array}{c} \vec{d} \\ \dot{\vec{d}} \end{array} \right\} \quad (10)$$

The equations of motion in state-space are then

$$\dot{\vec{x}} = A\vec{x} + B\left\{\begin{array}{c}\vec{f}\\ \vec{V}\end{array}\right\} \qquad (11)$$

$$\vec{y} = C\vec{x} \qquad (12)$$

where $\vec{y}$ is the measurements vector, and the vectors $\vec{f}$ and $\vec{V}$ contain the excitation forces and control voltages, respectively. The dynamics matrix A is given by

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}K & -M^{-1}C_d \end{array}\right] \qquad (13)$$

and the input matrix B is constructed as

$$B = \left[\begin{array}{c} 0 \\ \hline M^{-1}\left[\; \vec{F}_e^1 \,\middle|\, \cdots \,\middle|\, \vec{F}_e^{N_e} \,\middle|\, \vec{F}_a^1 \,\middle|\, \cdots \,\middle|\, \vec{F}_a^{N_a} \;\right] \end{array}\right] \qquad (14)$$

Assuming that the $MDOF_3$ located at the left edge of the l-th sensor is numbered k_l^l, and the one at the right edge is numbered k_r^l, Equation (7) can be approximated by

$$\begin{aligned}\tilde{V}_l = \\ -\frac{\eta_l}{\Delta_{\alpha_1}}\left[\left(x_{k_r^l+1} - x_{k_r^l-1}\right) - \left(x_{k_l^l+1} - x_{k_l^l-1}\right)\right] = \\ \frac{\eta_l}{\Delta_{\alpha_1}}\left[\left(x_{k_l^l+1} + x_{k_r^l-1}\right) - \left(x_{k_l^l-1} + x_{k_r^l+1}\right)\right]\end{aligned} \qquad (15)$$

where Δ_{α_1} is the distance between k_r^l and k_l^l, which may be assumed to be constant for all l without loss of generality. The measurement matrix C is constructed as

$$C = [\;\tilde{C} \mid 0\;] \qquad (16)$$

where the (l, m) components of $\tilde{C}$, $\tilde{C}_{lm}$ are given by

$$\tilde{C}_{lm} = \left\{\begin{array}{ll} \eta_l/\Delta_{\alpha_1}, & m = k_l^l + 1, \quad m = k_r^l - 1 \\ -\eta_l/\Delta_{\alpha_1}, & m = k_l^l - 1, \quad m = k_r^l + 1 \\ 0, & \text{otherwise} \end{array}\right. \qquad (17)$$

ACTIVE CONTROL OF STRUCTURAL VIBRATIONS

MODEL REDUCTION — The state space realization of the discrete structural model as given by the FEM is in all likelihood not minimal, since a number of MDOF's are aligned in the tangential directions. There are, however, no sensors that can sense displacements in these directions. This will cause the tangential or axial modes to be unobservable. In addition, in order to get a reliable approximation of the radiated acoustic power, an accurate representation of the spatial velocity field is needed, which in turn requires a large number of $MDOF_3$'s. However, due to the relatively small number of actuators, some of the discretized system modes will be uncontrollable.

The application of many controller design techniques to non-minimal systems can be numerically difficult. In addition, for reasons of computational efficiency and practical implementation considerations, systems with a low number of states are desirable. Low order systems can be obtained by using the balanced truncation technique. In this technique the system is transformed into coordinates where the controllability and observability gramians are diagonal and equal. This can be interpreted as if the states in the new coordinates are equally controllable and observable. Following the transformation, the states with Henkel singular values lower than a pre-specified tolerance ϵ are truncated. This may lead to a very substantial reduction in the number of states. Care has to be taken, however, that the lightly observable and controllable states do not include states we wish to control, since these states will be truncated, and will not be controlled. Careful placing of the actuators and sensors should overcome this problem. In general, sensors and actuators should be placed in areas of maximal bending moments, however, the optimal placement and sizing of sensors and actuators is the subject of ongoing research.

The Henkel singular values of the balanced system arranged in descending order are partitioned such that

$$\Sigma = \left[\begin{array}{cc} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{array}\right] \qquad (18)$$

where

$$\begin{array}{rcl} \Sigma_1 & = & \text{diag}\{\sigma_i\}, \; i = 1, \cdots, N_r \\ \Sigma_2 & = & \text{diag}\{\sigma_j\}, \; j = N_r + 1, \cdots, N_{states} \end{array} \qquad (19)$$

and $\sigma_j \leq \epsilon$. The balanced system is partitioned as follows

$$G_B(s) = \left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array}\right]; \dim(A_{11}) = N_r \times N_r \qquad (20)$$

where the following transfer matrix notation is used

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \triangleq C\,(sI - A)^{-1}\,B + D \qquad (21)$$

Since the original system is stable, the truncated system

$$G_T(s) = \left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array}\right] \qquad (22)$$

is balanced, asymptotically stable, and approximates the original system with a finite (known) accuracy [4, 5].

CONTROLLER DESIGN

The choice of a controller naturally depends on the control objective. When the control objective is to regulate the state response of an unforced structure to initial conditions or impulse excitation to zero, while minimizing the total vibrational energy in the structure, linear quadratic (LQ) controllers are used. In cases of persistent external disturbance to the structure, a better approach is $\mathcal{H}_\infty$ control, which minimizes the infinity norm of the output/input ratio for power bounded inputs.

Linear Quadratic (LQ) CONTROL — The LQ controller is designed to minimize a quadratic cost function

$$J = \int_0^\infty \left[\vec{x}(t)^T \boldsymbol{Q}\vec{x}(t)dt + \vec{u}(t)^T \boldsymbol{R}\vec{u}(t)\right] dt \quad (23)$$

where $\boldsymbol{R}$ is an input weighting matrix introduced to penalize the use of large control inputs. It is a well known result that the solution of Equation (23) is given by a linear constant gain feedback controller [9]

$$\vec{u}(t) = \boldsymbol{F}\vec{x}(t) \quad (24)$$

where the gain matrix $\boldsymbol{F} = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{S}_c$, and $\boldsymbol{S}_c$ is the solution of the controller Ricatti equation [4,9].

Linear quadratic regulation (LQR) assumes all states are measured. Since this is not the case here, a state estimator is needed. In this work a Luenberger observer is incorporated, where the observer gain matrix $\boldsymbol{L}$ is determined by pole placement [10]. This approach is valid when measurement and process noises are neglected. In the presence of noise, a stochastic estimator such as the Kalman filter should be employed.

Having placed the sensors and actuators so that the modes we are most interested in controlling are also the most observable/controllable, the state weighting matrix $\boldsymbol{Q}$ is chosen so that each state is penalized reciprocally to its controllability as [4]

$$\boldsymbol{Q} = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Sigma}^{-1} \quad (25)$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with the Henkel singular values of the truncted system in Equation (22). The closed-loop system architecture is shown in Figure 2.

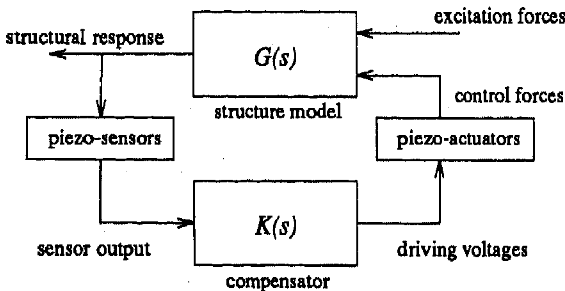


Figure 2: Schematic representation of controlled structure

$\mathcal{H}_\infty$ CONTROL — Steady state vibration can be stated as a problem of rejecting persistent disturbances in steady state. $\mathcal{H}_2$ control methods such as LQR are not ideally suited for this problem, since they minimize the "infinite horizon" $\mathcal{L}_2$ norm of the system response

$$||\vec{y}||_2 = \left[\int_0^\infty \vec{y}^T\vec{y}dt\right]^{\frac{1}{2}}$$

For persistently disturbed systems vibrating in steady state it is obvious that the $\mathcal{L}_2$ does not exist. A different approach is therefore needed to directly address the problem of persistent disturbance rejection.

Consider a linear dynamic system described by a generalized closed-loop transfer matrix $\boldsymbol{T}(s)$ as

$$\tilde{\vec{e}}(s) = \boldsymbol{T}(s)\tilde{\vec{d}}(s) \quad (26)$$

where $\tilde{\vec{d}}(s)$ and $\tilde{\vec{e}}(s)$ are power bounded scaled inputs and outputs, respectively as they appear in Figure 3. $\boldsymbol{P}(s)$ is the transfer matrix of the augmented open-loop plant, as explained in the following paragraphs. In this case the $\mathcal{H}_\infty$ norm of the closed-loop transfer matrix $\boldsymbol{T}(\jmath\omega)$ is defined as

$$||\boldsymbol{T}||_\infty = \sup_{\omega\in\boldsymbol{R}}\bar{\sigma}\left[\boldsymbol{T}(\jmath\omega)\right] \quad (27)$$

where $\bar{\sigma}(\jmath\omega)$ are the maximum singular values of $\boldsymbol{T}$. The $\mathcal{H}_\infty$ problem for power bounded input problems can be stated as: Find an internally stabilizing dynamic controller $\boldsymbol{K}(s)$, such that $||\boldsymbol{T}||_\infty \leq \gamma$, where γ = const [11]. The inputs to the system are assumed to be random zero mean processes, power bounded and scaled so that their power equals unity. Such scalng is denoted by $\tilde{(\cdot)}$. When such scaling is used, γ can without loss of generality be set to one.

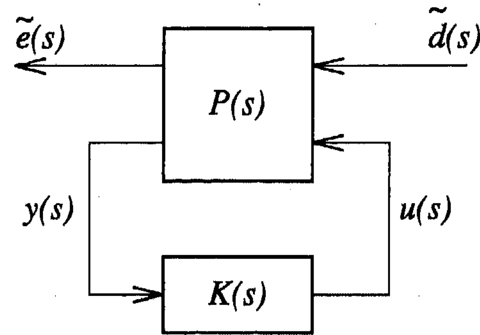


Figure 3: Closed-loop interconnection of controlled structure

In ASAC problems the external inputs $\vec{d}(s)$ may consist of mechanical or acoustic surface loads, as well as sensor or measurement noise. The output $\vec{e}(s)$ from the system is usually the displacement and/or velocity field of the vibrating structure.

Controller Design Setup — The control design objective is to obtain a constant gain controller of the sound radiation from a flexible plate. The interconnection model used for the controller design is presented in Figure 4. It incorporates the model of the plate dynamics (as given by finite element discretization), piezoceramic actuator dynamics, and various frequency weighting functions.

As can be seen, the inputs vector $\vec{d}$ to the augmented plant $\boldsymbol{P}(s)$ consists of two external and one control inputs: the external forces $f(s)$, measurement noise $n(s)$, and control inputs $u(s)$. The outputs vector of the closed-loop system $\vec{e}$ has two weighted components: controls $\vec{u}(s)$ and displacement field $d(s)$. The measurements $\vec{y}(s)$ are used as input to the compensator $\boldsymbol{K}(s)$.

The external force weighting parameter $\boldsymbol{W}_d(s)$ is a bandpass filter admitting only the frequencies of the

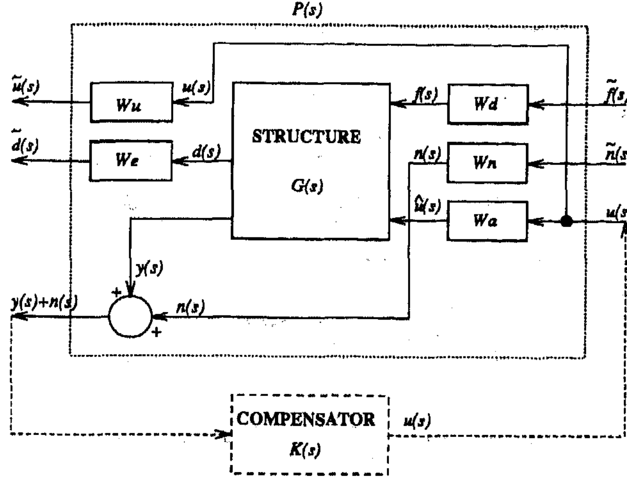


Figure 4: Open-loop architecture of generalized plant

external forces, and scaling them by their respective magnitudes. In the case of single frequency excitation $f(t) = F_a e^{j\omega_a t}$, a notch filter is used, as follows

$$W_d = \left(\frac{F_a}{2\zeta_a\sqrt{1-\zeta_a^2}}\right)\frac{\omega_a s}{s^2 + 2\zeta_a\omega_a s + \omega_a^2} \tag{28}$$

The measurement noise of the piezoelectric sensors is assumed to have a maximum amplitude of $\nu_{max} =$ 0.01[Volt], $\forall\omega$, and so the noise weighting factor is given by $W_n = \nu_{max} \boldsymbol{I}_{N_s}$. Piezoceramic actuators dynamic response can be obtained from the manufacturer, and is usually represented by a highpass filter, since they have a very limited DC capability. In this work, however no DC or low frequency actuation is required, and all performance requirements are limited to frequencies well within the actuators operating range. No generality is lost therefore by taking $W_a = \boldsymbol{I}_{N_a}$.

The voltage input to piezoceramic actuators is limited by the maximum allowable driving voltage, which may decrease with frequency. In some cases this voltage can be limited by the available power source, and not by the actuators themselves. In any case the controller design must take into account the limits on the control inputs. This is achieved by setting the weight W_u to $1/V_{max}\boldsymbol{I}_{N_a}$, where $V_{max} =$ 100[Volt] is the maximum allowable driving voltage.

The performance weighting parameter W_e determines the maximum displacements of the structure which can be tolerated after the application of control. It it thus set to $W_e = 1/d_{max}\boldsymbol{I}_{N_3}$, where d_{max} is the maximum allowable displacement.

EXAMPLES

Two examples are presented. In order to demonstrate the capapabilities of the coupled-field piezoelectric elements and the equivalent mechanical loading models proposed in this paper, a cantilever plate equipped with one actuator pair was used.

Demonstration of the controller design methodology was performed using a slender simply supported beam excited by one point force and controlled by three PZT actuator pairs. The beam was selected as a benchmark case because its vibrational modes are well known, and the results can be evaluated intuitively. The methodology is general and can be applied to any linear thin-walled structure.

THIN CANTILEVER PLATE — The experimental investigation was performed using a thin cantilever steel plate actuated by a single actuator pair as shown in Figure 5. The purpose of the experiment was to validate the coupled-field FE modeling technique, as well as the extended pin-force model presented previously. This was done by determining the first fundamental frequency of the plate, exciting it on-resonance using the piezoelectric actuators, and obtaining the steady-state response. As a measure of this response, the strain at the root of the plate was measured using a strain gauge couple at the clamp (see Figure 5). This strain was then compared to the strains found by the coupled-field FE model of the plate using the ANSYS[R] harmonic analysis capability. This was done both for electrical loads, and for the equivalent mechanical loads, found using the pin-force model. The process was repeated for several amplitudes of excitation, and the results are shown in Figure 6. Excellent agreement was found between the experimental results (denoted by (*)), and the results yielded by the coupled-field FEM (demoted by (o)). The equivalent mechanical loading used with a purely structural FEM (denoted by (+)), yielded an upper limit for the strain values.

SIMPLY SUPPORTED SLENDER BEAM — In this example we used an aluminum beam with three colocated sensor-actuator pairs, as shown in Figure 7. The beam was meshed into 480 SOLID42 two dimensional solid elements with 657 active DOF's, and a total of 33 MDOF's. The total number of states in this case was 66, and it was truncated to 6, so that the first three modes were represented in the truncated model. The truncation can be justified by looking at the natural frequencies and the frequency response of the beam. The frequencies yielded by the full FE (unreduced) model, the reduced and truncated model, and the idealized analytical expression which neglects the actuators and sensors and uses the Euler-Bernoulli model, i.e.

$$\omega_n = \left(\frac{n\pi}{L}\right)^2\sqrt{\frac{EI}{\bar{\rho}A}} \tag{29}$$

are grouped together in Table 1. As can be seen, the sensors/actuators mass and stiffness affect the natural frequencies of the beam. The truncated frequencies, however, are quite close to the values yielded by the full analysis. The actuators and sensors were placed so that their centers coincided with the maximum displacements in the third vibrational mode. Thus this mode was designed to be the most controllable/observable. Since the state weighting matrix $\boldsymbol{Q}$ was designed to favor the most controllable modes,

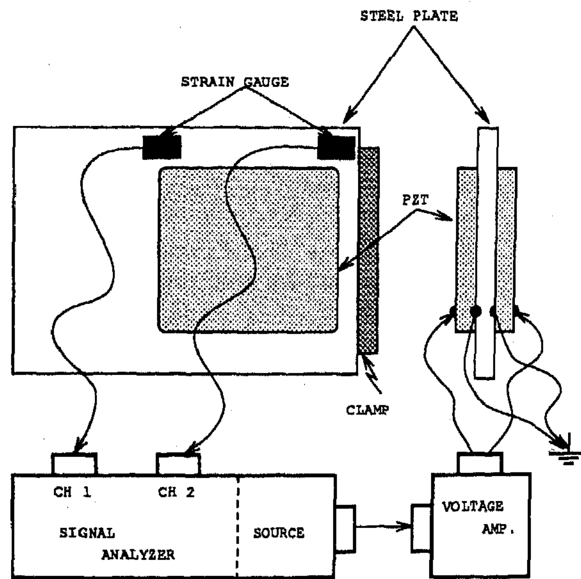


Figure 5: Experimental set-up

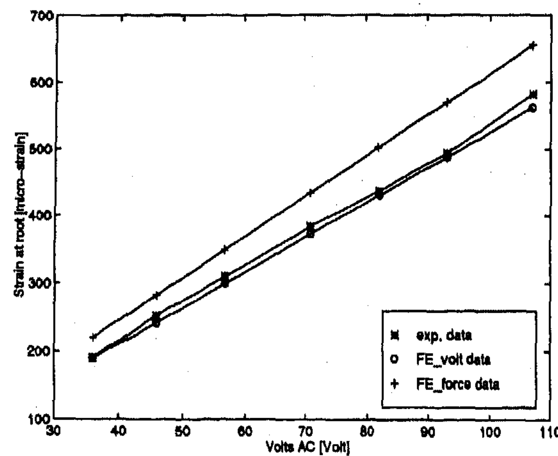


Figure 6: Strain at root of plate @ 40 [Hz]

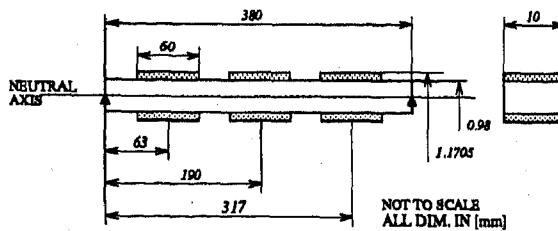


Figure 7: Simply supported beam with three actuator pairs

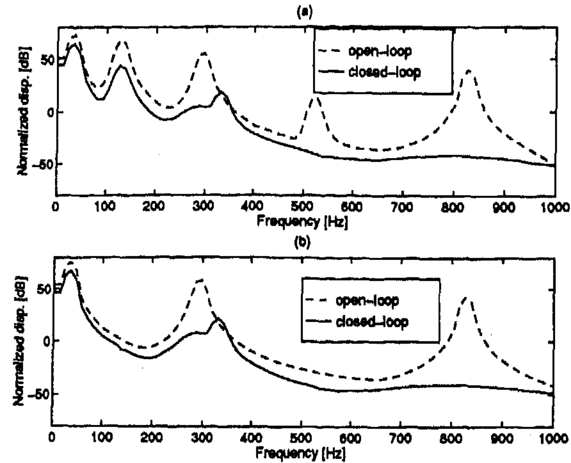


Figure 8: Normalized beam displacement

it is expected that the third mode vibrations will be attenuated the most. The matrix $\boldsymbol{R}$ was chosen so that the driving voltages never exceed 100 volts AC.

Table 1: Natural frequencies of simply supported actuated beam [Hz]

	analytical	FE full	truncated
f_1	31.34	32.45	32.45
f_2	125.36	130.10	130.13
f_3	282.05	294.11	297.05

Results — Figure 8 examines the spectral conetnt of the transient response of the beam to impulse excitation, as measured by the left (a), and middle (b), sensors. The measurement is scaled to a specified sensitivity of 10^{-6} meters, and presented in [dB]'s. Both sensors clearly show that the greatest reduction in displacement was achieved for the third mode, which was the mode we have designed for maximum attenuation through actuator placement and choice of $\boldsymbol{Q}$. The second mode is not measured by the second sensor, which is located at that mode's nodal location. In addition, we see that the third natural frequency was shifted upwards by the active control.

Tables 2, 3 present the open and closed-loop natural frequencies and damping ratios for the first three modes of the beam, respectively. A significant increase in the damping ratios is observed, as well as a shift of over 10% in the third natural frequency.

Table 2: Natural frequencies of a controlled beam

	ω_1 [Hz]	ω_2 [Hz]	ω_3 [Hz]
open-loop	32.6	130.1	294.1
closed-loop	32.6	129.6	332.5

An $\mathcal{H}_\infty$ controller was designed and applied to a beam similar to that in Figure 7, but with an overall length of 48 [cm], and an actuator length of 10 [cm]. The external force weighting filter was designed to admit forces at the third open-loop resonant frequency

Table 3: Modal damping ratios of a controlled beam ($\times 10^{-3}$)

	ζ_1	ζ_2	ζ_3
open-loop	0.10	0.41	0.92
closed-loop	10.0	51.0	20.0

of 200 [Hz], and scale them to an amplitude of 1 [N]. The open and closed-loop frequency responses of the beam scaled to 1 [μ] = 10^{-6} meters are presented in [dB]'s in Figure 9. As can be seen good control was achieved for all three fundamental modes of the beam, in accordance with the control design objective. It is interesting to note that for the non-resonant frequencies around 150 [Hz] control 'spill-over' occured, and the control effort led to higher displacements. This undesirable result can be reduced by incorporating the information which was lost due to the discretization of the continuous structure and to the balanced truncation of the states as an unstructured uncertainty in the $\mathcal{H}_\infty$ controller design. This is the subject of further research.

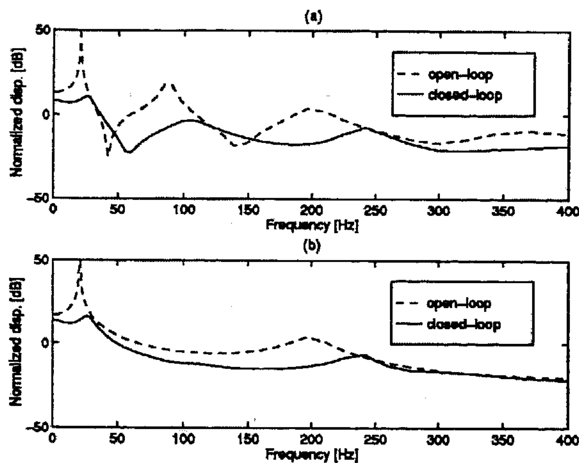


Figure 9: Normalized beam displacement; (a) – left sensor, (b) – midspan sensor

CONCLUDING REMARKS

A general methodology for the design of adaptive structures using the finite element method is presented. Coupled-field piezoelectric elements are shown to be a valuable tool in the initial stages of the design. Replacement of coupled-field elements with equivalent mechanical loads is shown to cause a relatively small loss of accuracy in the model.

The capability of the balanced truncation technique to reduce the order of the model by as much as an order of magnitude without losing significant information is demonstrated. FE sub-structuring in conjunction with balanced truncation are able to reduce large and complex models to orders previously obtainable solely through the modal analysis of idealized structures. This technique enables the use of the full range of FE capabilities in modeling geometrically and materially non-homogeneous structures for the low-order and accurate discretization of continuous structures. The capability of representing complex structures as relatively low-order and balanced systems in state space opens the way for the application of advanced tools for MIMO controller design, as well as for numerical simulations of the closed-loop systems.

Future work should include the design of robust $\mathcal{H}_\infty$ controllers, which take into account uncertainties in the structural dynamic model, as well as boundary conditions and excitation regimes. Proof-of concept experiments should be performed to validate the concepts proposed in this study.

References

[1] M.W. Lin, A.O. Abatan, and C.A. Rogers. Application of commercial finite element codes for the analysis of induced strain-actuated structures. In C.A. Rogers and G.G. Wallace, editors, *Proceedings of the Second International Conference on Intelligent Materials*, 1994.

[2] ANSYS. *ANSYS User's Manual - Revision 5.0.* Swanson Analysis Systems, 1992.

[3] B. Pletner and H. Abramovich. A consistent methodology for the modeling of piezolaminated shells: Part 2 - numerical and experimental validation. *AIAA Journal*, 1996. Submitted for publication.

[4] W. Gawronski. *Balanced Control of Flexible Structures.* Springer, 1996.

[5] B.C. Moore. Principal component analysis in linear systems, controllability, observability and model reduction. *IEEE Transactions on Automatic Control*, 26:17–32, 1981.

[6] J.M. Maciejowski. *Multivariable Feedback Design.* Addison-Wesley, 1989.

[7] B. Pletner and H. Abramovich. A consistent methodology for the modeling of piezolaminated shells: Part 1 - theory. *AIAA Journal*, 1996. Submitted for publication.

[8] R.L. Clark, M.R. Flemming, and C.R. Fuller. Piezoelectric actuators for distributed vibrations excitation of thin plates: A comparison between theory and experiment. *Journal of Vibrations and Acoustics, Transactions of the ASME*, 115:332–339, 1993.

[9] A.E. Bryson and Y.C. Ho. *Applied Optimal Control.* Hemisphere Publishing Corporation, New York, 1975.

[10] C.T. Chen. *Linear Systems Theory and Design.* Holt Saunders International Editions, 1984.

[11] J.C. Doyle, B.A. Francis, and A.R. Tannenbaum. *Feedback Control Theory.* Macmillan, 1992.

970586

Rolling Bearing Analysis Codes "BRAIN" - The Estimation of Rolling Bearing Performance for an Automotive Application

Hirotoshi Aramaki, Yuji Nakano, and Yoshio Shoda
NSK Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A computer program package named "BRAIN" has been developed to simulate the kinematics and the performance of rolling bearings under various running conditions. The calculation time necessary for running BRAIN software on a PC is very short. Various outputs can be obtained using BRAIN such as running torque, roller skew angle, roller slippage, and PV values.

Several experiments have been conducted to confirm the validity of BRAIN. The running torque of a four point contact ball bearing and that of a tapered roller bearing were measured. In addition, the skew of the roller in a needle bearing was measured. These experimental results were compared with the calculation results. The experiments and the calculations showed good agreement.

INTRODUCTION

Typically, a car contains over 100 rolling bearings. The wide range of bearing types includes ball bearings, needle bearings, cylindrical roller bearings, and tapered roller bearings. The quantity and diversity of rolling bearings indicate their vital contribution to the performance of modern automobiles. The requirements on automotive bearings have become more strict every year. Now, user requirements include compactness and low cost in addition to matching, or enhancing, the previous performance level.

As computer technology progresses, designers of rolling bearings rely more and more on computers as a simulation tool to design better bearings to match user requirements within a short time. Rolling bearing static analysis with a computer is very popular. It is essential for the bearing designer to know basic information such as internal load distribution and deflection of the bearing. However, to create high performance bearings, the bearing designer also needs more detailed information such as bearing torque, PV value, and slip rate of rolling elements. In such cases, dynamic analysis of a rolling bearing is required. The dynamic analysis of rolling bearings encompasses friction analysis, so it is much more difficult than the static case. ADORE, which is a commercially available software package, is an example of a well-known dynamic rolling bearing analysis program. Researchers can analyze the performance of roller bearings as well as ball bearings with ADORE[1]. ADORE is a very advanced and sophisticated tool. Since ADORE is a transient analysis program, it takes considerable time to find the steady state of the bearing performance.

We have developed a computer program package named "BRAIN" to simulate the kinematics and the performance of rolling bearings under various running conditions. When running BRAIN on a PC, the calculation time is short, so BRAIN is expected to become a frequently used simulation tool. BRAIN consists of several separate programs that are applicable to angular contact ball bearings (including deep groove ball bearings, thrust ball bearings), four point contact ball bearings, radial cylindrical (needle) roller bearings, thrust cylindrical (needle) roller bearings tapered roller bearings, and self-aligning roller bearings. Various outputs, such as running torque, roller skew angle, roller slippage, and PV values, can be calculated to assist in the design of bearings for advanced applications.

FEATURES OF BRAIN

The name BRAIN was made from the phrase "**BeaRing Analysis In Nsk**." BRAIN is a software package to simulate the performance of rolling bearings. BRAIN consists of six programs as shown in Table 1, namely, BALTAC (Ball bearing), B4PTAC (Four point contact ball bearing), CYLTAC (Cylindrical roller bearing), TNBTAC (Thrust cylindrical roller bearing), TAPTAC (Tapered roller bearing), and SELTAC (Self-aligning roller bearing). Therefore, most types of rolling bearings can be analyzed by BRAIN.

BRAIN has several attractive features. In conventional bearing analysis, an assumption of "pure rolling" is made, i.e. the "Jones control theory"[2] for ball bearings and pure rolling between rollers and inner/outer rings. However, the assumption of "pure rolling" is not rigorous enough especially for high-speed applications. The gyroscopic moment and centrifugal force acting on

rolling elements result in large slip at the contact area.[3),4)] In addition, the skew of the rollers in a roller bearing is usually neglected in conventional analysis although it strongly influences the bearing performance. In BRAIN, a bearing is analyzed on the basis of an equilibrium of forces and moments without resorting to simplifying assumptions of pure rolling, no tilt, and no skew motion of rolling elements.

Several important assumptions are made when using BRAIN.

(1)Inner ring and cage rotational speeds are constant.
(2)The cage center is fixed.(no eccentricity)
(3)The orbital speed of rolling elements and the cage rotational speed is identical.

How BRAIN simulation works might be easily visualized in your mind's eye as follows: Imagine that you are observing the bearing motion by a stroboscope whose frequency is adjusted to Zfc (Z: number of rolling elements, fc : cage rotational frequency). The rolling elements stay at the same positions. The calculation model is similar to that of Harris[5)]. The crowning profile of the roller is considered approximately using a slicing method. More exact analysis, e.g. edge load calculation, requires inputting BRAIN's calculation results, such as contact force, tilt, and skew angle of the roller into software that solves contact problems.

In BRAIN, it is easy to set constraint conditions on the inner ring movement. This function confers an advantage when calculating bearing performance under given misalignments and/or under a position preloading condition.

In many cases, deformed shaft and housing influence the bearing performance. For these cases, BRAIN can import the deformation data relating to the shaft and the housing calculated by FEA. The deformation is considered as a change in the internal clearance of the bearing. Therefore, the internal clearance varies axially and/or circumferentially in BRAIN.

COORDINATE SYSTEMS

The Coordinate frames defined in BRAIN are shown in Fig. 1. Inertial coordinate frame X*Y*Z* is fixed in space. When the outer ring is assumed to be stationary and fixed in a housing, the outer ring coordinate frame XYZ is identical with the inertial coordinate frame X*Y*Z*. External axial load is applied in the X direction. External radial loads are applied in the Y and Z directions. The shaft or housing rotates clockwise around the X axis. Each rolling element has its own local coordinate frame $x_j y_j z_j$, which rotates with the rolling element. All the coordinate systems are right handed ones. Azimuth angle of the jth rolling element is defined by the angle between Z axis and z_j axis. For convenience, radial load is usually applied in the Z direction. Therefore, rolling elements have their maximum contact load at an azimuth angle equal to zero.

FORCES AND MOMENTS ACTING ON ROLLING ELEMENTS

Fig. 2 shows forces and moments acting on rolling elements in a ball bearing. The contact forces between outer/inner ring and rolling elements are Q_1 and Q_2. The centrifugal force and gyroscopic moment are Qi and Mg respectively. Gravity is neglected in the analysis, which is thought to be negligible in automotive applications. Cage contact force Q_{ca}, friction F_{cx} and F_{cz} are considered at the cage pocket area.

Frictional forces work at all contact areas, such as F_{w1} and F_{w2}. For the estimation of frictional force and moment at ring/ball contact area, the elliptical contact area is divided into small slices. The frictional force and moment at the contact area are obtained by a numerical integration of frictional force on each slice. Coefficient of friction is determined by an empirical equation which is a function of the lubricant's viscosity, rolling velocity, slide/roll ratio, and contact pressure. The empirical equation was derived from a friction database based on results from two-disk test machines. Rolling resistance is estimated by Aihara's equation[6)], which use Goksem and Hargreaves' rolling traction force[7),8)]. The basis of the model about BALTAC is described in Aramaki et. al.[9)]. Coulomb friction is assumed for cage/rolling elements in contact with frictional forces F_{CX}, F_{CZ}. Cage land friction F_{CL} is estimated by a simple viscous model.

Fig. 3 shows forces and moments acting on the rolling element in a tapered roller bearing. The half cone angle and half roller cone angle are α and β respectively. The roller is sliced into thin disks. The contact force Q_1 and Q_2 are obtained by numerically integrating the contact force of each thin disk to consider the crowning profile. In tapered roller bearings and self-aligning roller bearings, flange contact force Q_f is also considered. The frictional forces are estimated by numerical integration of the frictional force on each thin disk of the rolling element as well as by the estimation of contact forces.

EQUILIBRIUM OF FORCES AND MOMENTS

BALL BEARINGS - From the balance of forces and moments, one can obtain the following equations for the rolling elements of the ball bearing shown in Fig. 2.

$$0 = -Q_{1,j}\sin\alpha_{1,j} + Q_{2,j}\sin\alpha_{2,j} + F_{w1,j}\cos\alpha_{1,j} - F_{w2,j}\cos\alpha_{2,j} + F_{cx,j} \quad (1)$$

$$0 = F_{y1,j} - F_{y2,j} + F_{d,j} + Q_{ca,j} \quad (2)$$

$$0 = -Q_{1,j}\cos\alpha_{1,j} + Q_{2,j}\cos\alpha_{2,j} + Q_I - F_{w1,j}\sin\alpha_{1,j} + F_{w2,j}\sin\alpha_{2,j} + F_{cz,j} \quad (3)$$

$$I_B(d\omega_{x,j}/dt) = -M_{1x,j} - M_{2x,j} + M_{r1x,j} + M_{r2x,j} - D_a F_{cz,j}/2 \quad (4)$$

$$I_B(d\omega_{y,j}/dt) = -M_{gy,j} + D_a(F_{w1,j} + F_{w2,j})/2 \quad (5)$$

$$I_B(d\omega_z/dt) = M_{gz,j} - M_{1z,j} - M_{2z,j} + M_{r1x,j} + M_{r2x,j} + D_a F_{cx,j}/2 \quad (6)$$

Eq. (1)-(3) are force balances and (4)-(6) are moment balances with respect to x_j, y_j, and z_j directions respectively. Additional contact loads, frictional forces and moments should be considered with four point contact ball bearings. As shown in above equations, the

acceleration term (left hand side) in the force balance equations are neglected. The angular acceleration $I_B(d\omega/dt)$ in (4)-(6) are approximated in the following manner.

$$I_B(d\omega/dt) = I_B(d\omega/d\varphi)(d\varphi/dt)$$

$$= I_B(\omega_j - \omega_{j-1})/(\varphi_j - \varphi_{j-1})\omega_c \quad (7)$$

From the cage assumptions of steady rotation and no eccentricity, the following equation is obtained for the cage moment balance in the rotational direction.

$$0 = -0.5\, D_m\, \Sigma\, Q_{ca} + D_{cr}\, F_{cL} \quad (8)$$

where, D_m : pitch diameter of the bearing
D_{cr} : cage land diameter

The inner ring is assumed to be rotated with a constant speed so that the following five equations describe the five degrees of freedom for the inner ring.

$$F_X = \Sigma\, (Q_{2,j}\sin\alpha_{2,j} - F_{w2}\cos\alpha_{2,j}) \quad (9)$$

$$F_Y = \Sigma\, (Q_{2,j}\cos\alpha_{2,j} - F_{w2}\sin\alpha_{2,j})\sin\varphi_j \quad (10)$$

$$F_Z = \Sigma\, (Q_{2,j}\cos\alpha_{2,j} - F_{w2}\sin\alpha_{2,j})\cos\varphi_j \quad (11)$$

$$M_Y = -\Sigma\, \{ (-0.5\, D_m \sin\alpha_{2,j})Q_{2,j} + (0.5\, D_a - 0.5\, D_m \cos\alpha_{2,j})\, F_{w2,j} \} \cos\varphi_j \quad (12)$$

$$M_Z = \Sigma\, \{ (-0.5\, D_m \sin\alpha_{2,j})\, Q_{2,j} + (0.5\, D_a - 0.5\, d_m \cos\alpha_{2,j})\, F_{w2,j} \} \sin\varphi_j \quad (13)$$

where left hand side F_X, F_Y, F_Z, M_Y, M_Z are external forces and moments acting on the inner ring. In the program, any of the above equations can be omitted to impose a constraint on the inner ring.

Above 6 x (rolling elements) + 6 equations are solved simultaneously using the Newton-Raphson method.

ROLLER BEARINGS - The equilibrium equations are derived for roller bearings. As an example, the following equations are obtained for the tapered roller bearing shown in Fig. 3.

$$0 = -Q_{1x,j} + Q_{2x,j} + Q_{fx,j} + Q_{ix} \quad (14)$$

$$0 = F_{y1,j} + F_{y2,j} + F_{fy,j} + F_{d,j} + Q_{ca,j} \quad (15)$$

$$0 = -Q_{1z,j} + Q_{2z,j} + Q_{fz,j} + Q_{iz} + F_{cz,j} \quad (16)$$

$$I_x(d\omega_{x,j}/dt) = M_{1x,j} + M_{2x,j} + M_{fx,j} + M_{r1x,j} + M_{r2x,j} + M_{rfx,j} + M_{cz,j} \quad (17)$$

$$0 = M_{1y,j} + M_{2y,j} + M_{fy,j} + M_{gy,j} \quad (18)$$

$$0 = M_{1z,j} + M_{2z,j} + M_{fz,j} + M_{gz,j} + M_{cx,j} \quad (19)$$

where M_f : moment due to flange contact
M_c : moment due to cage/roller contact

Excluding the x rotation, the other angular acceleration terms (on left hand side) are ignored.

Eq. (8) is applicable even for the moment balance of a roller bearing. Double row type bearings can be calculated in CYLTAC (cylindrical roller bearings) and SELTAC (self-aligning roller bearings) so that eq. (8) is applied to each cage when two cages are assembled in the bearing. One can also obtain the equilibrium equations for the inner ring in the same way as in eq.(9) - (13).

Some self-aligning roller bearings have floating rings. In this case, force and moment balance equations for the floating ring are necessary in addition to the above equations. Since use of self-aligning bearings in an automobile is very rare, we omit an explanation about the floating ring.

CALCULATION EXAMPLES

As already described, BRAIN can analyze the performance of most rolling bearings. In the present paper, the analyses of a four point contact ball bearing, a radial needle bearing, and a tapered roller bearing are shown.

The use of four point contact ball bearings in an automobile is not popular since they would generate high power loss and severe ring/ball wear due to large spin friction. However, their forward/reverse axial load supporting capability is very attractive in terms of cost and compactness. The estimation of the PV values at ball/ring contact points under radial/axial combined loading condition is a difficult task in a conventional analysis. BRAIN can estimate PV value distribution at each contact area as well as bearing power loss. Therefore, optimization of four point contact ball bearings becomes feasible.

Many needle bearings are used in automotive applications thanks to their compactness. Due to the skew motion of the rollers, however, high power loss is a potential problem. In roller bearings, skewing is defined as an angular rotation of the roller axis in a plane tangent to its orbital direction with respect to the axis of the contacting ring according to Harris[3]. The skew motion of the roller is influenced by many factors, such as lubricant and geometry of the bearing. Although the estimation of the skew of the roller is difficult work, it is very important to analyze the bearing performance including the skew influence.

The final example is a tapered roller bearing. Many tapered roller bearings are used in automobiles. The running torque of tapered roller bearings at relatively low speed can be estimated well by Aihara's equation[6]. However, the drag loss may constitute a relatively large portion of the bearing power loss when the bearing runs at high speed with plenty of feeding oil. A relatively large tapered roller bearing was tested to measure the running torque.

FOUR POINT CONTACT BEARING ANALYSIS - The performance of radial and thrust ball bearings is calculated with the BALTAC program, especially to evaluate the bearing torque (power loss) and PV value at

the ring/ball contact area. For three point and four point contact ball bearings, B4PTAC has been developed for the analysis of gothic-arch shaped ring grooves. B4PTAC is based on BALTAC, and it utilizes many of the same subroutines as BALTAC does.

Fig. 4 shows a schematic view of the experimental apparatus used to measure the running torque. The experiment was conducted to compare the measured bearing torque and the one calculated with B4PTAC. Radial and axial loads were applied to the test bearing through hydrostatic bearings. As the inner ring rotates, the bearing's frictional torque works to turn the bearing housing. The force necessary to restrain the housing was measured and converted to bearing torque. The outer ring temperature was also measured by a thermocouple. The speed range in the experiment was from 1000 rpm to 10000 rpm.

A four point contact ball bearing was used in the experiment. The bore diameter and outside diameter of the bearing are 40mm and 90 mm, respectively. The rest angle of the bearing is 60 deg. 12 balls are inserted in the bearing. The experimental conditions are shown in Table 2. Case A is a pure axial loading condition. Case B is a combined axial and radial loading condition.

The comparison of experiment and calculation results under case A is shown in Fig. 5. Solid line and broken line in the figure show experimental results and calculated ones respectively. Since no radial load is applied, the test bearing works as a conventional angular contact ball bearing. Below about 3000 rpm, the torque increased rather sharply with speed. However, the slope of the torque became smaller as speed increased. Reduction of oil viscosity due to temperature rise seems to play an important role in this phenomenon. The experiment agrees with the calculation excellently. The discrepancy between the experiment and the calculation was below 10%.

The comparison of the experiment and the calculation under case B is shown in Fig. 6. The measured running torque under radial/axial combined load (case B) was about 30% higher than that under pure axial load (case A). Under case B, the agreement between the experiment and the calculation is reasonable. The calculation is seems to always be below the experiment. Although the maximum discrepancy between the experiment and the calculation is about 16%, the tendency of bearing torque to vary with speed is well-simulated by the calculation.

Fig. 7 shows the PV value distribution with the ball azimuth φ under a pure axial loading condition (case A). This PV value is the peak PV value (pressure P$\times$slip velocity V) along the major axis of elliptical contact between the ball and inner/outer ring. Black plots in the figure show axial loading side grooves (outer ring No.2 and inner ring No.1 groove) and white ones show unloading side grooves (outer ring No.1 and inner ring No.2 groove). Under this load condition, a ball always makes only one contact point with the outer and with the inner ring. Therefore, PV values on the unloading side grooves are zero. In such cases, balls can rotate with the minimum spin friction. PV values at inner ring No.1 are about 0.2 [GPa·m/sec].

Fig. 8 shows the PV value distribution under the combined axial and radial loading condition (case B). The PV value has its maximum at a radial loading position (φ = 0). The inner ring has a higher contact pressure and a higher PV value than the outer ring has. The maximum PV value is 1.5 [GPa·m/sec], and it appears at the radial loading position of the inner ring groove No.1. The PV values under case B are much higher than those under case A. In the present calculation, the peak PV value in the ball/ring contact area appears near the edge on the major axis of the contact ellipse due to the spin velocity.

NEEDLE BEARING ANALYSIS - Roller skew angle and axial force generated by roller skewing in a single row needle bearing is analyzed by CYLTAC. Input data for the calculation is shown in Table 3.

A schematic view of the test part is shown in Fig. 9. Needle rollers were inserted without a cage between the housing and the shaft. The shaft was rotated and the housing was stationary. The external radial load and moment were applied to the housing. Two eddy current type gap sensors were aligned axially and mounted in the housing. The housing was fixed by a thin plate and strain gauges were attached to the thin plates.

Roller skew angle was determined by the phase difference in peak signals from the two gap sensors. Axial force was measured by strain gauges on a thin plate. A positive sign for a skew angle means that a roller passed over the sensor 2 before the roller passed over the sensor 1. The directions of shaft rotation and axial force are shown in Fig. 9.

Fig. 10 shows the skew angle of the roller at the maximum loading position (azimuth = 0 deg). Black plots in the figure show the calculated skew angles and white ones show the measured results. The absolute values of measured skew angle are also shown in the figure. In the experiment, the roller skew angle changed from positive to negative suddenly at around 3000 rpm. The absolute values of skew angle were at around 1 deg and almost constant in the speed range of 1000 - 6000 rpm. In the calculation, the roller skew angle is around +1 deg. It is still not clear why the roller changed its sign at 3000 rpm in the experiment. The rollers might have two stable positions (+1 deg and -1 deg) for the roller skew motion.

Fig. 11 shows the axial force generated by the skew motion of the roller. For reference, it also shows the negative of the measured axial forces for the region above 3000 rpm. Accordingly with the change in the sign of the skew angle at 3000 rpm, the axial force also changed its sign in the experiment. The absolute value of the axial force decreased with increasing speed. Though the signs differed in the region above 3000 rpm, the magnitude of the experimental and calculated values showed good agreement.

TAPERED ROLLER BEARING ANALYSIS - The performance of tapered roller bearings such as running torque, roller revolution, cage rotation, roller skew angle, is analyzed by TAPTAC. Input data of the calculation is

shown in Table 4. The test bearing is HR32217 which has a bore diameter of 85 mm.

The experimental apparatus was similar to that shown in Fig. 4. However, the shaft was arranged horizontally instead of vertically as shown in Fig. 4. Jet lubrication system was adopted in the experiment and a mineral oil having a viscosity of 32 mm^2/s at 40°C was used. Three jet nozzles were arranged to feed the oil to the bearing.

Fig. 12 shows the relationship between bearing torque and speed. Plots show the measured torque while the broken line shows the calculated torque. The measured torque agrees with the calculated one when the oil feeding rate was 1 l/min. On the other hand, the calculated torque shows higher values than the measured one, especially in the mid-speed range when the oil feeding rate was 2 l/min. This difference is attributed to drag loss, since good agreement existed when fed with a small amount of oil. Although the maximum discrepancy is 20% at most, more research effort is necessary to estimate the drag loss better.

CONCLUSION

A new rolling bearing analysis program package named BRAIN has been developed. BRAIN software gives results in several minutes when run on a PC. BRAIN can calculate the kinematics and the performance of most rolling bearings. The calculation results are useful, especially for advanced applications involving rolling bearings.

The experiment with the four point contact ball bearings was conducted to compare results with those from BRAIN. The running torque of the four point contact bearing was well-simulated by BRAIN. The PV distribution of the four point contact ball bearing was also calculated. The PV value of four point contact ball bearings is an essential design parameter in any attempts to reduce the power loss and to prevent severe wear.

A needle bearing experiment was also conducted. The skew angle of the rollers and resultant axial force were measured and compared with those from BRAIN. The measurements agreed with the calculations although their signs were opposite of each other in the high speed region.

The running torque of a tapered roller bearing was measured and compared with the calculation from BRAIN. The tendency of the measured torque was well-simulated by BRAIN. The agreements between the experimental and calculated results were good.

REFERENCES

1) Gupta, P.K., Advanced Dynamics of Rolling Elements, Springer-Verlag New York Inc.,1984.

2) Jones, A.B., "Ball Motion and Sliding Friction in Ball Bearings," ASME Journal of Basic Engineering, Vol.3, 1959, pp.1-12.

3) Harris, T.A., Rolling Bearing Analysis, 3rd Ed., John Wiley & Sons,Inc.,1991.

4) Touma, K., Kawamura, H., Kawakita, K., "Ball Motion in High-Speed Angular Contact Ball Bearings," Proc. of JSLE Int. Trib. Conf., 1985, pp.585-590.

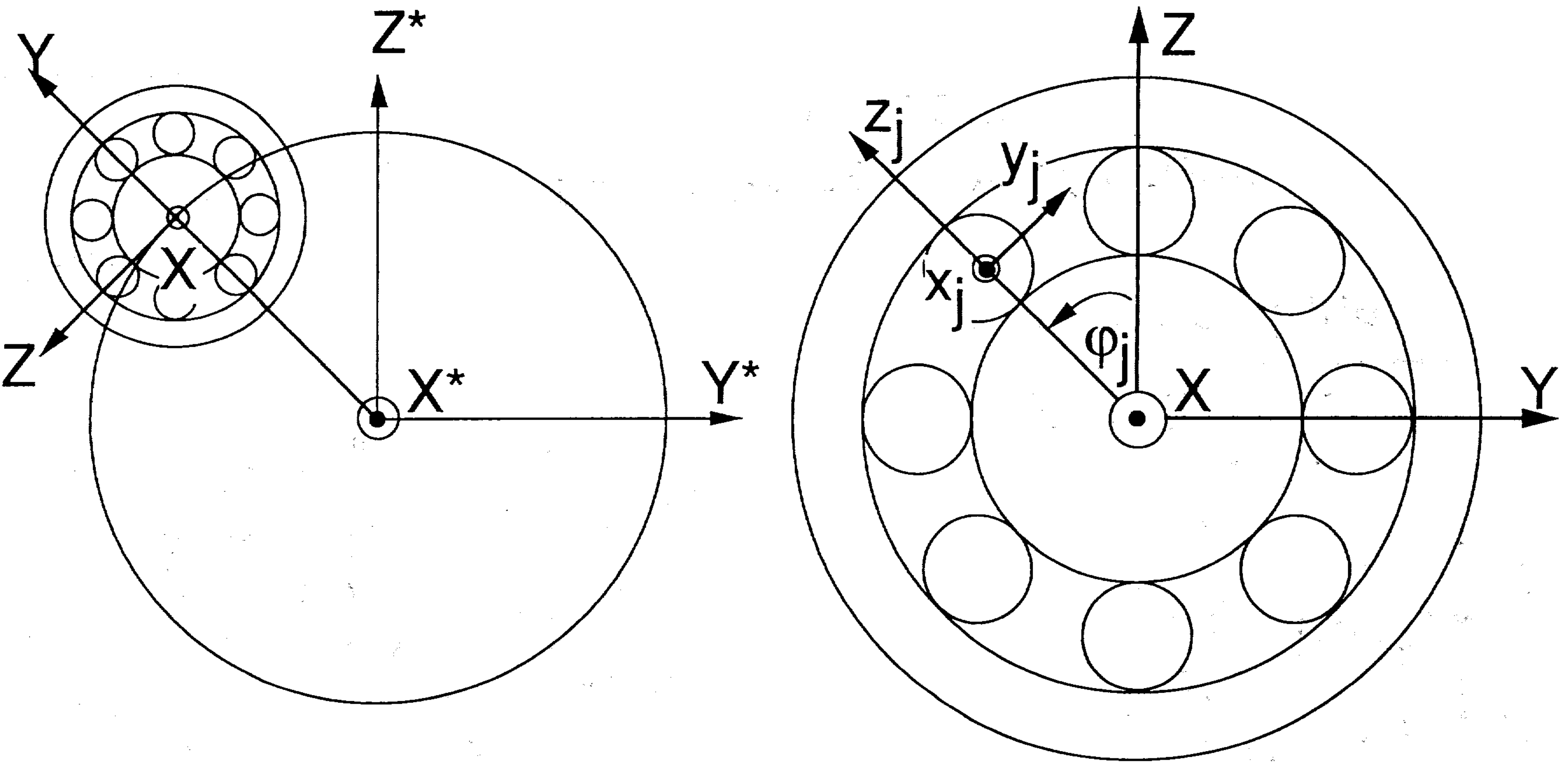
5) Harris, T.A, and Mindel, M.H., "Rolling Element Bearing Dynamics," Wear, Vol.23, 1973, pp.311-337.

6) Aihara, S., "A New Running Torque Formula for Tapered Roller Bearings Under Axial Load," ASME Journal of Tribology, Vol.109, 1987, pp. 471-478.

7) Goksem, P.G., and Hargreaves, R.A., "The Effect of Viscous Shear Heating on Both Film Thickness and Rolling Traction in an EHL Line Contact," ASME Journal of Lubrication Technology, Vol.100, 1978, pp.346-358.

8) Goksem, P.G., and Hargreaves, R.A., "An Approximate Equation for Predicting the Isothermal and Non-isothermal Starved Rolling Traction in an EHL Line Contact," ASME Journal of Lubrication Technology, Vol.100, 1978, pp.472-478.

9) Aramaki, H., Shoda, Y., Morishita, Y., Sawamoto, T., "The Performance of Ball Bearing with Silicon Nitride Ceramic Balls in High Speed Spindles for Machine Tools", ASME Journal of Tribology, Vol.110, 1988, pp.693-698.



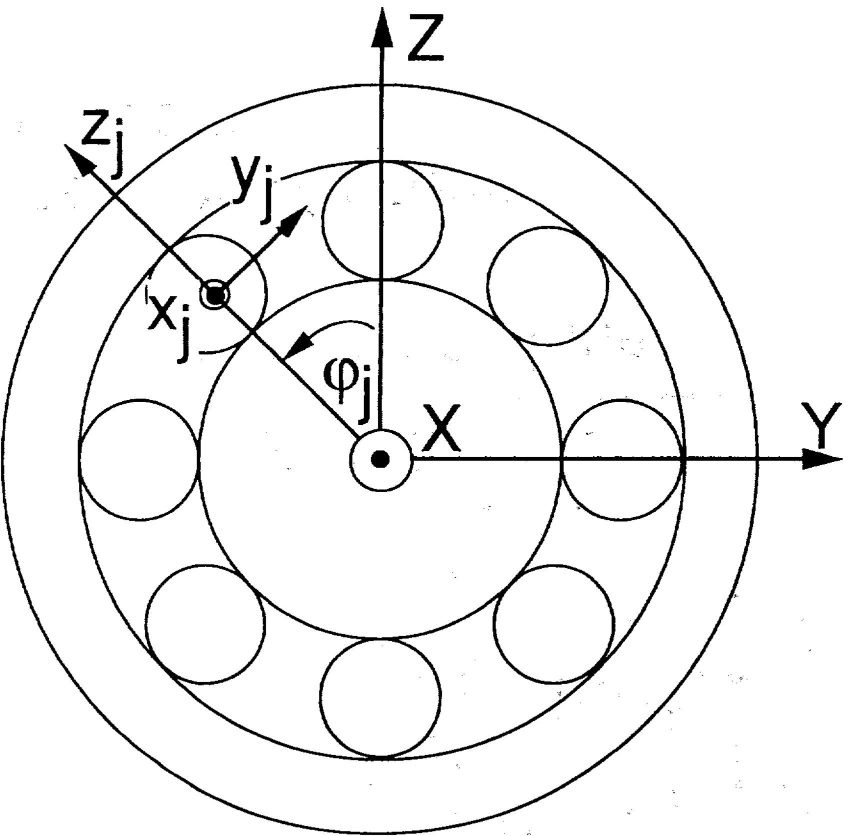


Fig. 1 Coordinate System

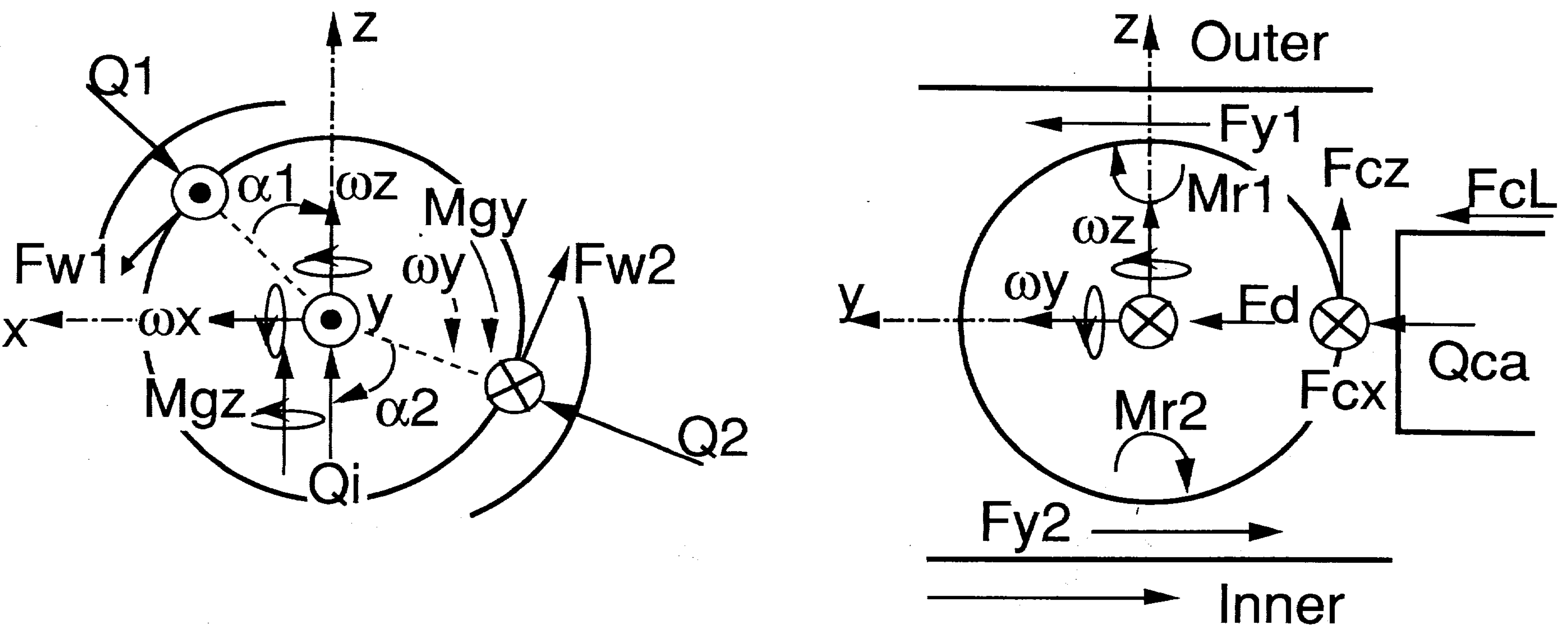


Fig. 2 Forces and Moments
(Ball Bearing)

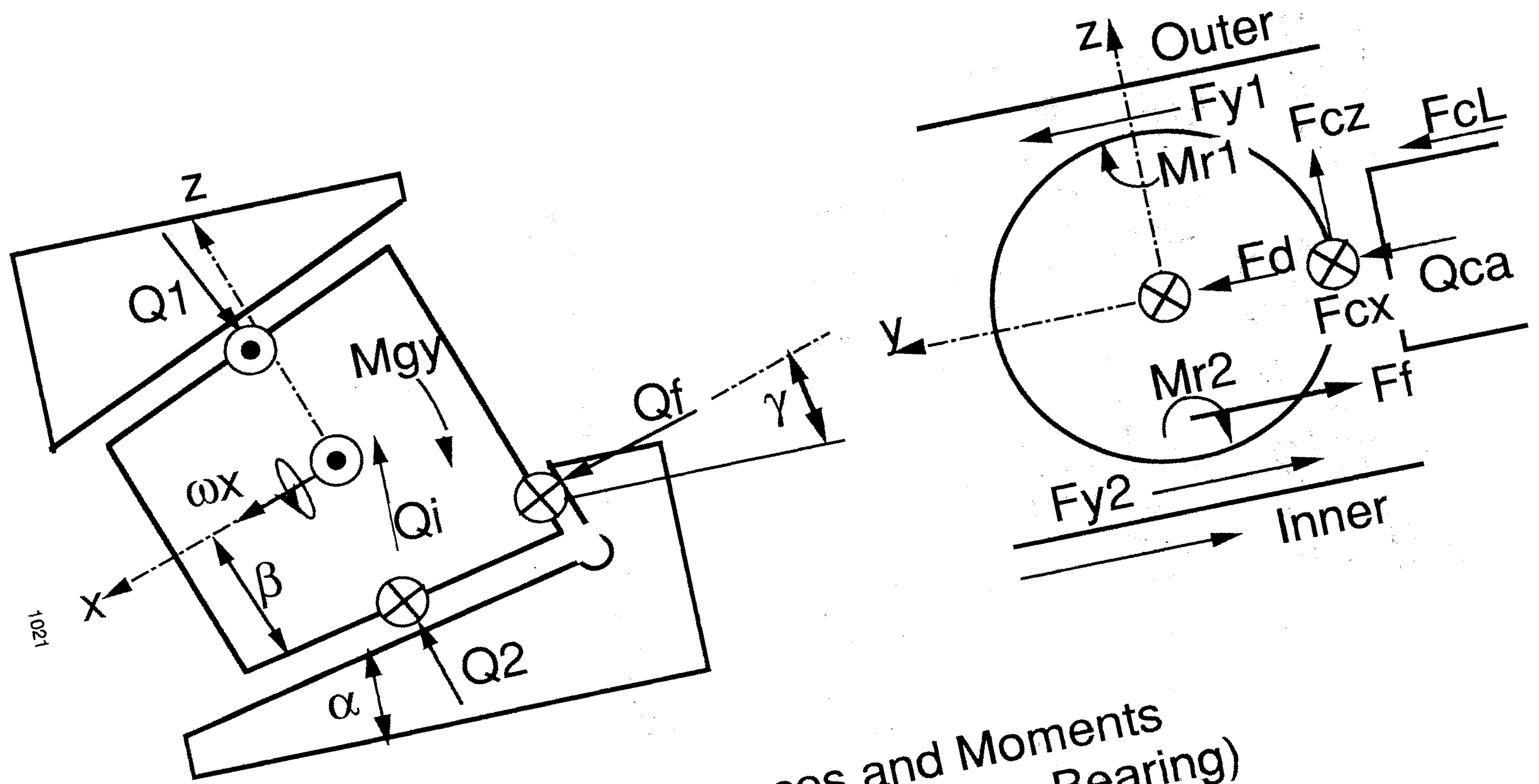


Fig. 3 Forces and Moments
(Tapered Roller Bearing)

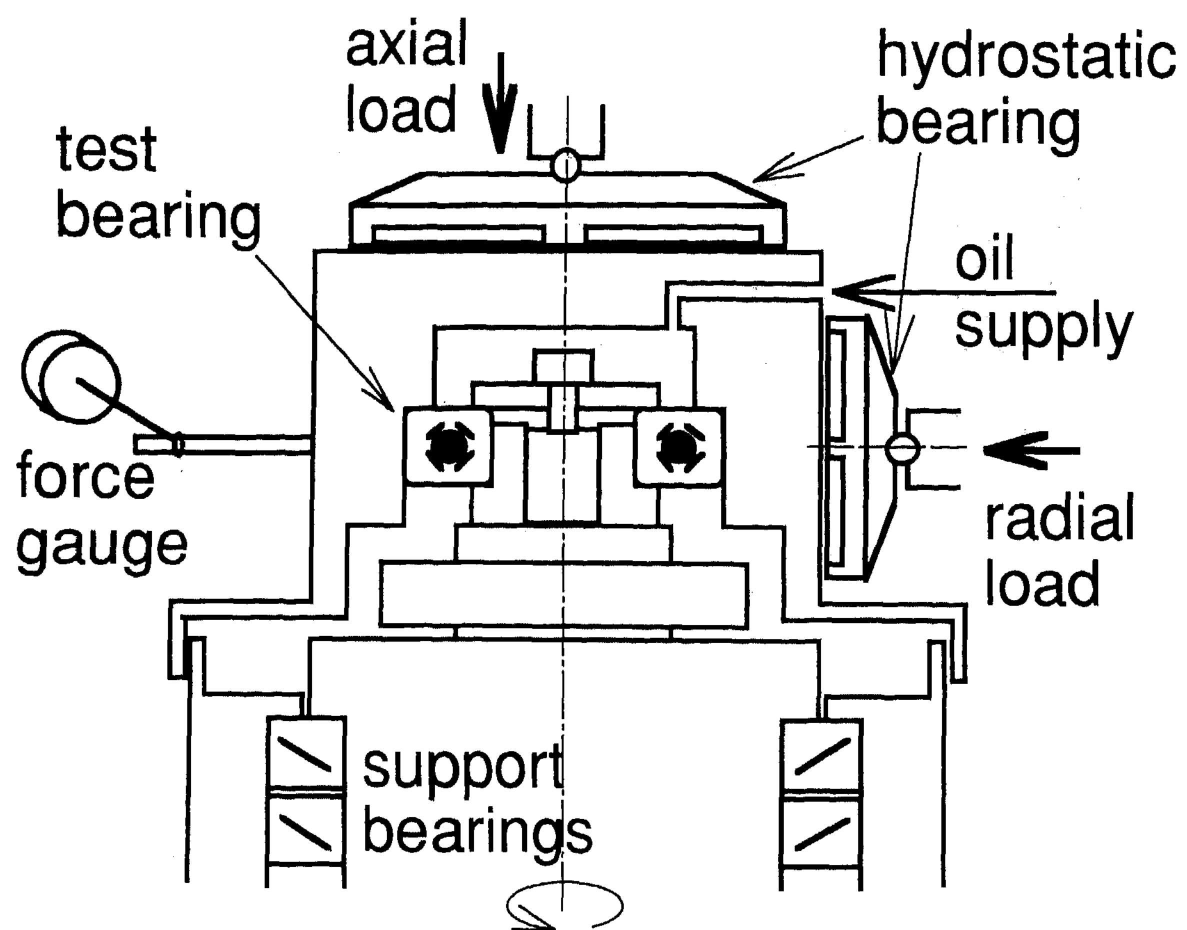


Fig. 4 Running Torque Test Rig

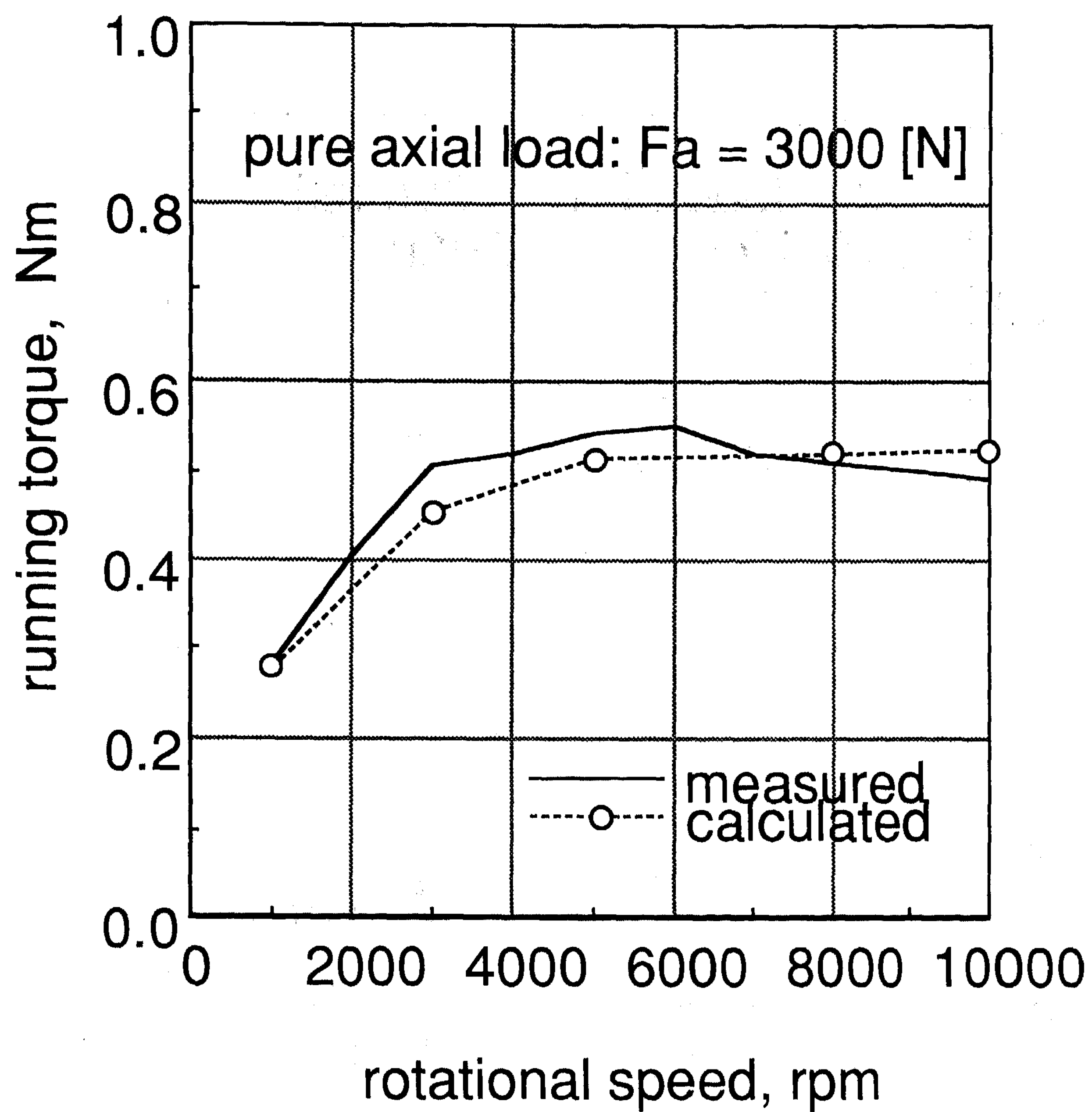


Fig. 5 Running Torque of Four Points Contact Bearing (case A)

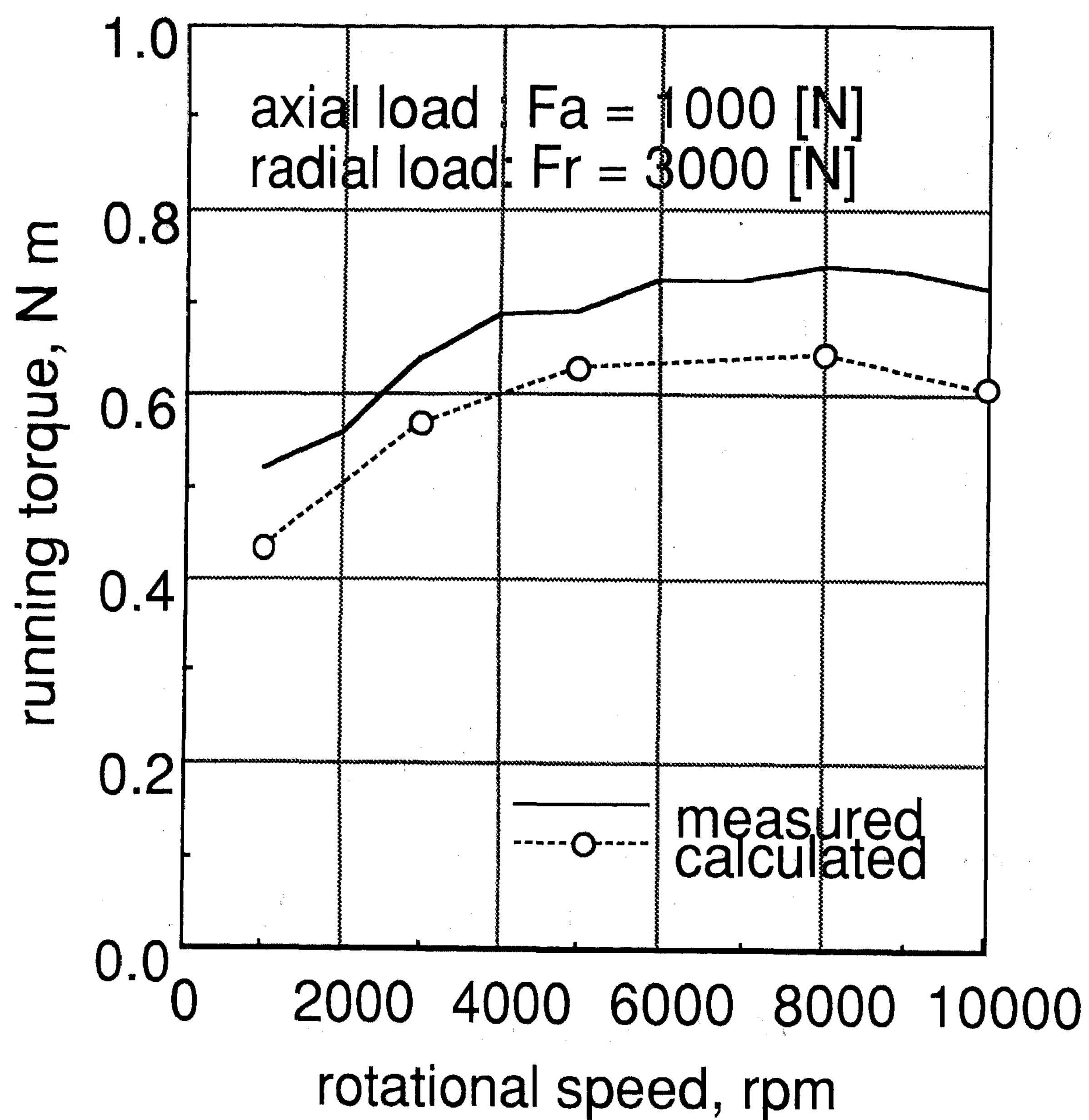


Fig. 6 Running Torque of Four Points Contact Bearing (case B)

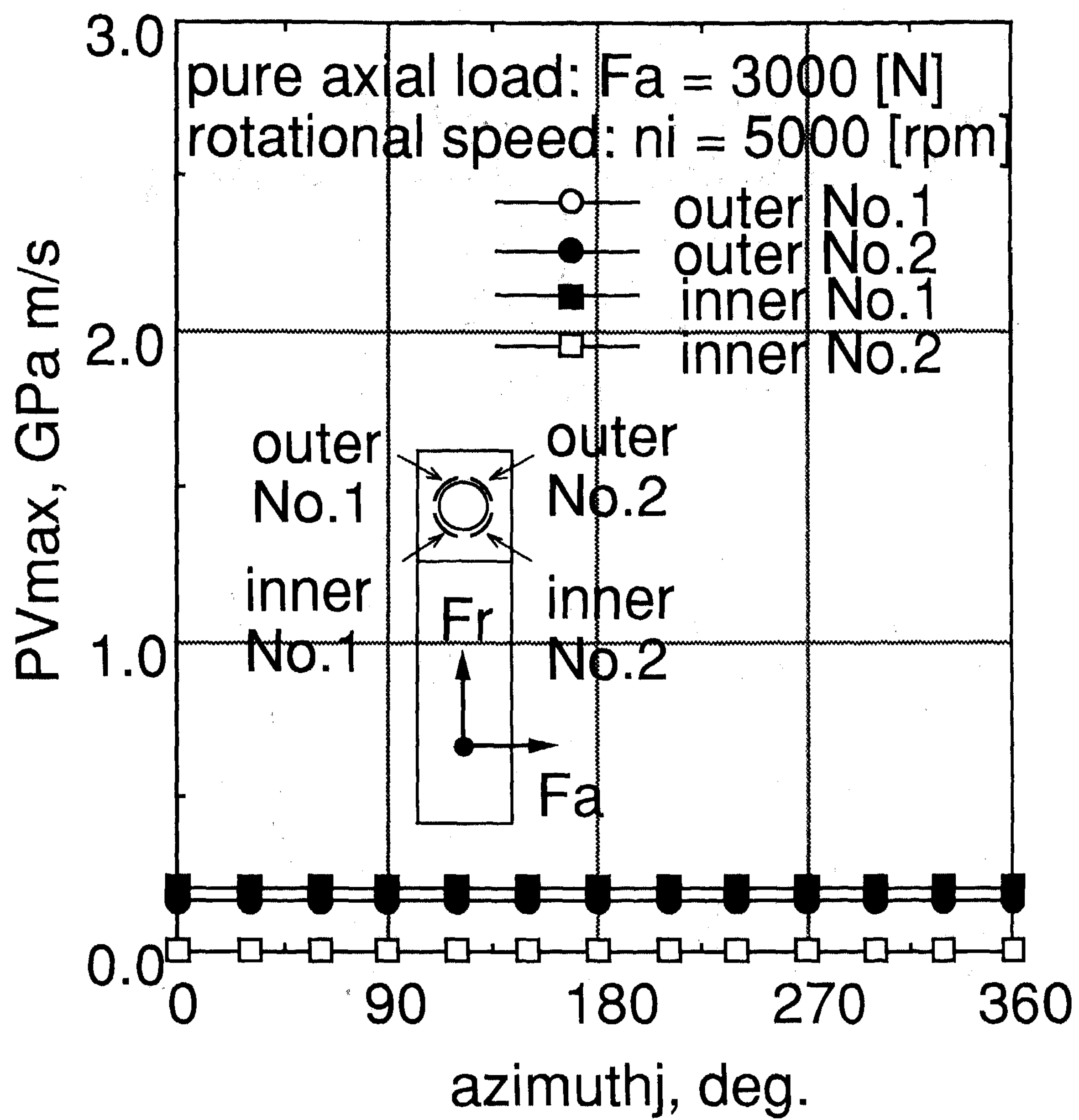


Fig. 7 PV value (case A)

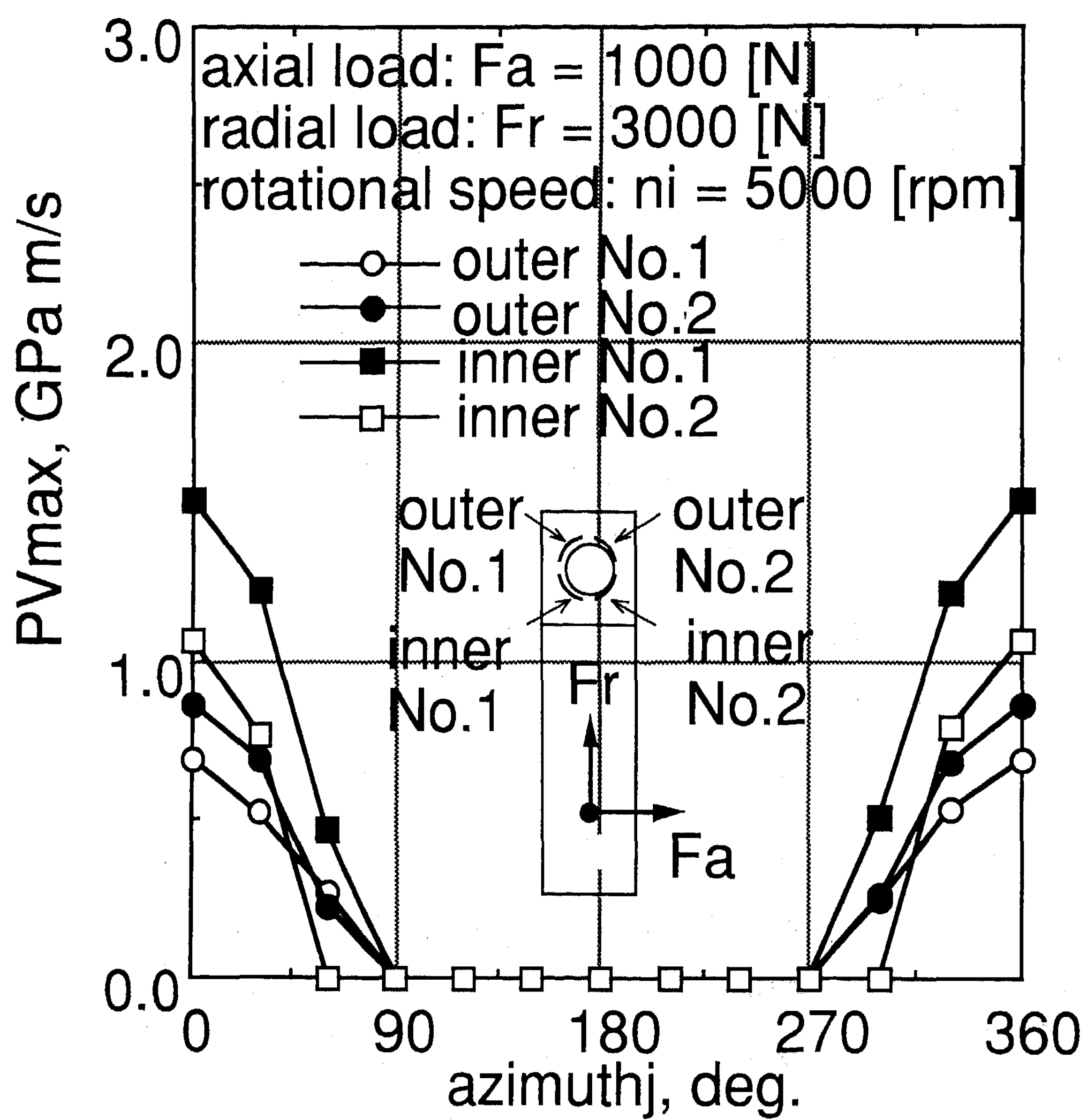


Fig. 8 PV value (case B)

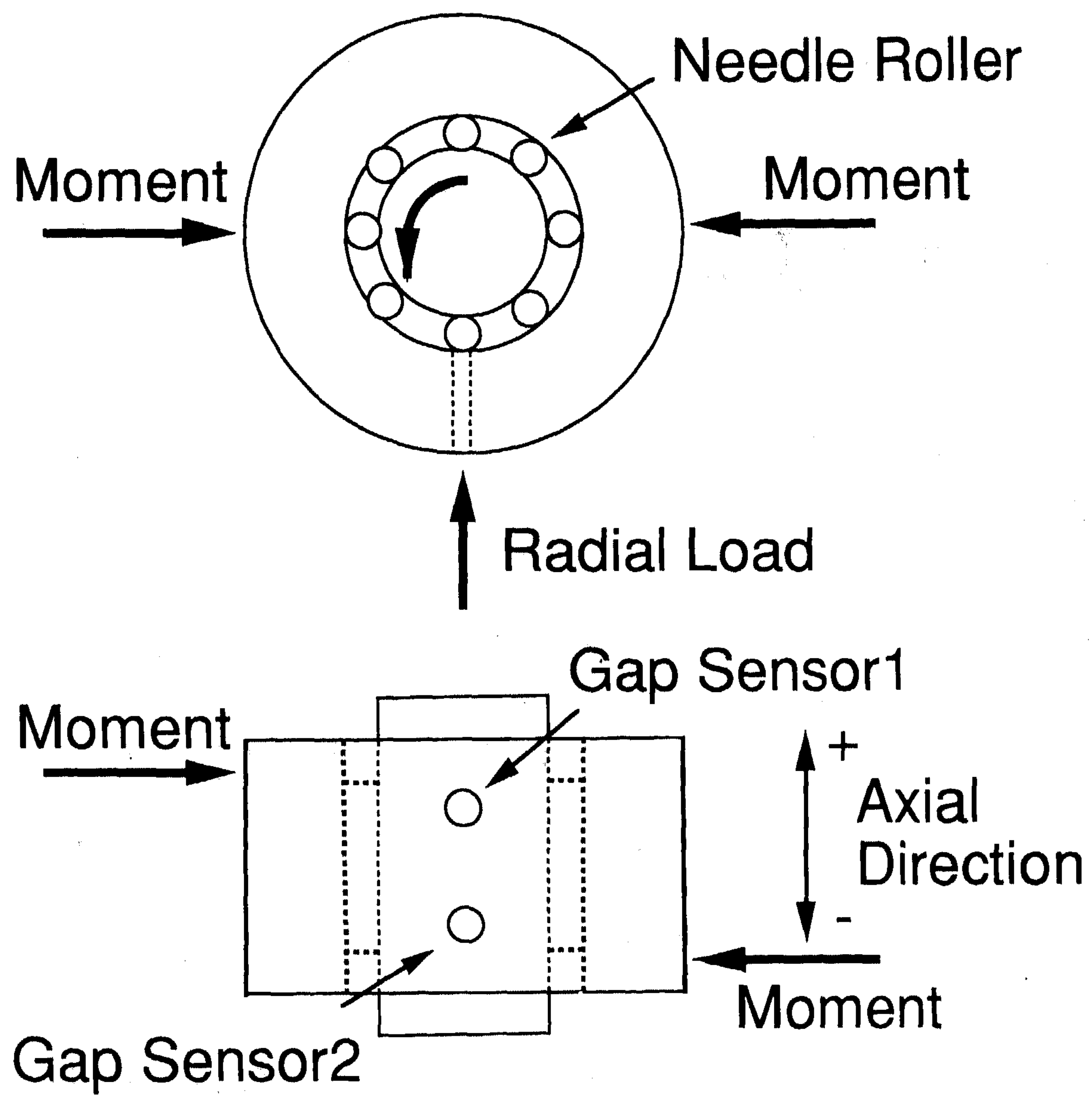


Fig. 9 Schematic View of the Test Part

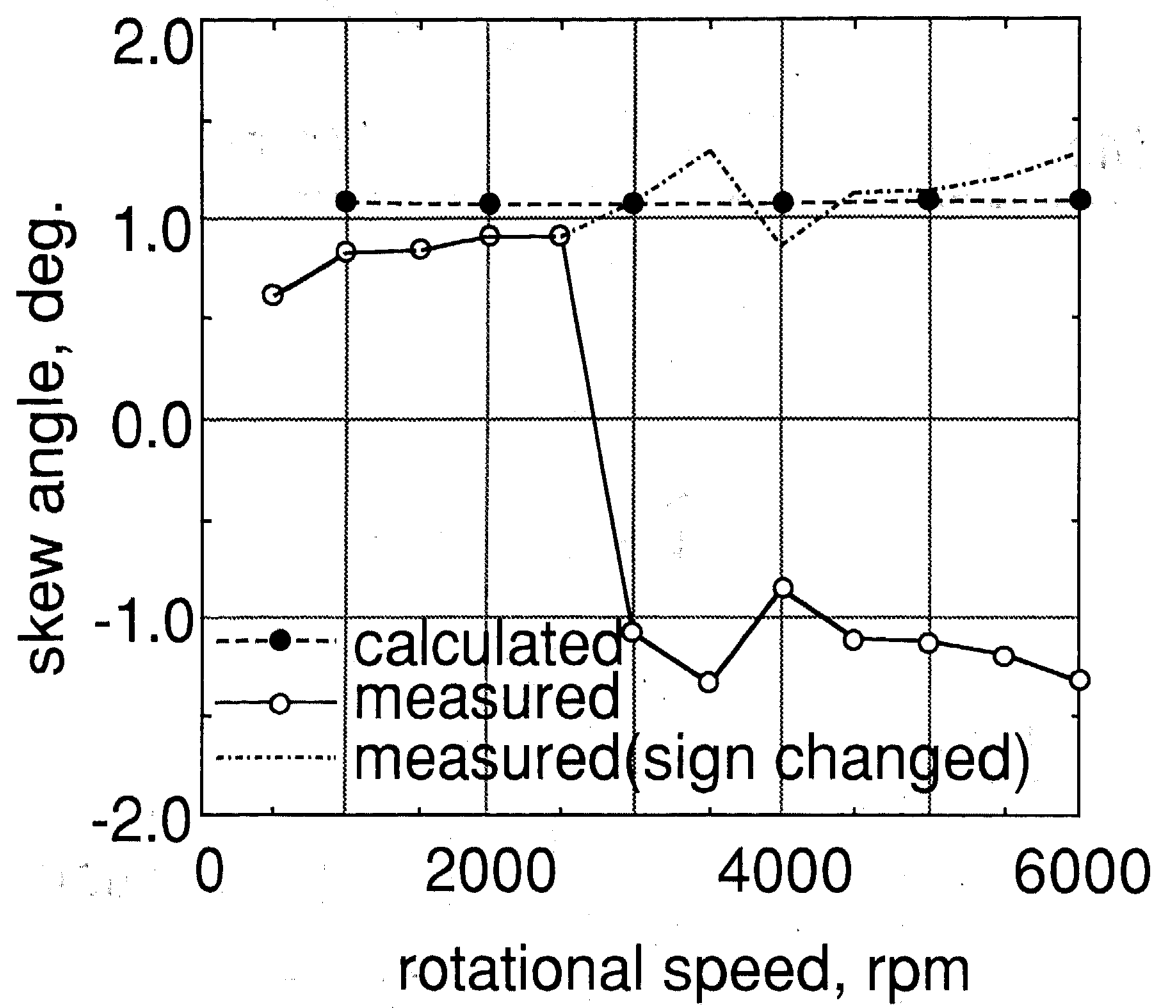


Fig. 10 Skew Angle of the Needle

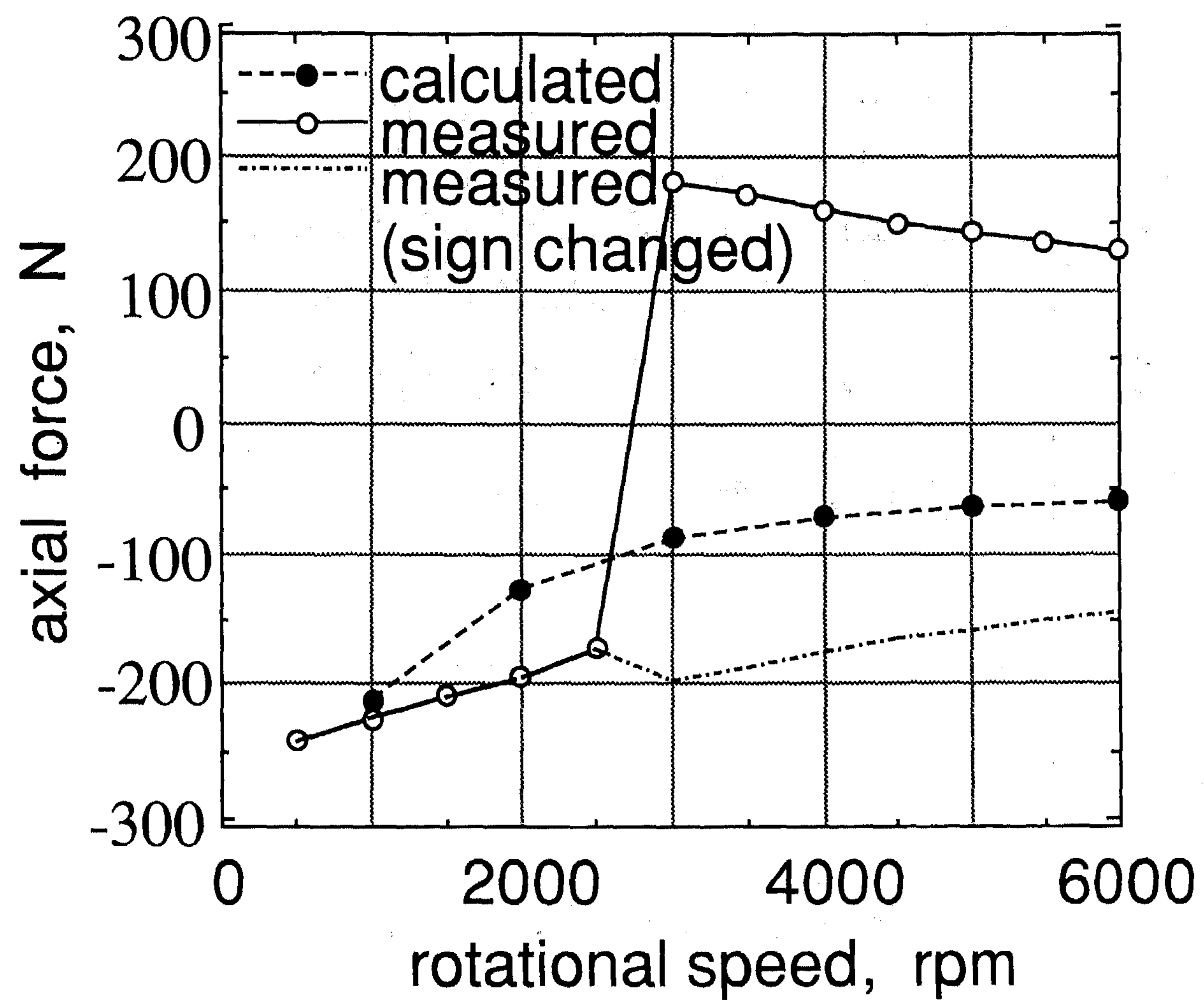


Fig. 11 Axial Force and Speed

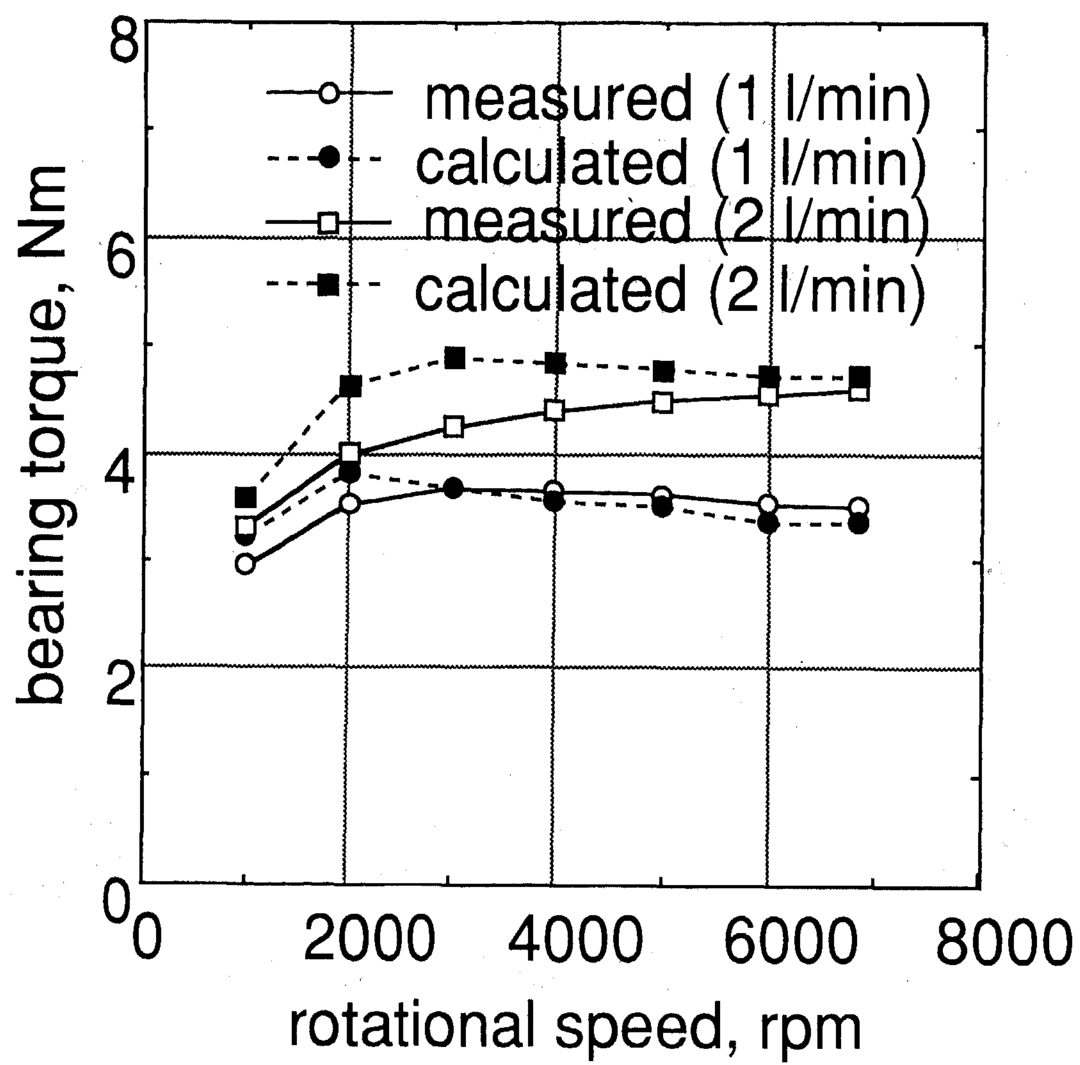


Fig. 12 Bearing Torque and Speed
(tapered roller bearing)

Table 1 BRAIN

	PGM Name	Bearing Type
BRAIN	BALTAC	deep groove, angular, and thrust ball bearings
	B4PTAC	three point and four point contact ball bearings
	CYLTAC	cylindrical roller bearings
	TNBTAC	thrust needle bearings
	TAPTAC	tapered roller bearings
	SELTAC	self-aligning roller bearings

Table 2 Experimental Condition

	case A	case B
Radial Load (N)	0	3000
Axial Load (N)	3000	1000
Speed (rpm)	1000 - 10000	
Lubricant	turbine oil (32 mm^2/s at 40 degC) oil feeding rate 100ml/min(25 degC)	

Table 3 Roller Bearing Specifications

Roller	ϕ 5 x 35, z = 9
Shaft Diameter	ϕ 25
Load	radial load Fr = 5000N moment M = 5 Nm
Speed	500 - 6000 rpm
Oil Temperature	95 degC

Table 4 Tapered Roller Bearing Specifications

Bearing	HR32217J
Bore Diameter	ϕ 85 mm
Outside Diameter	ϕ 150 mm
Width	38.5 mm
Roller Number	19

970589

Development of Dynamic Simulation Models of Seated Reaching Motions While Driving

Xudong Zhang and Don B. Chaffin
University of Michigan

Deborah Thompson
Chrysler Corporation

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A research effort was initiated to establish an empirical data base and to develop predictive models of normal human in-vehicle seated reaching motions while driving. A driving simulator was built, in which a variety of targets were positioned at typical locations a driver would possibly reach. Reaching motions towards these targets were performed by demographically representative subjects and measured by a state-of-the-art motion analysis system. This paper describes the experiment conducted to collect the movement data, and the new techniques that are being developed to process, analyze, and model the data. Some initial findings regarding the role of torso assistive motion, the effect of speed used in completing a motion on multi-segment dynamic postures, and illustrative results from kinematic modeling are presented.

INTRODUCTION

The proliferation of computer technology has been renovating conventional ergonomics. The design as well as the analysis of human-machine systems such as a driver-vehicle system is becoming more and more computerized. As this trend continues, the development of human models that realistically predict how people behave and interact with the system is of great importance. Not only can such models facilitate quantitative, easy-to-conduct ergonomic analysis, but also can economize the design by providing recommendations and changes early in the process prior to costly prototyping.

Most of the existing human models that can be used to facilitate vehicle design and ergonomic assessment are static biomechanical models [1,2,3,4]. Although there are a limited number of computer-aided design programs that feature dynamic human models (e.g., JACK), these models usually are not based on empirical studies of preferred human movements. Rather, they use overly simplified heuristics or rough estimates of human kinematics, and therefore the realism of rendered human dynamic motions remains untested and questionable. There is a pressing need for dynamic human models that are empirically validated for a typical range of driving tasks and for people of varied anthropometry, gender, and age.

Recognizing this need as well as the emerging new techniques to measure, analyze, and model complex 3D human movements, the Center for Ergonomics at the University of Michigan and Chrysler Corporation have initiated a collaborative research project. The first phase of this project has three primary goals: 1) to establish an empirical database of in-vehicle human reaching movements (concurrent with driving); 2) to derive models that accurately predict dynamic, multi-segment postural profiles during such movements; 3) to develop a computer-aided design (CAD) software tool which generates graphical animation of normal human movements predicted by the database and the models. The specific purpose of this paper is to describe the experiment conducted to acquire three-dimensional seated reaching movement data and the techniques to analyze and model the resulting large scale, complex data. Some preliminary findings and results achieved are also presented as illustrations.

EXPERIMENTAL PROCEDURES

SUBJECTS - Thirty-eight subjects including 20 males and 18 females, demographically representative in terms of anthropometry and age, served as subjects and were paid for their participation. Table 1 provides some descriptive statistics for the age and gross anthropometry of these subjects.

REACHING-WHILE-DRIVING TASKS - A driving simulator has been developed at the Center for Ergonomics of the University of Michigan. It consists of a mock-up vehicle and a graphically rendered driving scene on a big screen monitor integrated with the primary controls (e.g., steering wheel, acceleration pedal, etc.) of the vehicle. As illustrated in Figure 1, forty-one targets grouped into six areas that emulate the typical location zones that a driver would normally reach to, are presented in the simulated vehicle. These six areas include left side corner (door or mirror control), headliner, shift console, instrument panel (radio &

climate control), glove box, and passenger side door. The number of targets within these areas was 3, 5, 5, 18, 8, and 2, respectively. Note that the instrument panel and glove box areas can be moved in and out, which doubles the number of targets in these two areas and also somehow emulates two different size vehicles. The three targets placed at the left side corner are used to evaluate the left-right motion symmetry.

Table 1.
Descriptive Statistics of Age and Gross Anthropometry Information of the 38 Subjects

Attribute	Mean	Std. Dev.	Minimum	Maximum
Age	35.3	15.4	19	72
Stature (cm)	171.3	11.1	152.8	201.2
Weight (lb)	155.6	35.8	100	293

While the subjects were performing driving tasks, they were intermittently signaled to reach towards the 41 targets in a discrete, natural, preferred manner. Each target was reached twice, and the order of 82 reaching movements was generally randomized (reaches to the in and out positions of the instrument panel and glove box area targets, and to the three left side targets, were blocked but randomized within the blocks). Instructions and a training session were provided to the subjects prior to the reaching task performance that was actually measured. The seat position was adjusted as individually preferred at the beginning of the experiment but remained fixed thereafter.

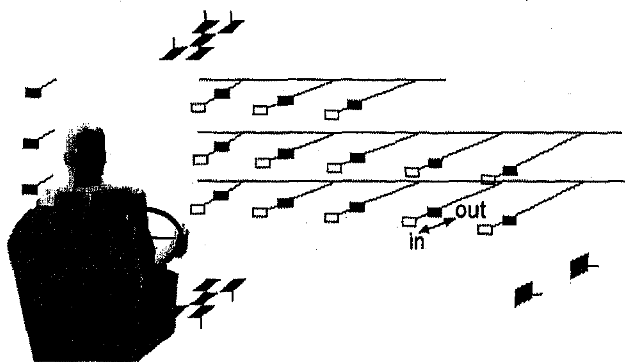


Figure 1. A simulated vehicle where 41 targets are positioned in six areas: left side corner, headliner, center console, instrument panel, glove box, and passenger side door. The instrument panel and glove box area targets can be moved in and out to create twice as many targets.

DATA ACQUISITION - Eleven spherical reflective markers were placed on palpable body landmarks or on rigid fixtures attached to the body identifying the right hand, right wrist, right elbow, right & left shoulder, head, suprasternale, right & left anterior-superior iliac spine (ASIS). The three markers on the right hand and one on the right elbow were moved onto the left hand and elbow during left side reaching movements. As illustrated in Figure 2, these markers formed a virtual stick-figure representation of the biomechanical system of our particular concern. This representation is also often referred to as a linkage representation, with the assumption of each segment being a rigid link. The kinematics of this linkage describe the dynamic postures of the torso, right upper extremity, neck and head. An opto-electronic motion analysis system (MacReflex™, Qualisys Inc.) with four cameras was employed to measure the movements of the markers at a sampling frame rate of 25 Hz. The system has the capability of tracking the captured motions as a succession of 2D images (four images corresponding to four cameras for each instant time frame) and then delivering the 3D marker Cartesian coordinates of in a time series.

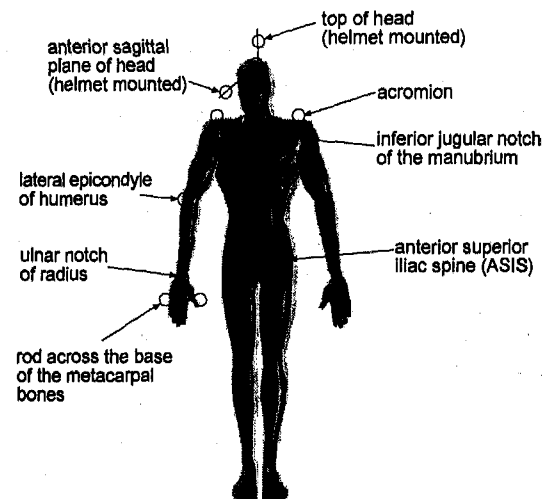


Figure 2. The placement of 11 reflective markers which describe the kinematics of the torso and right upper extremity. The three markers on the right hand and one on the right elbow were moved onto the left hand and elbow during left side reaching movements.

ANALYSIS & MODELING

JOINT ANGLE DERIVATION - In order to enhance the realism of our linkage representation and to better describe the kinematics using more truthful angles, the coordinates of surface markers were transformed into those of internal joint centers for the shoulder, elbow, and wrist joints. There was no existing kinematic algorithm readily applicable to carry out this transformation for our specific marker placement situation, since most algorithms require at least three markers on each body segment. A new, experimentally confirmed, heuristic method was developed to accomplish the transformation [5,6]. Based on the enhanced linkage, a comprehensive set of joint angles was defined [7] and computed. These joint angles include absolute angles needed for completely determining a posture (i.e., linkage configuration) and some relative angles (e.g., elbow included angle) that are more revealing and easier to visualize.

STATISTICAL MODELING - It is intended to present our empirical data base in the form of a robust population statistical model which predicts movement profiles for in-vehicle reaching motions. This model uses the derived angular profiles as dependent variables, and the target location and subject individual attributes as independent variables. This major statistical modeling effort takes place at the Statistics Department of The University of Michigan. A new modeling approach based on regression analysis for functional responses has been established [8].

In addition to the principal statistical effort of generating a robust population model, a multitude of statistical analyses have been conducted or are being explored, using subsets of the collected data, to address specific issues. Such issues include the effects of target location on the amount as well as the direction of torso assistive motion, the effect of the speed used in completing a reaching movement on dynamic postures, the effects of individual anthropometry on the preferred seat position reflected by the ASIS location, etc. The results from these analyses can either aid in developing the general statistical dynamic posture prediction model or be used singly to provide useful design information.

KINEMATIC MODELING - In parallel with the statistical modeling effort, kinematic modeling is also being undertaken. The linkage representation described earlier is simplified as a four-segment chain with seven degrees of freedom to kinematically represent the upper torso and right upper extremity. Figure 3 portrays this simplified linkage which consists of the torso, right clavicle, right upper arm, and right forearm (including the hand which is treated as a rigid extension of the forearm). The degrees of freedom possessed by this linkage account for torso flexion/extension, torso medial/lateral lending, torso twisting (modeled as clavicle rotation), shoulder flexion/extension, shoulder ad/abduction, humeral rotation, and elbow flexion/extension.

What posture prediction poses is a classic problem of solving kinematic redundancy in biomechanics. That is, given an instantaneous hand location, how the configuration of the linkage is determined from an infinite number of possibilities to position the body segments? Optimization has been a viable tool for solving this redundancy problem, based on the theory that an inherent optimal strategy is followed by people in posture control. Previous three-dimensional models concerning seated reaching postures were static, relying on tedious moment-to-moment optimization procedures [3,9,10], and therefore have limited practicality. We propose a unique velocity-domain approach for solving the kinematic redundancy problem in a dynamic fashion. We view it as a problem of determining how changes in hand coordinates are allocated among changes in degrees of freedom, and theorize that people employ an optimal allocation strategy to minimize the total weighted energy-type effort in dynamic posture control. Weighting parameters assigned to individual body segments quantify the "cost" of their participation in a movement. More detailed descriptions of this approach can be found in Zhang and Chaffin [11].

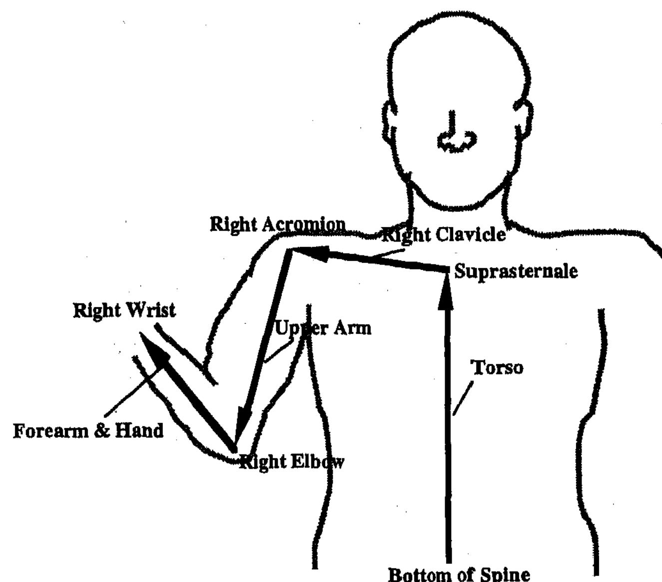


Figure 3. A four-segment seven-degree-of-freedom linkage to model the upper torso and right upper extremity.

RESULTS

STATISTICAL ANALYSES - The primary statistical analysis that promises to deliver a robust statistical model predicting the behavior of in-vehicle reaching movements is ongoing. A number of "satellite" analyses however have been concluded. One analysis which examined the movement speed effect on postural profiles revealed an insensitivity of instantaneous postures to speed variation—a difference in the speed mode used to complete a movement merely had a scaling effect on the angle displacement as well as velocity profiles. This analysis therefore suggested excluding the speed as a modeling element. Another analysis showed that the torso motions towards the targets located in the instrument panel and glove box areas have interesting and predictable characteristics. The torso was found to be a very adept prime mover in reaching targets located within or beyond the arm length. The amount and direction of torso motion can be well predicted ($R^2 = 0.85$-0.95) by linear regression models using independent variables derived from target location and distance with respect to a driver. This latter analysis helped convince ourselves that there is some inherent strategic planning involved in modeling of human movements.

OPTIMIZATION-BASED INVERSE KINEMATIC MODEL - The proposed model is being subjected to empirical tests using a subset of acquired movement data. Preliminary results indicated an averaged (across different movements and subjects) mean error (across the entire movement and seven joint angles) of 3 degrees in the predicted angular profiles with respect to the measured one. Such a level of accuracy compares favorably with what was achieved by previous posture prediction models most of which were two-dimensional and/or static. Figure 4 is an illustration of

predicted angular profiles versus the measured ones for a reaching movement towards a target in the glove box area. The mean error of prediction (across the entire motion and 7 joint angles) was 4.5 degrees for that particular trial.

It was also found that the weighting parameters (which quantify the allocation strategy among the body segments) demonstrated some difference between, but also congruence within, types of reaching movements driven by different in-vehicle target locations. Future efforts will continue in validating the model using more experimental data. Particularly, the emphasis will be placed on synthesizing the model parameters from one set of subjects and applying them to another set of subjects. Some statistical means will also be needed to summarize these parameters and present them for population prediction purposes.

CONCLUSION

Researchers and practitioners in the field of ergonomics are cognizant of the trend toward computerized ergonomic analyses to aid the design of increasingly complex systems. However, the lack of a realistic, three-dimensional, dynamic model of human movements has been the roadblock to the quest for accurate and comprehensive computer-aided ergonomics. Our research effort has made an initial but important step toward the development of such models. To our knowledge, it is the first major effort of studying and modeling three-dimensional, multi-segment human reaching movements under a realistic context. Based on the encouraging results we have harvested so far, we are fully confident of accomplishing the stated goals. We are also prepared for expanding the current effort to a variety of system configurations and human physical tasks other than driving.

Our effort is unique in that we are able to enjoy the advantage of having two concurrent modeling approaches. Considerable benefit has been and will continue to be gained from the cross-fertilization between statistical and kinematic modeling approaches.

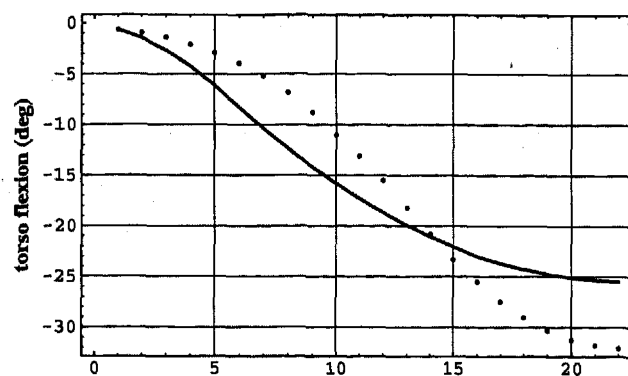
Beyond the practical motivation of developing an advanced dynamic simulation tool, strategic issues related to human motion control also underscore the theoretical significance of studying and modeling human reaching movements in a systematic fashion. It is believed that the findings gained from this effort will contribute to advancing knowledge in that respect as well.

ACKNOWLEDGMENTS

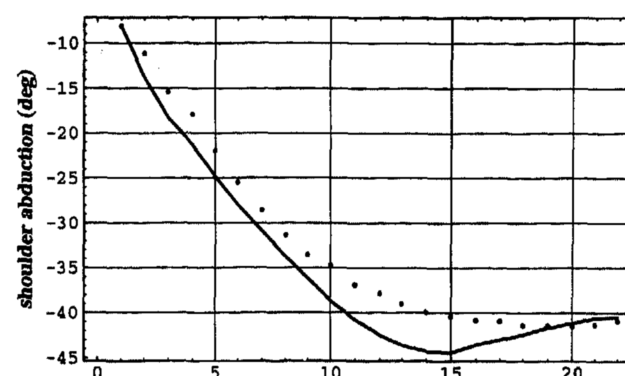
This project is supported by Chrysler Corporation Challenge Fund. Other active members of the project include Jim Foulke, Ulrich Raschke, Chuck Woolley of the Center for Ergonomics, and Julian Faraway of Statistics Department at the University of Michigan. Pat Terrell provided assistance in rendering graphical illustrations.

REFERENCES

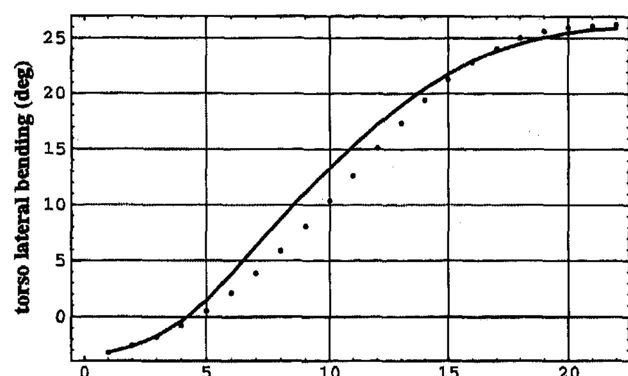
1. Hubbard, R.P., Haas, W.A., Boughner, R.L., Canole, R.A., and Bush, N.J. (1993). New Biomechanical Models for Automobile Seat Design. SAE Technical Paper 930110, Society of Automobile Engineers, Inc., Warrendale, PA.
2. Case, K.B., Porter, J.M., and Bonney, M.C. (1986). SAMMIE: A Computer Aided Design Tool for Ergonomists. Proceedings of the 30th Human Factors Society Annual Meeting, Dayton, OH.
3. Ryan, P.W. (1970). Cockpit Geometry Evaluation (Joint Army-Navy Aircraft Instrument Research Report 700201). The Boeing Company, Seattle, Washington.
4. Peacock, B.J. and Karwowski, W. (Eds.) (1993). Automotive Ergonomics. Taylor & Francis, London.
5. Nussbaum, M.A., Zhang, X., Chaffin, D.B., Stump, B.S., and Raschke, U. (1996). A Reduced Surface Marker Set for Upper Limb Kinematics: Heuristics and Optimization. Proceedings of the American Society of Biomechanics 20th Annual Meeting, Atlanta, GA.
6. Zhang, X., Nussbaum, M.A., and Chaffin, D.B. (1996). Calibrating Instantaneous Helical Axes to Identify the Elbow and Shoulder Joints for Movement Studies. Proceedings of the American Society of Biomechanics 20th Annual Meeting, Atlanta, GA.
7. Center for Ergonomics (1996). Angle Definitions for Seated Operator Motion Simulation. Technical Report, The University of Michigan, Ann Arbor, MI.
8. Faraway, J.J. (1995). Regression Analysis for a Functional Response. Technical Report #263, Department of Statistics, The University of Michigan, Ann Arbor, MI.
9. Kilpatrick, K.E. (1970). Models for the Design of Manual Work Station. Ph.D. Dissertation, The University of Michigan, Ann Arbor, MI.
10. Jung, E.S., Kee, D., and Chung, M.K. (1995). Upper Body Reach Posture Prediction for Ergonomics Evaluation. International Journal of Industrial Ergonomics, 16, 95-107.
11. Zhang, X. and Chaffin, D.B. (1996). An Optimization-Based Differential Inverse Kinematics Approach for Modeling Three-Dimensional Dynamic Postures during Seated Reaching Movements. Technical Report (TR 96-15), Dept. of Industrial & Operations Engineering, The University of Michigan, Ann Arbor, MI.



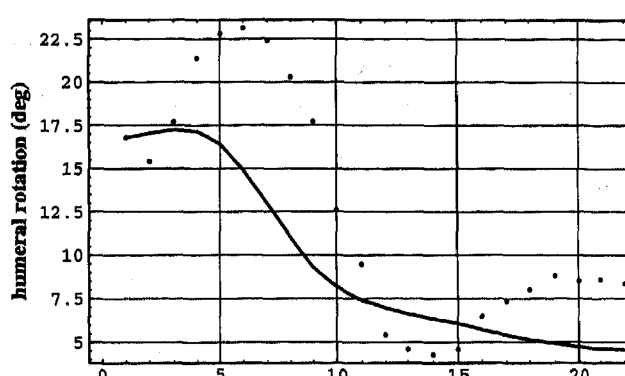
(a)



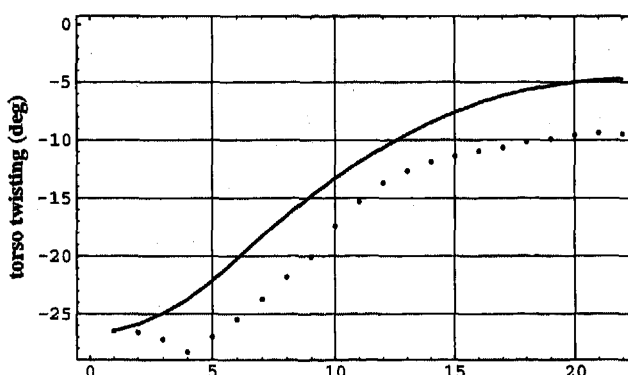
(e)



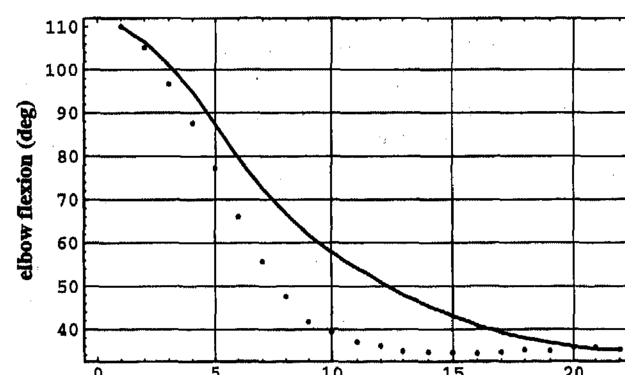
(b)



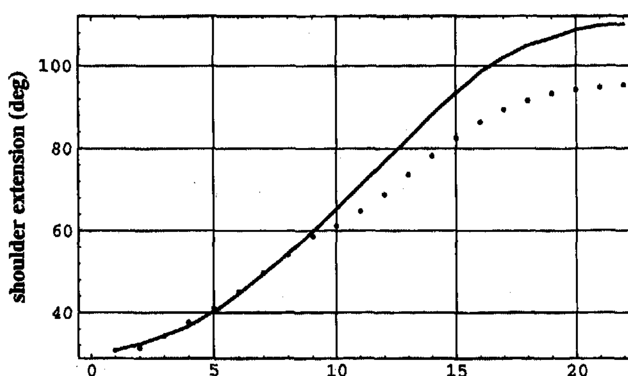
(f)



(c)



(g)



(d)

Figure 4. Predicted angular profiles (solid) by the proposed kinematic model versus the measured ones (dotted) for a reach to the glove box area: (a) torso flexion; (b) torso lateral bending; (c) torso twisting; (d) shoulder extension; (e) shoulder abduction; (f) humeral rotation; (g) elbow flexion. The mean error of prediction (across the entire motion and 7 angles) for this particular trial was 4.5 degrees.

970590

Development of Human Back Contours for Automobile Seat Design

Barry L. Frost III, Robert P. Hubbard, and Robert L. Boughner
Michigan State University

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Driver and passenger comfort, as related to automotive seats, is a growing issue in the automotive industry. As this trend continues, automotive seat designers and developers are generating a greater need for more anthropometrically accurate tools to aid them in their work.

One tool being developed is the JOHN software program that utilizes three-dimensional solid objects to represent humans in seated postures. Contours have been developed to represent the outside skin surfaces of three different body types in a variety of postures in the sagittal plane. These body types include: the small female, the average male, and the large male.

INTRODUCTION

Comfort in automotive seats is an issue that has increased in interest over the past decade. Seats are being contoured to promote spinal curvature and designed with variable contours to promote movement and variation of spinal curvature and posture. Historically, seat designers have created seats mainly with templates on paper. The use of templates has become increasingly difficult in the design of highly contoured seats. Testing the seat design for comfort could not be done until a physical prototype was built. With the advent of computers and three-dimensional solid modeling programs, seat designers developed the ability to generate surface contours. If human back contours can be represented with computers, seats that fit people's contours can be designed. While not sufficient for comfort, the ability to fit seats to human contours is one of the necessary steps for creating comfortable seats. The use of computers in seat design also reduces the need for costly physical prototypes.

The following paper discusses several additions that have been made by the Michigan State University Biomechanical Design Laboratory to the JOHN software program, developed using the I-DEAS solid modeling software from Structural Dynamics Research Corporation. The first addition was the improvement of the existing JOHN contours. The software program was then rewritten to allow contour generation for any amount of movement of the lumbar spine in the sagittal plane. New contours for the large male and small female were also added to the program. The JOHN program was originally introduced at the Society of Automotive Engineers International Congress in 1992 by Hubbard, et al., as paper no. 930110, titled "New Biomechanical Models for Automobile Seat Design" [1]. This paper will begin with a historical overview of the JOHN model examining four previous key developments. Each of the three new additions will then be discussed, followed by a conclusion.

BACKGROUND

In the early 1980's, the University of Michigan Transportation Research Institute (UMTRI) conducted an anthropometric study for crash dummy development by determining how the occupant is positioned within the automobile [2]. In their study, UMTRI placed people in a simplified model of the driver's side of an automobile and used a combination of seat surface castings and photographs of anthropometric landmarks on the subjects to determine exterior skin contour points. Anthropometric data were collected for a wide variety of individuals and they were classified and averaged into three categories: the mid-sized male [3], the small female, and the large male [4]. An example of the contour data collected for the average adult male with external and internal landmarks is shown in Figure 1.

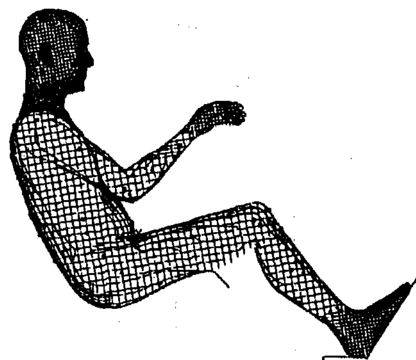


Figure 1. UMTRI Average Male Contour

In the late 1980's, Michigan State University (MSU) began developing computer models of the

human musculo-skeletal system for seat design [1]. The first work at MSU was done by Haas [5], who developed a two-dimensional computer model, called JOHN, to determine the relative position of the head, thorax, and pelvis of an average human male, as shown in Figure 2.

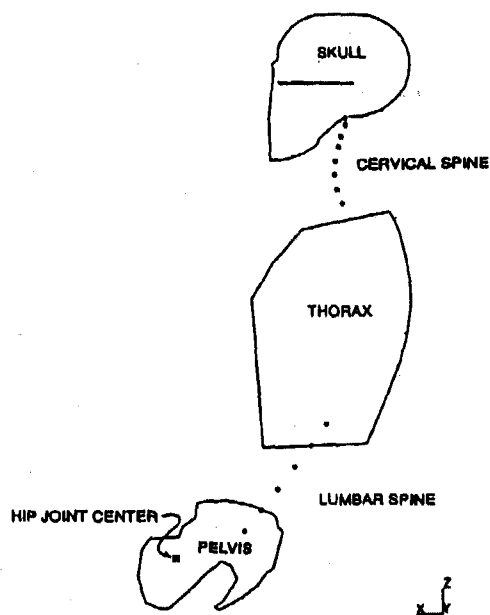


Figure 2. Two-Dimensional JOHN Computer Model

The JOHN software program utilizes solid objects to represent humans in seated postures. The head, thorax, and pelvis segments in the Haas model were connected by a series of points depicting joint centers in the cervical, thoracic, and lumbar spinal regions.

During this time at MSU, two anthropometric variables were defined for the JOHN model. Torso Recline Angle (TRA) is defined as the angle, measured from vertical, of the line that passes from the T_{12}/L_1 spinal joint center to the L_5/S_1 spinal joint center, as seen in Figure 3. The second is an internal anthropometric variable known as Total Lumbar Curvature (TLC). A back with a flat lumbar section where all the spinal joint centers are collinear is defined as TLC=0°, Figure 4. Given a reference line on the pelvis or ribcage that is parallel to the TLC=0° line, TLC can then be determined for any amount of lumbar curvature by measuring the angle between the reference line and a line passing between the T_{12}/L_1 and L_5/S_1 spinal joint centers, Figure 5. TLC is defined as positive when the back of the lumbar spine is concave.

The next study at MSU, by Boughner [6], took the two-dimensional program of the adult male developed by Haas and created a three-dimensional solid model program, Figure 6. Boughner created models of the entire major skeletal structure using data

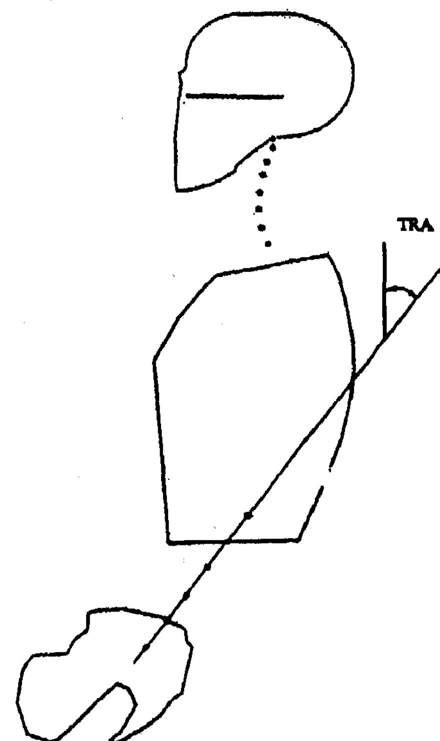


Figure 3. Torso Recline Angle (TRA)

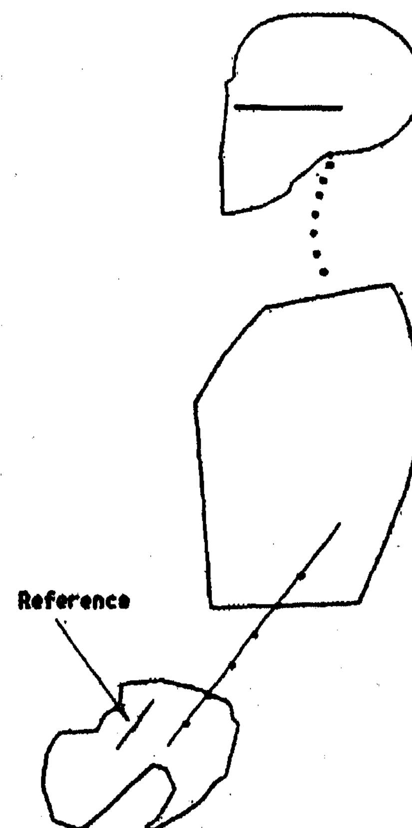


Figure 4. Total Lumbar Curvature = 0°

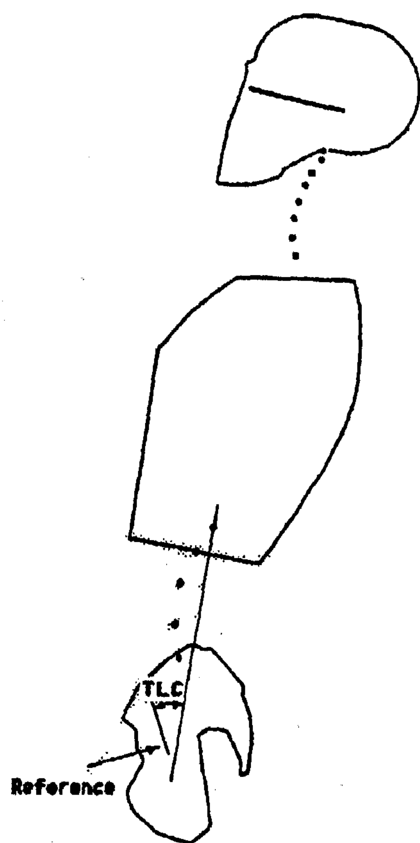


Figure 5. Total Lumbar Curvature > 0°

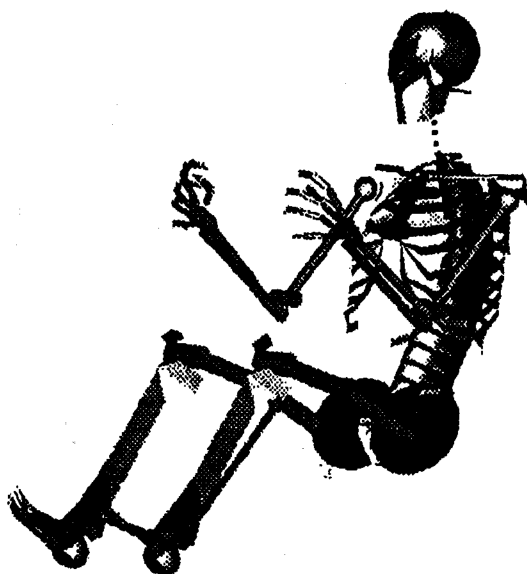


Figure 6. Three-Dimensional JOHN Computer Model

from several anatomy texts [6,7,8] and data from the UMTRI study [3]. Boughner further refined his program by adding the ability to model leg muscle geometry and calculate hamstring length for the JOHN model.

In another investigation, Bush [9] developed models of the *erector spinae* muscles, the muscle groups that lay directly to either side of the spinal column and external skin contours for the back of the JOHN model, Figure 7. Tissue thickness of the model was determined by examining cross-sections of cadavers presented by Koritké and Sick [10]. Bush developed skin contours for the JOHN model at discrete values of Total Lumbar Curvature (TLC), specifically at values of -10°, 0°, 10°, 20°, 30°, and 40°. Because three-dimensional renderings of the UMTRI body contours were not available at that time, Bush was limited to examining these two contours at specific points only and was not able to perform an in-depth comparison between his contours and the UMTRI contours.

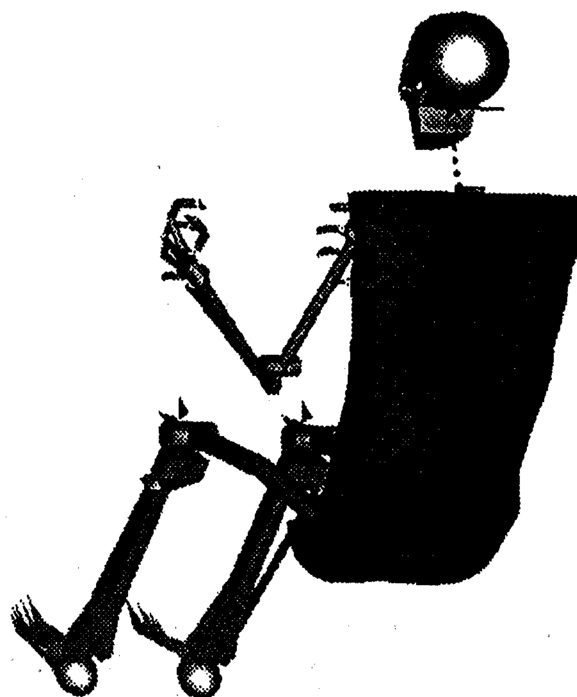


Figure 7. JOHN Model with Bush Contour [9]

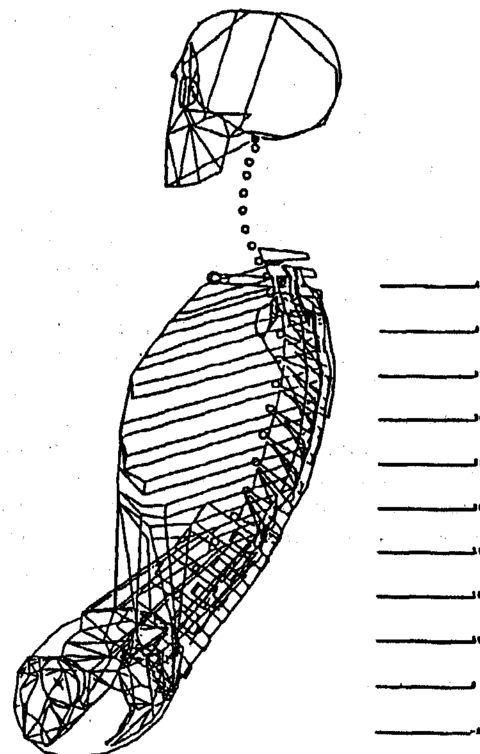


Figure 8. JOHN Model with Bush Cross-Sections [9]

METHODS

REDEFINITION OF THE JOHN CONTOUR

In his research, Bush performed a comparison between his contours and the contours developed by UMTRI [3]. Even though the body surface landmark data from the UMTRI study had been available for some time, without computer models Bush was limited to analyzing his data in comparison to UMTRI's at a few surface landmarks. More recently, complete three-dimensional UMTRI contours became available as computer files. With the new contour files, the complete UMTRI contour data could be analyzed on the computer at every point.

To analyze the differences between the contour created by Bush and the UMTRI contour, both contours were placed over the JOHN musculo-skeletal system. The UMTRI contour represents an average adult male in a single posture with a flat lower back, or straight lumbar spine (TLC = 0°). Therefore, the UMTRI contour was compared with Bush's contour for a TLC = 0°.

Cross-sections of the contours in the back region were created through each of the vertebrae down the spinal column, seen in Figure 9. Each cross-sectional plane was defined by the mid-plane between the top and bottom of each vertebrae. In the lower buttock region, the cross-sectional planes were defined as the y-z plane, also known as the vertical transverse plane, where y represents the width dimension and z is the vertical height dimension. The y-z planes were created at distances from the H-point in a range from 50 mm to -70 mm, at 20 mm increments from front to rear. In the upper buttock region the cross-sectional planes were defined by rotating the vertical transverse plane 45° about the y axis. Three sections were made at distances, measured perpendicular to the plane from the H-point, of 50, 70 and 90 mm in the rearward direction.

Once the new vertebral cross-sections were created, the Bush contour line was compared to the UMTRI contour line at each cross-sectional level. In the upper body, sections of the back varied no more than 14 millimeters in the posterior direction and the differences between the contours in the lateral direction varied between 3 to 39 millimeters. In the lower body, or buttock region, the Bush contour and the UMTRI contour had differences in the inferior direction ranging between 17 and 26 millimeters and lateral differences varying between 14 to 47 millimeters.

New contour lines were then drawn at each vertebral level. While the UMTRI contour [3] was used as a guide to the overall size of the new contour, an attempt was made to keep the muscle and bone definition shown in Bush's contour [9]. In sections where there was a large difference between the positions of the Bush and UMTRI lines, the new contour was positioned closer to the UMTRI contour while attempting to maintain the basic shape of the Bush contour.

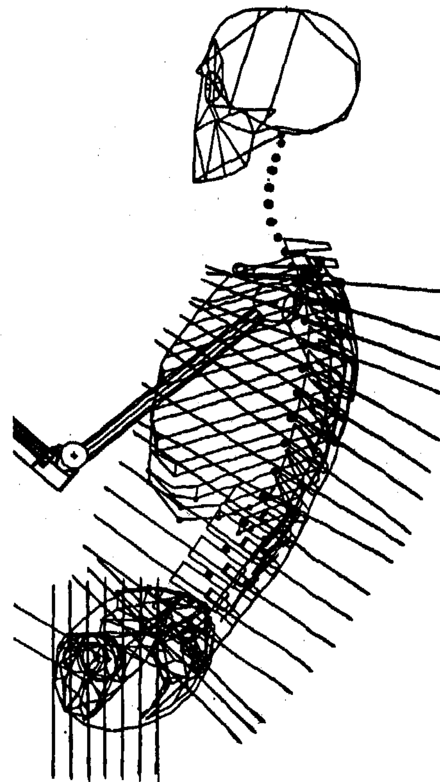


Figure 9. JOHN Model with Cross-Section Locations

The approach for representing body contours had been to try to model JOHN as if the model had no deformation of the skin surface, or zero forces acting upon the body. The goal of undeformable contours was difficult to obtain because every model examined in the development of JOHN was in a deflected, seated posture. Therefore, several approximations were used to move closer to the goal. Allowances were made due to UMTRI's method of data collection including using casting tape, which spanned across concave contours of the back and buttocks region. Allowances were also made for the averaging technique used by UMTRI to obtain the final contour. The averaging technique used by UMTRI caused several concave regions to become less severe in their concavity.

Once the new contours were defined, point coordinates along each contour line were calculated according to the reference axes of the JOHN system. Each cross-sectional line was defined by the same number of points to create fully defined vertical lines down the length of the model. The coordinates were then entered into the JOHN model. Splines were calculated for each cross-section and down each vertical line to create a wireframe mesh, Figure 10.

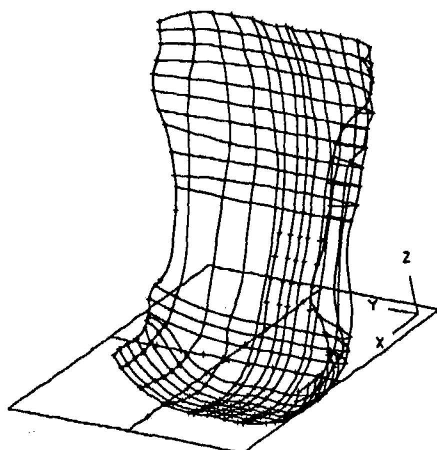


Figure 10. JOHN Contour Wireframe Mesh

A solid object skin surface was developed from the wireframe mesh. Due to the constraints of our computing system, I-DEAS would only allow approximately 20 splines to be chosen at one time in the generation of skin contours. Not all of the splines could be chosen to create the solid. Therefore, all of the vertical splines were chosen and cross-sectional splines were only chosen at each of the following levels:

1. Third thoracic vertebrae level (T_3),
2. Eleventh thoracic vertebrae level (T_{11}),
3. Third contour line below the fifth lumbar vertebrae level (UB_3), and
4. Seventh or lowest contour line in the lower buttocks region (LB_7).

The four cross-sectional splines were chosen as they are nearly equidistant along the length of the contour. The final new skin contour can be seen in Figure 11.

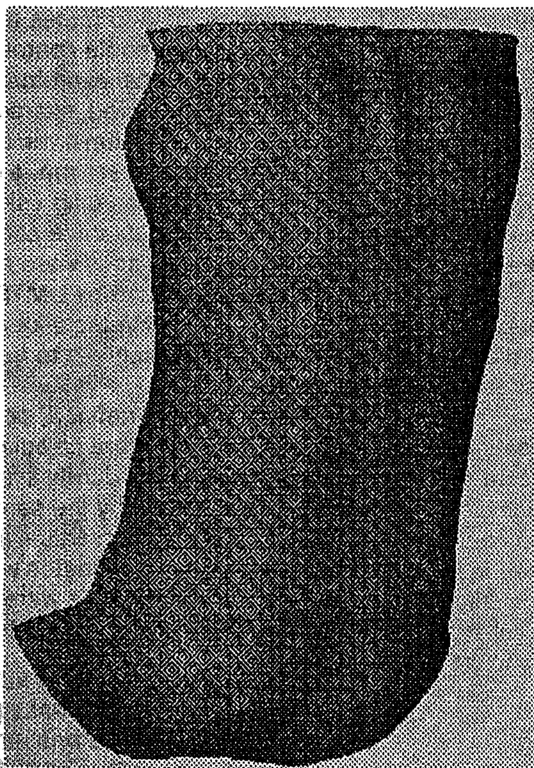


Figure 11. The New JOHN Contour

Even though more information on human contours are being developed by current research at MSU, the new contour described above is the best contour possible with the currently available data for TLC = 0°. Once the contour data comes available from MSU and other sources, the JOHN contours can be further validated and refined. The importance of having highly representative contours can not be overstated. By using better contours, seat designers can design seats to have a better fit.

CREATION OF "VARIABLE" CONTOURS

The second objective was to design and implement a method to create contours at any value of TLC and TRA. In the original development of contours for the JOHN average adult male by Bush [9], contours were created for six different values of TLC. While keeping the amount of TRA at a constant 35°, the JOHN model was placed at angles of TLC equal to -10°, 10°, 20°, 30°, and 40°. At each TLC value, Bush created horizontal cross-sections as he had done for the 0° TLC model. Once again, contour lines had been drawn over the modeled muscles and bones with an offset thickness.

A limitation of the Bush contours was that seat designers only had the ability to evaluate seats with those six contours. If a designer wanted to design a seat for a TLC value other than one of the 10° increments, they would not be able to set the musculo-skeletal model at that TLC value, but would rather have to add an overlying contour at a 10° increment nearest the TLC value.

By examination of the ribcage and its skin surface, it was assumed that points on the skin surface over the thoracic and pelvic regions tend to remain in place relative to the rib cage and pelvis, respectively, over a wide range of motions in the sagittal plane. For this reason, points were attached to the rib cage, pelvis, spine, and muscles of the JOHN musculo-skeletal model. These skin points would then move with those objects as JOHN's posture changed to different values of TLC. Once the designer has set the musculo-skeletal system in a preferred posture, points can then be splined through and the contours for that exact selected posture would be created in the same manner as the 0° TLC contour. This process allows for both a systematic and repeatable method for contour generation.

A new articulation program was developed based upon Boughner's program [6], in which the musculo-skeletal objects and skin surface points articulate about the lumbar spinal joint centers. Once this motion is completed, the points are splined and a solid contour is generated from the splines. The program then rotates the musculo-skeletal system, with the new contour, to the specified TRA and aligns the *ischial tuberosities* to their original position. Leg and arm motion is then manipulated at the request of the user.

By allowing the contours to articulate with the musculo-skeletal system, contours can be made for any degree of TLC. The seat designer no longer has to

design a seat for occupants only at TLC values of -10, 0, 10, 20, 30, and 40 degrees.

SMALL FEMALE AND LARGE MALE CONTOURS

Even though JOHN is a useful representation of the average U.S. man, designers have the need to determine if seats will also accommodate people who are tall or short in stature, or large or small in corpulence. Due to that need, two more models were developed by MSU in addition to the JOHN model. In order to model tall, heavy men, the JERRY model was developed to represent the 95th percentile male in stature and weight. To represent small women, a 5th percentile (stature and weight) female model was developed called JANE. These models were based on studies by UMTRI [4] that established skeletal landmark locations and body contours for tall, heavy men and short, light women.

The large male and small female contours were developed by scaling from the average male contour. By using the average male model as the basis, the small female contour and the large male contour would then be related in a known way, creating a basis for future models different from the 95th male, 50th male and 5th female sizes.

Scaling factors were developed by examining the differences between the dimensions for the UMTRI average male and the dimensions for the UMTRI small female and large male. By examining the distances between different skeletal landmarks and comparing the related distances between the three models, scaling factors were established. The torsos of the UMTRI contours were separated into three segments: thoracic, lumbar and pelvic with different scaling factors chosen for each segment. This was done because of differences in the relative proportions of these three body segments between the male and female.

Scaling factors for the thoracic region were determined by examining the distances between the following points defined by UMTRI [2], seen in Figure 12:

1. *Mesosternale* to the T_4 *spinous process*
2. Lowest (inferior) point on the right 10^{th} Rib to the lowest (inferior) point on the left 10^{th} Rib
3. C_7 *spinous process* to the T_{12} *spinous process*

These distances were chosen because they lie mainly in the direction of each of the coordinate directions. The scaling factors between the points were calculated as shown below in Table 1. The scaling factors in each of the coordinate directions were calculated by dividing the distances for the small female and the large male by the related distance of the average male.

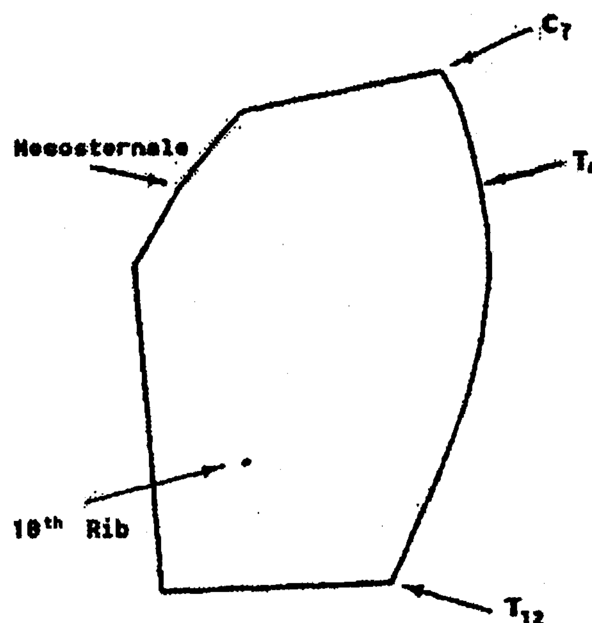


Figure 12. Thoracic Scaling Points

Table 1. Thoracic Scaling Factors

	Mesostem. - T_4 sp. pro.	Right 10^{th} rib - left 10^{th} rib	C_7 sp. pro. - T_{12} sp. pro.
JOHN Distance (mm)	195.1	312	343.6
JANE Scaling Factor	0.77	0.83	0.95
JERRY Scaling Factor	1.1	1.24	1.1

Scaling factors for the pelvic region were determined by examining the distances between the following points, seen in Figure 13:

1. L_5 *spinous process* to the center of the line between the right and left Anterior Superior Iliac Spine (ASIS)
2. Right ASIS to the left ASIS
3. *Iliocristale* to the Inferior Tuberosity projection on the mid-sagittal plane

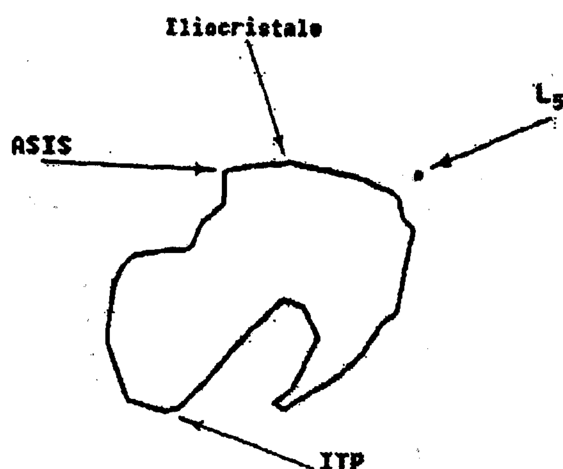


Figure 13. Pelvic Scaling Points

Again these dimensions were chosen due to their similarities in direction with the coordinate axes. Table 2 shows the scaling factors in the pelvic region.

Table 2. Pelvic Scaling Factors

	L_5 sp. pro. - ASIS	Right ASIS - left ASIS	Iliocrist.- Inf. Tub. pro.
JOHN Distance (mm)	164.6	222	231.7
JANE Scaling Factor	0.89	0.93	0.89
JERRY Scaling Factor	1.15	1.09	1.15

Once the scaling factors had been established for both the thoracic and the pelvic regions, a systematic approach was developed to place the scaled regions in the proper orientation. It was apparent that there were postural differences between the UMTRI contours of the small female, the average male, and the large male. These differences in posture were determined by scaling the torso segments of JOHN. The scaled torso segments were then fit to the UMTRI surface landmarks of the large male and small female. The UMTRI contours showed that large men tend to slouch more than average-sized men with a -3° TLC difference between the two. The small female contour showed a 15° more erect posture (+15° TLC) relative to the average male. In order for the scaled JOHN contours to be in the orientation of the UMTRI contours, the JOHN contour was first articulated and then scaled to match the posture of the UMTRI contours.

When an object or point is to be scaled in I-DEAS, two parameters are required: the scaling factor and the point from which the object or point will be scaled from. When an object or point is scaled from any single point, the scaling procedure will cause all other points to move. Because all of the points in the JOHN model were scaled from the origin, located at the H-point, each of the scaled regions required a translation to move them back into the correct position.

In order to realign the newly scaled pelvic regions, the x value of the L_5 *spinous process* point and the z value of the Inferior Tuberosity projection point were each compared to their related values in each of the three models. These values were used because it was assumed that the height of the seat surface and the depth of the seat back were well defined due to the methods used in the UMTRI study [2]. The skin points were translated to align the L_5 *spinous process* and the Inferior Tuberosity projection.

To position the newly scaled thoracic regions, it was assumed that because the C_7 *spinous process* is a easily palpated point, the location was accurately measured by UMTRI. This point was then compared to the newly scaled C_7 spinal point to determine translation values.

Scaling factors for the lumbar region were determined after the both the thoracic and pelvic regions were scaled and positioned. By examining the distances between the newly positioned points of the thoracic and pelvic regions, scaling factors for the lumbar regions were established. Because each of the models had a TLC=0° posture, the points on the lumbar spinal processes are positioned in a straight line in the sagittal plane. This linearity, therefore, causes the scaling factor for the x (anterior) direction to be the same as the scaling factor for the z (superior) direction. To determine this scaling factor, the distance was calculated from the newly placed T_{12} *spinous process* point to the newly placed point defining the midpoint of the top upper buttock (UB) contour line. Table 3 shows the calculated distances and the resulting scaling factors. It should be noted that these scaling factors are smaller than those for the pelvic and the thoracic regions for both the large male and the small female. This may be explained by Boughner's three-dimensional model [6], which was based upon the two-dimensional model developed by Haas [5]. Even though Haas lowered the position of the pelvis relative to the thorax from the UMTRI position for the average male, creating a longer lumbar spine in the JOHN model, the positions of the large male and small female pelvic regions were never fully examined. It is likely that the pelvises in both the UMTRI small woman and large man were also positioned too high, as was discovered for the UMTRI average man.

Table 3. Lumbar Distances and Scaling Factors

	JOHN (mm)	JANE (mm)	JANE Scaling Factor	JERRY (mm)	JERRY Scaling Factor
T_{12} sp. pro. - UB midpnt.	206.4	162.3	0.78	212.2	1.02

Information about the y (lateral) direction in the lumbar region was not available from the UMTRI data

because all of the points in that region lie on the mid-sagittal plane (y = 0). In the JOHN model, the points defining the skin contour in the lumbar region covers only a narrow area down the center of the spine and over the *erector spinae* muscles directly to each side of the vertebral column. Therefore, it was assumed that using the y scaling factor calculated for the thoracic region, Table 1, would suffice over the narrow width of the lumbar region for both the small female and the large male.

The placement of the lumbar region was determined by the final placement of both the pelvic and thoracic regions. The points associated with the lumber region were rotated and translated into position by examining the position of the point associated with the T_{12} *spinous process* and the midpoint of the top upper buttock contour. The angle of the lumbar region was determined by calculating the angle between a line drawn from the T_{12} point to the pelvic spinal point and a line in the -x (posterior) direction. Once the lumbar region was rotated to the calculated angle, the region was translated into position such that the point associated with the T_{12} spinous process was placed in its original position. Table 4 summarizes the calculated angle of the lumbar region and the amount of translation to move the region into position for both the small female and the large male.

Table 4 Lumbar Translation Values

	JANE	JERRY
x (anterior) translation (mm)	-22.5	-31.155
z (superior) translation (mm)	16.6	5.924

Oblique views of the JANE and JERRY contours at a value of 0° TLC are shown in Figure 14. In comparing each of the new contours to the contours created by UMTRI for the small female and the large male, the respective contours match to within several millimeters of one another. Figure 15 is a front and side view of JANE's contour overlaid onto the UMTRI small female contour. The front and side view overlay of JERRY's contour with the UMTRI large male is seen in Figure 16.

Each of the new contours of JANE and JERRY are representative of the size and shape of the small female and the large male and each maintain muscle and bone definition throughout the contour. The JANE and JERRY contours can also be articulated into any degree of TLC and TRA, as their contours are based upon surface points as in the JOHN contour. By scaling the contours of JANE and JERRY from the JOHN contour, each contour is now related to the other. The seat designer now has the ability to model seats for not just the average male, but can design seats that fit a larger array of occupant sizes.

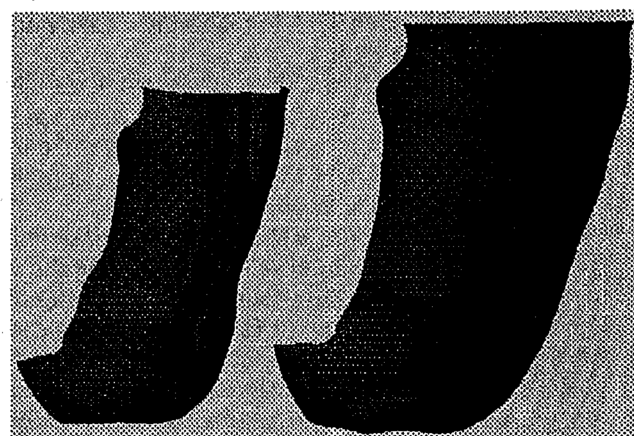


Figure 14. JANE and JERRY Contours

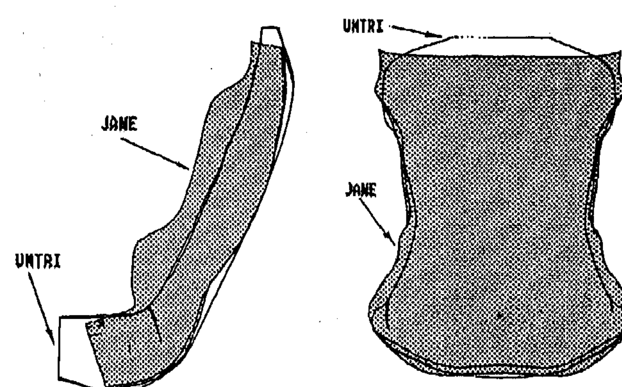


Figure 15. Comparison of JANE and UMTRI Contours

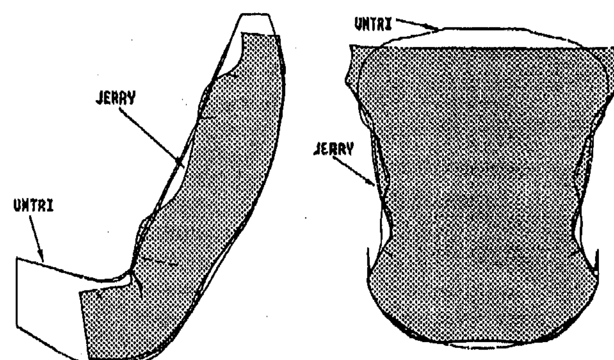


Figure 16. Comparison of JERRY and UMTRI Contours

CONCLUSION

The external contour of the JOHN model was reconfigured to be a more anatomically representative model of an undeflected skin surface. The new contour was developed using the contour created by Bush [9] and from studies done by UMTRI [2,3]. The new JOHN contour, which is partially based on the UMTRI average male contour, has a more representative size but also retains the muscle and bone definition noted in Bush's model.

A contour articulation program for each of the adult skeletal models was created. The external contours of the small female, the average male, and the large male can now be articulated in the sagittal plane to

any value of TLC and TRA to follow the underlying musculo-skeletal systems.

External contours for both the small female and the large male were developed from the JOHN average male model. Both of the JANE and the JERRY contours are scaled from the JOHN contour, thus making all of the contours related to each other.

With a more representative size and definition for the model, the designer will be able to use the model to create seats that fit the occupant better. Being able to use contours at any value of TLC and TRA, the designer will be able to fit occupants in seats in a larger variety of postures. By allowing the designers to model not only the average male but also the small female and the large male, he or she can design a seat that will fit a larger segment of the population. Fit, while it is by no means the sole component, is one of the most important elements to designing comfortable seats. If seat design tools, like the JOHN family of models, can better represent how a large variety of people fit in seats in different positions, seats designed with those tools should be more comfortable.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of this work by the Automotive Systems Group of Johnson Controls, Inc.

REFERENCES

1. Hubbard, R.P., W.A. Haas, R.L. Boughner, R.A. Canole and N.J. Bush, "New Biomechanical Models for Automobile Seat Design", paper no. 930110, International Congress, Society of Automotive Engineers, 1992.

2. Schneider, L.W., D.H. Robbins, M.A. Pflüg and R.G. Snyder, "Development of Anthropometrically Based Design Specifications for an Advanced Adult Anthropomorphic Dummy Family, Volume 1", UMTRI-83-53-1, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C., 1983.

3. Robbins, D.H., "Anthropometric Specifications for the Mid-Sized Male Dummy, Volume 2", UMTRI-83-53-2, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C., 1983.

4. Robbins, D.H., "Anthropometric Specifications for Small Female and Large Male Dummies, Volume 3", UMTRI-83-53-3, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C., 1983.

5. Haas, W.A., "Geometric Model and Spinal Motions of the Average Male in Seated Postures", Masters Thesis, Michigan State University, 1989.

6. Boughner, R.L., "A Model of Average Adult Male Human Skeletal and Leg Muscle Geometry and Hamstring Length for Automotive Seat Designers", Masters Thesis, Michigan State University, 1991.

7. Clemente, C.D., Anatomy - A Regional Atlas of the Human Body, Urban & Schwarzenberg, 1981.

8. Hollinshead, W.H. and C. Rosse, Textbook of Anatomy, 4th Edition, Harper and Row, 1985.

9. Bush, N.J., "Two-Dimensional Drafting Template and Three-Dimensional Computer Model Representing The Average Adult Male In Automotive Seated Postures", Masters Thesis, Michigan State University, 1992.

10. Koritké, J.G. and H. Sick, Atlas of Sectional Human Anatomy, Volumes 1 & 2, Urban & Schwarzenberg, Baltimore, MD, 1983.

970592

Kinetic Computer Modeling of Human Posture in Automotive Seats

David F. Ekern, Christopher J. Gedraitis, and Robert P. Hubbard
Michigan State University

Neil J. Bush
Johnson Controls, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

To assist automotive seat development and evaluation, a technique for predicting the posture of seated occupants has been developed. The method involved modeling the torso geometry and articulation of a mid-size male, based on information presented in SAE paper number 930110 [1]. This mid-size male model, known as 2-D JOHN, was developed in a commercial kinetic modeling software and used in a comparative seat evaluation study between a current production automotive seat and a prototype articulating seat. The 2-D JOHN model was supported a greater range of postures, defined as Total Lumbar Curvature (TLC) and Torso Recline Angle (TRA), in the prototype seat than the automotive seat.

INTRODUCTION

An important design goal of the automotive industry for the past 40 years has been to accurately locate the position of seated occupants within the interior space of the automobile. Knowing occupant positioning is very important for factors such a visibility, safety restraint positioning, and comfort. The position of the occupant in the seat and interior has a strong influence on comfort, because the occupant position will affect posture and seat pressure.

Since the early 1960's, the main tools used in the design and development of automotive seating have been the 2-D drafting template (Figure 1) and 3-D testing manikin (Figure 2) specified in SAE standard J826 [2]. Though originally designed only for locating the occupant within the automotive interior, the SAE J826 tools have been used extensively to design seat shapes and thereby affect the posture and comfort of seated occupants. However, the SAE tools do have a major drawback in that they have a combined thigh-pelvis segment and torso segment that together only articulate about the H-point (the hypothetical center of motion of the hip joint), so that unlike true human motion, the SAE tools do not have the ability to represent a change in lumbar curvature. In addition, the posture that the torso segment represents is a slumped posture, with a flat back in the lumbar region.

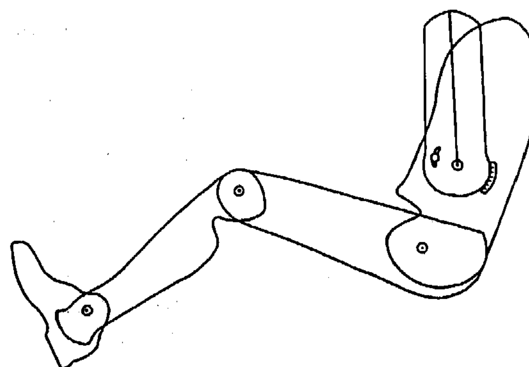


Figure 1 - SAE 2-D Drafting Template

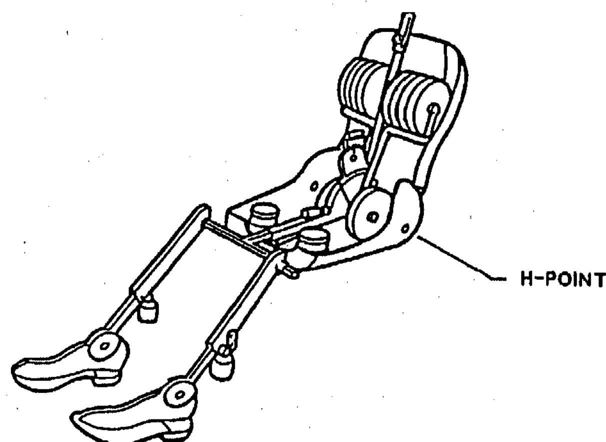


Figure 2 - SAE 3-D Testing Manikin

Due to these limitations, the J826 3-D manikin is unable to fit well in many automotive seats with large amounts of lumbar prominence. If the 3-D testing manikin is placed into a seat with a large amount of lumbar prominence, the manikin is unable to conform to

the shape of the seat back and the torso segment will either pivot about the lumbar area as it is forced back into the seat, which causes the thigh-pelvis segment to shift forward in the seat, or the torso segment will not be able to recline into the seat at all due to the large amount of lumbar prominence pressing on the segment.

In recent years, biomechanical research at Michigan State University (MSU) has been directed toward the development of human body models for automotive seat design. The main thrust of the work at MSU has been in improving the representation of human torso shape and posture by introducing lumbar curvature as a movement factor. To accomplish this, the human body models developed at MSU have been designed to represent more body segments than the SAE models. The models were designed to have separate thorax, lumbar, pelvis and thigh segments.

The human body model of the average size male (approximately 5'9", 165 lbs.) which was developed at MSU was named JOHN, to recognize the support of Johnson Controls, Inc. Details regarding the development of JOHN are described in SAE paper number 930110 [1], but descriptions of the positioning parameters are detailed here. The position of JOHN is described through the use of three variables, Total Lumbar Curvature (TLC), Torso Recline Angle (TRA), and Hip Joint Center (HJC). To represent the articulation of the human torso from a slumped to an erect posture, a one-to-one relationship of motion between the pelvis and thorax was selected [3]. For example, a 5° rotation of the pelvis corresponds to a 5° counter rotation of the thorax. This relationship of thorax to pelvis rotation was defined as TLC. The zero reference for TLC (0°) is a straight lumbar spine, with the angular orientation of the pelvis and thorax based on the seated occupant posture defined in a study by the University of Michigan Transportation Research Institute (UMTRI) [4, 5]. Figure 3 shows the pelvis and thorax articulation as the JOHN model moves from 0° TLC to 40° TLC. TRA is the measure of overall torso recline and is defined as the angle between vertical and a line passing through the top and bottom lumbar joint centers. These joint centers are the twelfth thoracic/first lumbar (T12/L1) and the fifth lumbar/first sacral (L5/S1), respectively. Figure 3 shows the 2-D JOHN model with a TRA of 30°. The HJC is the representation of the center of rotation between the femoral head and the acetabular socket of the pelvis.

The previous work at MSU concentrated on modeling the geometry and motion, or kinematics, of the human body in seated postures. For this study, a computer model was developed that was kinetic, predicting both the motion and forces acting on the body segments, while continuing to represent the posture of the JOHN model.

The 2-D JOHN kinetic computer model was developed with the aid of Working Model™, a commercial software product. This program is used for design of mechanical systems. To simulate mechanisms, it is possible to constrain the model bodies with elements such as pin and rigid joints, gears, and springs. Gravity and other forces can be applied to bodies in the software, and the program will automatically detect and represent collisions between bodies.

The development of this model represents a step forward in understanding and modeling the force-posture relationship of seated humans in automotive seats. By understanding this relationship, the iterative process of re-designing prototypes to support particular occupant postures can be shortened through use of the 2-D JOHN model. Predicting human posture in the design stage will allow seat designers to compare and refine seat designs on the computer before making prototypes, saving both time and money.

Specifically, the objectives of this study were to 1. develop the 2-D JOHN kinetic computer model and 2. undertake a comparative study between a current production seat and an articulating prototype seat using the 2-D JOHN model in order to study the effect of different seat designs on the final posture of the model.

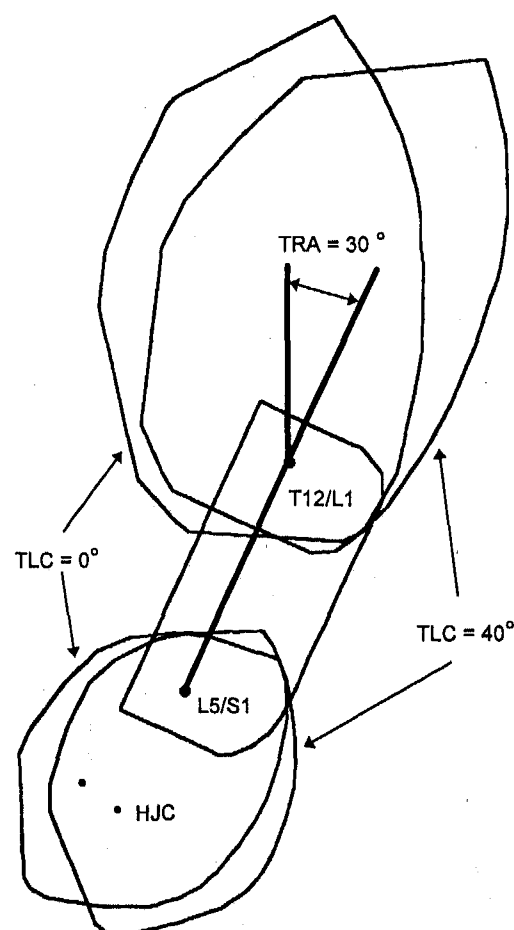


Figure 3 - 2-D JOHN model at a constant 30° TRA showing 0° TLC and 40° TLC

METHODS

2-D JOHN KINETIC MODEL CONSTRUCTION - The first step in the construction of the 2-D JOHN kinetic computer model was to develop the torso segments. Numerous contour points were selected from pre-existing 2-D CAD drawings of the pelvis, lumbar and thoracic body segments that were developed for the JOHN model, and these points were entered into the Working Model™ software as the vertices of polygons. The body segment contours represented the skeletal geometry of the three segments surrounded by soft tissue. The contours were adapted by Hubbard, et. al. [1] and Bush [6], based on data obtained from UMTRI [4, 5], and a corrected pelvis location that had been used in the development of the JOHN model in which the lumbar spine was lengthened by 30 mm. This anthropometric data was also used to locate the joint centers on the torso body segments. In Working Model™, points were placed on the body segments at the HJC, L5/S1, and T12/L1 joint centers (Figure 4).

After the torso had been completed, the legs for the 2-D JOHN model were constructed. The average male legs from the SAE J826 tools were imported into Working Model from a CAD computer file. Unlike the typical practice of using the SAE 95th percentile leg lengths with the 50th percentile body, the average male legs were used for this study. A consistent anthropometry was preferred because the study was concerned with seat issues, and not package issues. Because the 2-D JOHN model included a pelvis segment, the SAE thigh-pelvis segment was reshaped to represent only a thigh. Once the thigh had been reshaped, the joint centers were located on the segments in Working Model™.

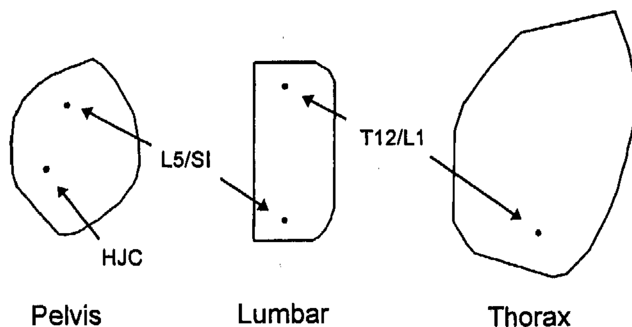


Figure 4 - 2-D JOHN kinetic model torso segments

With all the body segments sized appropriately and the joint center locations specified, the 2-D JOHN model was assembled with the motion constraints. Gear circles were created and rigidly pinned to the pelvis and thorax body segments. The lumbar segment was joined to the gear circles, creating the T12/L1 and L5/S1 joints. The pelvis, thigh, shank and foot segment were joined together, creating the HJC, knee joint and ankle joint. Finally a gear linkage was placed between the two gear circles, creating the 1:1 counter-rotation mechanism that controlled the torso articulation.

After initial trial simulations had been run with the 2-D JOHN model, it became apparent that another constraint was needed to keep the model from slumping so far that it would achieve a posture beyond the range of normal human movement. As a solution to this problem, a rope was added between the lower rear of the thorax and upper rear of the pelvis, that stopped the 2-D JOHN model from achieving a posture below -10° TLC. Figure 5 shows the constructed 2-D JOHN model.

BODY SEGMENT CENTERS OF MASS - When the 2-D JOHN model had been completed, the next step was to assign the center's of mass (COM's) for each body segment. For ease in modeling, a sagittal, symmetrical posture was assumed. As a result, the masses of the 2-D JOHN model thigh, shank and foot segments are the combined mass of the respective right and left body segments. In addition, because this study was focused on the effects of seat geometry, and not interior packaging, it was decided to eliminate the steering wheel in the model. This negated any possible effect of the wheel on the posture of the model due to the moment placed about the lower back when the arms are held out to grasp the wheel. An equivalent COM was calculated for the thorax, combining the mass of the head, neck, thorax both arms, with the arms oriented so that they were straight down by the sides of the body.

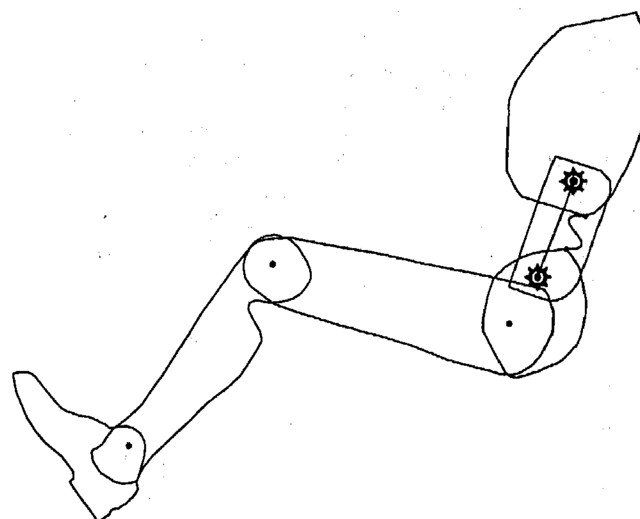


Figure 5 - 2-D JOHN model showing gear and rope constraints

The COM for the foot, shank, thigh, pelvis and lumbar segments were located utilizing average size seated male data developed in the UMTRI study [4, 5], and the COM of the body segments in Working Model™, were adjusted to correspond with the UMTRI data.

The equivalent COM for the thorax was calculated utilizing data of the average size male from the UMTRI seated occupant study [4, 5]. The head, neck, and thorax were kept in the same orientation as described in the UMTRI study. The upper and lower arm body segments, and the respective COM's, were re-oriented so that they hung straight down (vertical) the

sides of the body from the shoulder joint (glenohumeral joint). After the thorax equivalent COM location was calculated, the COM of the thorax segment in Working Model™ was repositioned to the calculated location (Figure 6).

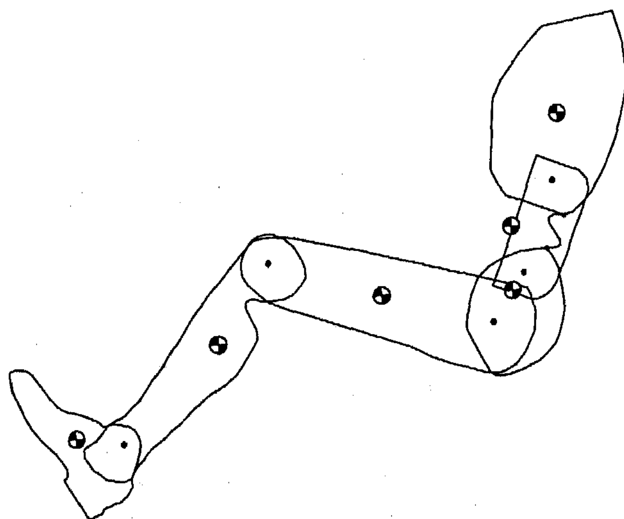


Figure 6 - 2-D JOHN showing location of COM's

CONSTRUCTION OF SEAT MODELS - When the 2-D JOHN model had been completed, it was used in a study to compare two different seat designs. The purpose of this study was to determine which seat geometry could support a greater range of postures, which are TLC values, at different levels of torso recline (TRA). The two seats that were compared were a prototype articulating seat, called the Biomechanically Articulating Chair (BAC) [7,8], and a current production automotive seat with an adjustable lumbar support, from a mid-size sedan automobile.

BAC model development - The BAC was designed to support the body in a wide range of postures by matching the motions of the chair to the motions of the body, defined through the JOHN model. There are 3 pieces to the BAC seat model; a pelvis support that cradles the back and bottom of the pelvis and rotates under the ischial tuberosities, a thorax support that pivots in the middle to upper back and allows rearward rotation of the top of the shoulders, and a recline bar that connects the pivots of the pelvis and thoracic supports. The full-size prototype of the BAC includes two independent thigh supports which were not included in the BAC seat model for modeling purposes. Like the 2-D JOHN model, the thorax and pelvis supports of the BAC move together in a one-to one relationship where a rotation of the pelvis support results in an equivalent counter-rotation of the thorax support. For this reason the position of the BAC supports is described in terms of the TLC and TRA promoted by the seat. The coupled rotation of the pelvis and thorax supports results in a change of TLC, and a seated person's TRA can be changed by adjusting the angle of the recline bar.

To develop the BAC in the Working Model™ simulation, dimensions of the thorax support, pelvis support and recline bar angle were measured from a full size prototype of the BAC. In addition, the angular relationship between the thorax and pelvis support were measured using a full size 2-D template of the JOHN model [6]. These dimensions and angles were used to construct and orient the rigid seat support surfaces in the simulation. The seat was then assembled with pin and rigid joints (Figure 7).

Mid-size sedan seat model development - To develop the mid-size sedan seat model (Figure 8), a design drawing for the seat was obtained from the seat manufacturer. The first step in developing the mid-sized sedan seat model in Working Model™ was to construct the visual references. To accomplish this, points from the seat frame and undeflected foam contours were measured from the design drawing and used to construct polygon bodies, with the H-pt of the design drawing as the origin of the coordinate system. After the thin, undeflected contours were created they were anchored to the background of the model. The seat frame was anchored to the background with a pin joint located at the recliner position.

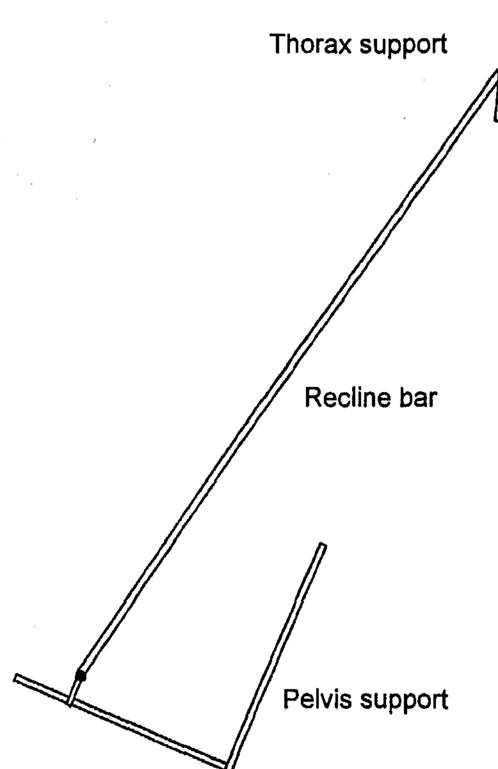


Figure 7 - Picture of BAC seat model

The second step involved developing the deflected seat contours for the seat cushion and seat back. The contour shape used for the deflected seat cushion was a reference line from the design drawing that represented the interface between the deflected seat cushion and a fully loaded SAE 3-D testing manikin. Seventeen points along this reference line were

measured from the design drawing and used to create a thin polygon to represent the deflected seat cushion contour. Then, seven small, thin rectangles were created and oriented on top of the deflected contour, following it as smoothly as possible. These seven polygons were anchored to the background and were the rigid seat cushion surfaces that the 2-D JOHN model contacted. The piecewise contour was used instead of the one-piece contour in order to improve the speed of the simulations, which ran slowly with the one-piece contour.

The seat back contour represented deflection of the known foam thickness in four regions. These regions were the top of the frame (upper thorax support), mid-back wire tie down, lumbar support and rear pelvis wire tie-down. To develop the contour, the distance between the undeflected contour and the back of the foam was measured from the seat drawing in these regions. The maximum deflection into the seat at these four regions was estimated as 65% of the foam thickness. The deflections at the four regions were calculated and then four rectangle bodies were created and positioned appropriately in Working Model™ to represent the deflected seat back surface at each of the regions.

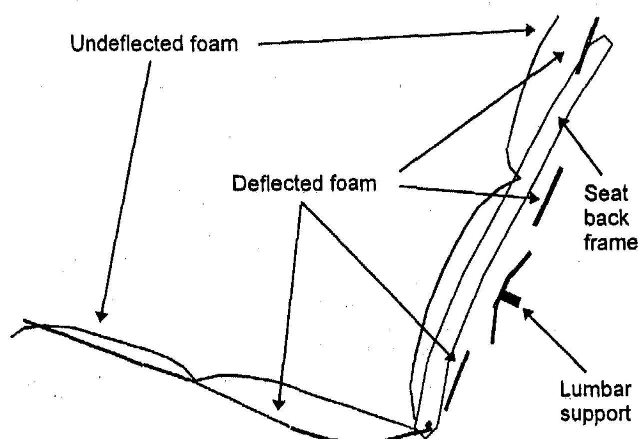


Figure 8 - Picture of the mid-size sedan seat model

The final step in the development of the BAC and mid-size sedan seat models was to specify the material properties for all contact surfaces. The two material variables were coefficient of friction and coefficient of restitution. There is little data on the coefficient of friction (μ) between a clothed human and an automotive seat, and so to determine the coefficient of friction to be used, a series of simulations were run using the 2-D JOHN model and the mid-size sedan seat model. From the results of these simulations, a μ of 0.7 was chosen as the smallest value where no slip between the 2-D JOHN model and the seat was observed.

Due to the nature of the modeling software the collisions between the 2-D JOHN model and the seat model bodies were simulated as rigid collisions. In order to counteract the elastic rebound between the rigid seat surface and 2-D JOHN model bodies and to more closely simulate a real person sitting into an automotive seat, the coefficient of restitution was set at 0.02.

SEAT EVALUATION PROTOCOL - With the 2-D JOHN model and seat models completed, study conditions and a simulation protocol were developed. To study the ability of the seat models to support different postures, the effects of two different seat factors on the posture of the 2-D JOHN model were simulated. The seat factors chosen were:

1. amount of lumbar curvature promoted, defined for the JOHN model as TLC.
2. angle of the torso support relative to vertical, defined through the JOHN model as TRA.

The two variables of the mid-size sedan seat that produced such posture changes were the seat back recline angle, and amount of lumbar prominence. The seat frame recline angle was defined as the angle between the a line parallel to the lumbar region of the seat frame and vertical, and the amount of lumbar prominence was defined as the distance the lumbar support was forward of a fully retracted position, along a line perpendicular to the seat frame. The corresponding variables for the BAC were the recline bar angle, and the angular relationship between the thorax and pelvis supports. The recline bar angle was defined as the angle of the recline bar with respect to vertical, and the relative angular position of the thorax support to the pelvis support was defined as the TLC measured with the JOHN 2-D articulating template.

In order to study whether the mid-size sedan and BAC could support a wide range of postures, different values of the mid-sized sedan and BAC seat variables were chosen and placed on different axes of a matrix, forming a set of study conditions. Mid-sized sedan seat frame angles and BAC recline bar angles were chosen to support upright, reclined and intermediate angles of torso recline. Values of lumbar support prominence for the mid-sized sedan model and angular orientation of the pelvis support to the thorax support for the BAC were chosen that would support slumped, erect and intermediate postures. The two variables affecting torso recline and lumbar curvature for each seat were placed on the axes of their respective matrix and when the possibilities were combined, the 3x3 grid resulted in nine seating conditions for each seat.

The mid-sized sedan seat design drawing was used to determine the starting values for the matrix. The design position of the lumbar paddle was defined as 0 mm of prominence, which was the fully retracted position. The design angle of the seat frame in the lumbar region was 28°, which corresponded to an SAE 3-D testing manikin torso angle of 24°. This angle was used as an intermediate value of seat frame angle. Upright and reclined values of the seat frame angle were selected as 23° and 33° respectively. Intermediate and

erect levels of lumbar prominence were defined as 12.5 mm and 25 mm.

To support an upright through reclined torso postures in the BAC seat model, values of the recline bar angle were selected that would promote a 23°, 28°, and 33° TRA in the 2-D JOHN model. As discussed earlier, the mid-sized sedan seat variable, amount of lumbar prominence, corresponded to the BAC variable of angular relationship between the pelvis and thorax supports, which was the TLC promoted by the seat. To model from a slumped to an erect posture, values of this angular relationship were selected that would promote TLC's of 0°, 20° and 40°. Recall that a TLC of 0° is defined for the 2-D JOHN model as a straight lumbar spine, corresponding to a slumped posture.

When the matrix of study conditions had been completed, a simulation protocol was developed and specified as follows:
1. the 2-D JOHN model was oriented in a 40° TLC posture
2. the TRA of the JOHN model was adjusted to match the predicted torso angle promoted by the seat for that particular seat set-up
3. because the 2-D JOHN model pelvis rolled rearward on the ischial tuberosities as the pelvis collided with the seat surface, the 2-D JOHN model was positioned with the HJC located 33 mm forward and 6 mm above the design H-pt location for the mid-sized sedan seat, and above the ischial tuberosity pivot for the BAC
4. the simulation was started.

Due to the effects of gravity and the mass of the body segments, the model collided with the seat surfaces, and its final posture and position was determined by the location of the seat supports and the mechanism controlling the movement of the 2-D JOHN model. When the model did not have a change in TLC and TRA for 20 consecutive time-steps, the simulation was manually stopped and the TLC and TRA values were recorded. The simulation was reset and the seat variables were until all nine simulations for each seat model had been completed.

After the 18 simulations described above had been completed and the data analyzed, another four simulations were performed using the mid-size sedan seat model. These simulations were run to study the effect of the upper thorax support location on the final posture of the model. The seat configuration used for these simulations was the one that produced the greatest amount of lumbar curvature in the mid-size sedan seat model. The effect on the 2-D JOHN model posture due to the proximity of the upper thorax support was studied by moving the upper thorax support rearward, perpendicular to the seat frame, in 10 mm intervals. The upper thorax support was moved back until the 2-D JOHN model was no longer in contact with the upper thorax support after becoming stable.

RESULTS

The results for the mid-size sedan simulations and the BAC simulations are shown in Table 1 and Table 2 respectively [9].

Figure 9 displays the results for the mid-size sedan simulation at 28° frame angle, with maximum lumbar prominence. Figure 10 shows the simulation results from the BAC seat model set up to promote a 28° TRA and 20° TLC.

Table 1 - Mid-sized sedan simulation results

Lumbar prom. / Seat frame ang.	0 mm	12.5 mm	25 mm
23°	TLC = -10° TRA = 25°	TLC = -10° TRA = 24°	TLC = -10° TRA = 21°
28°	TLC = -2° TRA = 30°	TLC = -8° TRA = 29°	TLC = 3° TRA = 29°
33°	TLC = -1° TRA = 35°	TLC = 5° TRA = 38°	TLC = 14° TRA = 38°

Table 2 - BAC simulation results

TLC / TRA	0°	20°	40°
23°	TLC = -1° TRA = 22°	TLC = 20° TRA = 23°	TLC = 40° TRA = 23°
28°	TLC = -2° TRA = 27°	TLC = 21° TRA = 28°	TLC = 40° TRA = 28°
33°	TLC = -2° TRA = 32°	TLC = 21° TRA = 33°	TLC = 39° TRA = 33°

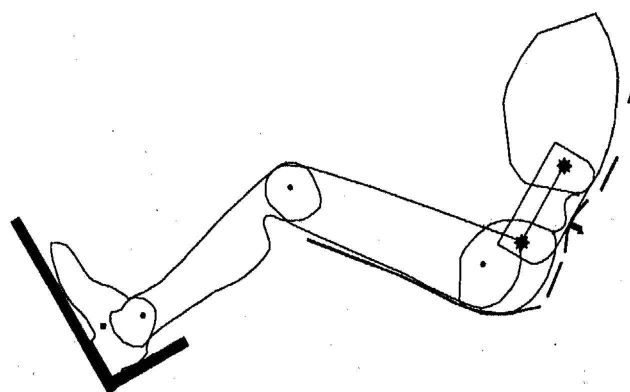


Figure 9 - Mid-size sedan simulation results for 28° frame angle, maximum lumbar prominence

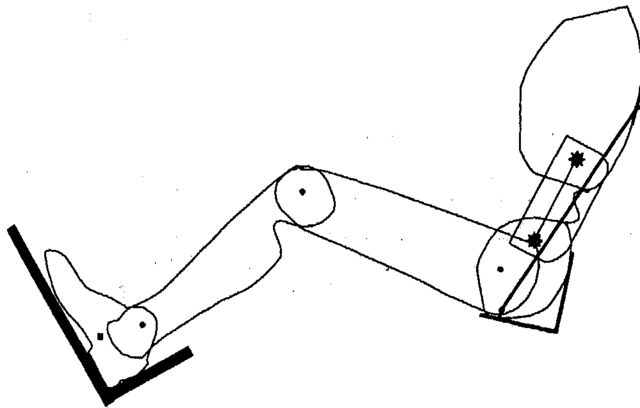


Figure 10 - BAC simulation results for 28° TRA promoted, 20° TLC promoted

Figure 11 summarizes the results from the mid-sized sedan and BAC seat simulations in graphical form. The graph plots the change in TLC versus the change in TRA for a constant seat recline angle. Therefore each line on the graph has three points, one point for each level of lumbar curvature simulated at each seat recline angle. The numbers 1,2, and 3 next to points on the plotted lines represent the amounts of lumbar curvature promoted by the seat for a particular seat recline angle. A slumped posture is represented by the number 1, an intermediate posture by number 2 and an erect posture by number 3.

As can be seen from Figure 11, the three lines representing the BAC simulations are all horizontal, and long. This shows two things, first, that TLC change was independent of TRA change, meaning that the BAC was able to support a range of TLC values while holding the TRA of the JOHN model constant. This was possible for every level of seat recline angle tested, 23°, 28°, and 33°. The second point is that the BAC was able to support a very wide range of TLC values, from 0° to 40°, at each level of seat recline. This is the ideal situation for a seat, supporting a wide range of lumbar curvature at any level of torso recline angle.

The results from the mid-size sedan simulations show a different story. Looking first at the results from the 23° seat frame angle, it can be seen that the line is vertical, opposite that of the BAC results, and is constant at -10° TLC. This demonstrates that at an upright seat frame angle, the mid-sized sedan seat was unable to promote any amount of lumbar curvature, and that as the lumbar prominence increased, the only result was that the TRA of the 2-D JOHN model decreased. At 28° seat frame angle, the line shows that the TLC of the JOHN model changed from -2° to -8° back to 3° TLC as the lumbar prominence increased, but the TRA of the 2-D JOHN model remained relatively constant. So at design position, the TLC of the JOHN model can be changed with increased lumbar prominence, but the change is erratic, and the posture that the 2-D JOHN model maintains is still very slumped, around 0° TLC. At 33° seat frame angle, a very reclined seat set-up, the results show that the mid-size sedan seat is able to promote a medium amount of lumber curvature, showing a total change of 15°, while keeping the TRA of the 2-D JOHN model fairly constant. However, this may not be a chosen driving posture, due the recline angle of the seat.

The results from the supplemental study to evaluate the effect of the upper thorax support on the final posture of the 2-D JOHN model are added to the graph of Figure 11 and are shown in Figure 12. The graph shows that moving the upper thorax support rearward 30 mm allows the TLC of the 2-D JOHN model to increase an additional 11°. Allowing the upper thorax support to move allows the number of TLC values possible in the particular seat set-up to almost double, with a minimal change in the TRA of the 2-D JOHN model. This shows that if the final posture of the model is supported, or restricted, by the upper thorax support then allowing the upper thorax support to move rearward can have a dramatic effect on the final posture of the 2-D JOHN model.

DISCUSSION

The main observation from the simulations was the effect of the pelvis support on the final posture of the 2-D JOHN model. In all nine BAC simulations the pelvis comes to an immediate stop when it contacts the pelvis support. The TRA and thorax position continue to adjust slightly, but TLC change stops soon after the pelvis stops. The different amounts of TLC possible in the BAC are directly dependent on the starting distance between the rear of the 2-D JOHN model pelvis segment, in an erect 40° TLC posture, and the pelvis support. When this distance is very small, as in the case of all three simulations to support 40° TLC, the pelvis support constrains any rotation of the pelvis, thereby keeping the TLC at 40°. When the distance is the largest, as with the cases for 0° TLC, the pelvis will rotate rearward, resulting in a decease in TLC, until it comes into contact with the pelvis support. At that time the rotation of the pelvis stops, and the TLC does not change any further. The closer the pelvis support is to the rear of the 2-D JOHN model pelvis when the model contacts the seat surface and begins to slump, the greater the final TLC of the model.

In both the BAC and mid-sized sedan seat, the location of the upper thorax support is the second most important factor in determining the final posture of the 2-D JOHN model. If the upper thorax support is too far forward, relative to the pelvis support, then the 2-D JOHN model will come into contact with the upper thorax support before the pelvis support and the 2-D JOHN model will be unable to obtain an erect posture. The contact with the upper thorax support stops all rearward torso translation of the model, the change in TRA, therefore, any motion that occurs must be a rotation of the thorax and pelvis, or TLC change. This results in a more slumped posture. One of the reasons that the 2-D JOHN model could not obtain a large value of TLC in the

mid-sized sedan seat was due to the contact interference with the upper thorax support. In contrast, the upper part of the 2-D JOHN model thorax was free of obstruction in the upper thorax region.

From the simulation results, it was observed that the human torso can be thought of as two structures and a linkage, the structures being the thorax and pelvis, and the linkage the lumbar segment. Controlling the posture of the torso by influencing the position of the structures, like the BAC does, is much more effective than trying to control the posture of the torso by pushing on the lumbar region, like in the automotive seat. It seems clear from the simulation results that, in order to control the posture of the seated occupants, it is necessary to replace the current method of a single lumbar support with a more influential support system. The new system would involve a rear pelvis support, lower thorax support and an ability to let the upper part of the shoulders to move rearward. If a seat model can support a simplified mechanistic model of the human body, with no muscle forces, in an erect posture, then it can surely support a relaxed human body in a real seat.

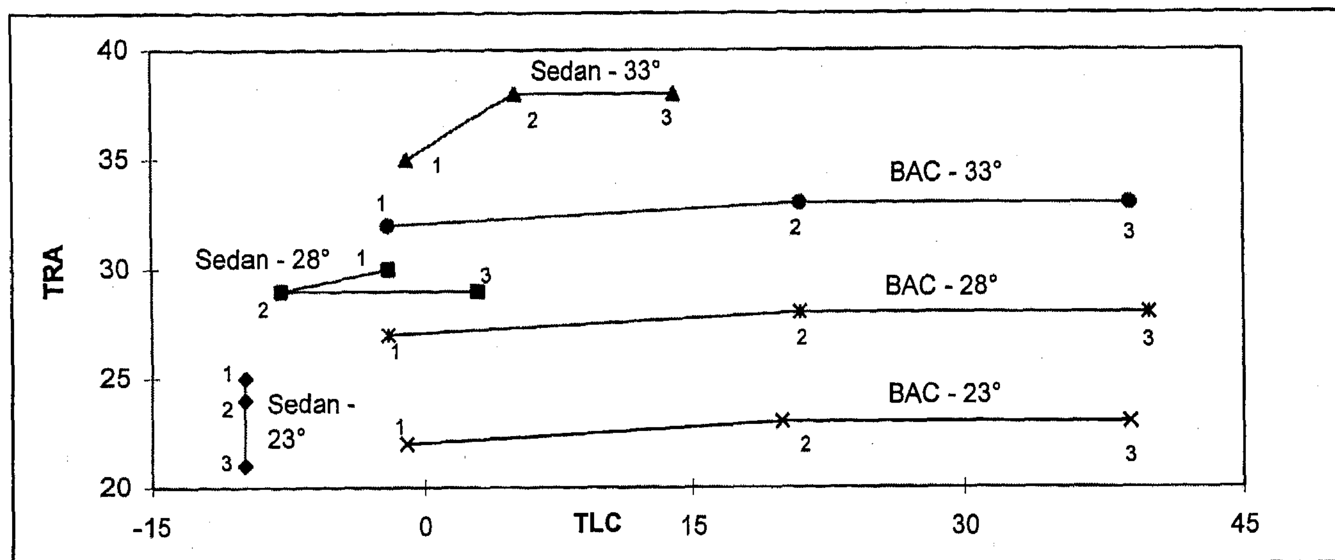


Figure 11 - Results from the BAC and mid-size sedan simulations

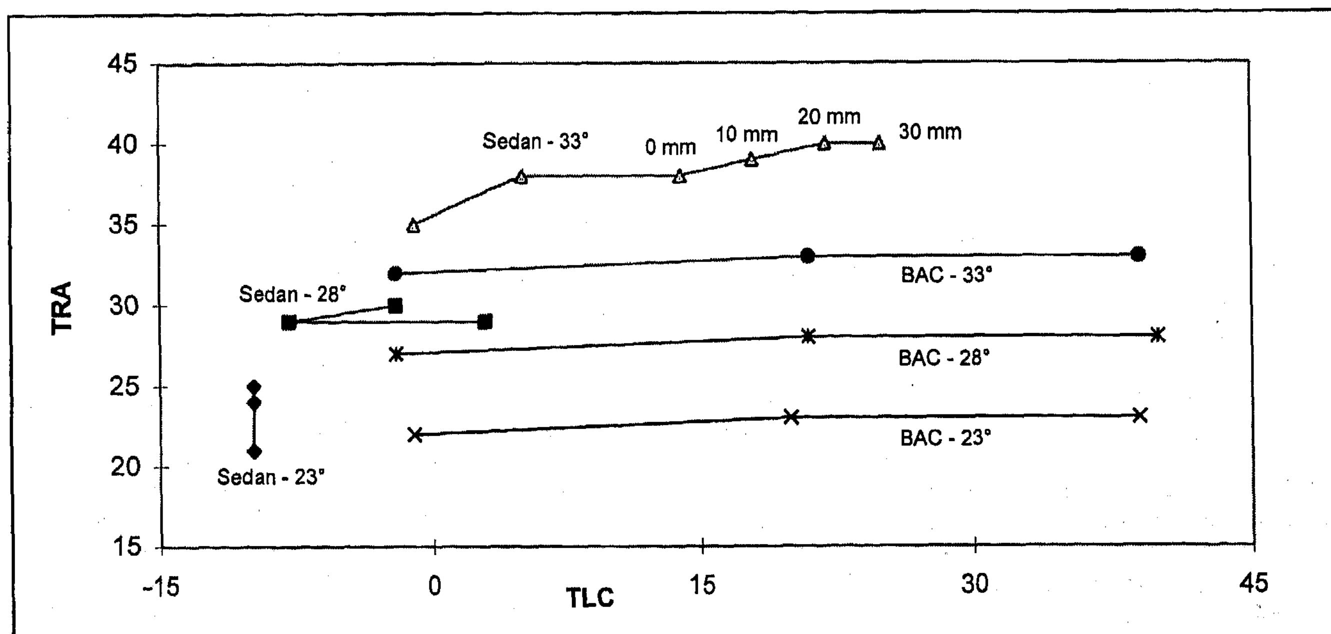


Figure 12 - Results from the upper thorax placement simulations

CONCLUSIONS

This paper described the development of a kinetic computer model of the human body. This model is called JOHN and was based on the model described in SAE paper 930110 "New Biomechanical Models for Automotive Seat Design". The model was used in a study to compare between two seat designs, one of which was a prototype articulating seat, and the other was a current production automotive seat. The construction of the 2-D JOHN model and seat models were described, and the seat testing protocol was detailed. The results from the simulations were discussed, showing that the prototype seat was able to support a wider range of postures than the automotive seat. Finally, a new method of supporting the human body in seated postures was described.

The ability to represent and simulate the interaction between a seat model and the 2-D JOHN kinetic human body model is a step forward in understanding the relationship between the posture of seated humans and the forces acting between the occupant and the automotive seats. Automotive designers now have a tool that can allow them to see the postural effect on the 2-D JOHN model by modeling different seat contours and posture controlling mechanisms. This effect can be related to the posture, and ultimately comfort, of a seated person. Through the use of the 2-D JOHN kinetic model, seat designers can modify seat contours on the computer, thereby saving time and money by reducing the amount of prototyping needed.

ACKNOWLEDGMENTS

The authors would like to thank the support of the Automotive Systems Group of Johnson Controls, Inc. for their support of this research.

REFERENCES

1. Hubbard, R.P., W.A. Haas, R.L. Boughner, R.A. Canole, N.J. Bush, "New Biomechanical Models for Automobile Seat Design", paper no. 930110, Internat. Cong., Soc. Auto. Eng., 1993.

2. SAE Standard J826, "Devices for the Use in Defining and Measuring Vehicle Seating Accommodation", Body Eng. Comm., Soc. Auto. Eng., Approv. Nov. 1962.

3. Haas, W.A., "Geometric Model and Spinal Motions of the Average Male in Seated Postures", Thesis for the Master of Science Deg., Dept. of Metal., Mechanics and Mat. Sci., Mich. State Univ., 1989.

4. Schneider, L.W., D.H. Robbins, M.A. Pflug, and R.G. Snyder, "Development of Anthropometrically Based Design Specifications for an Advanced Adult Anthropometric Family", Vol. 1, Fin. Rep. No. DOT-HS-806-715, U.S. Dept. of Trans., NHTSA, 1985.

5. Robbins, D.H., "Anthropometric Specifications for Mid-sized Male Dummy", Vol. 2, Fin. Rep. No. DOT-HS-806-716, U.S. Dept. of Trans., NHTSA, 1985.

6. Bush, N.J., "Two-dimensional Drafting Template and Three-dimensional Computer Model Representing the Average Adult Male in Automotive Seated Postures", Thesis for Master of Science Deg., Dept. of Metal., Mechanics, and Mat. Sci., Mich. State Univ., 1992.

7. Hubbard, R.P., "Seat with Biomechanical Articulation", U.S. Patent Appl. # 08/259,662, filed June 14, 1994.

8. Hubbard, R.P., Gedraitis, C.J., "Biomechanically Articulated Seat Concept and Prototypes" paper no. 970591, Internat. Cong., Soc. Auto. Eng., 1997.

9. Ekern, D.F., "Two-dimensional Kinetic Modeling of Human Posture in Automotive Seats", Thesis for Master of Science Deg., Dept. of Mat. Sci. and Mechanics, Mich. State Univ., 1996.

970593

Measuring and Modeling of Human Soft Tissue and Seat Interaction

Richard H. Setyabudhy, Akram Ali, Robert P. Hubbard, Clifford Beckett, and Ronald C. Averill
Michigan State University

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Deformations of soft tissues and seat cushion foam are significant factors in determining the interface contours between the seat and the back of the thigh. This paper describes the measurement of forces, deformations, and contours of people's thighs and seat cushion materials. The goal of this work is to represent the human interactions with seats. A two-dimensional, plane strain finite element method was used to develop a contact model between the cross section of the human mid-thigh and flat surfaces, which can be a flat, rigid surface or a flat, foam cushion of various thicknesses and densities. Results of human and seat interactions for various subjects were measured, modeled, and compared. The present work showed a good agreement between experiments and models for various subjects and foam densities. The important results showed that the stiffness of the foam does not depend on the foam thickness.

INTRODUCTION

Even though the behavior of living tissues is known to be non-linear and viscoelastic (1), the assumption of linearly elastic soft tissues is commonly used in finite element models (2,3,5). The contact mechanics can be categorized as a highly non-linear problem because geometry, boundary conditions and material properties are non-linear, therefore, by using linear material behavior in the finite element model the computation time can be reduced.

Applying non-linear properties of human soft tissues in finite element model becomes of growing interest as more powerful computer hardware and software become available. Vannah, et al. (7) used Jamus-Green-Simpson strain energy function to represent bulk muscular tissue between tibial plateau and ankle in their model. The geometry of the model was using CT scans with horizontal cuts at 20 mm intervals. A quasi-static test was conducted on bulk muscular tissue. The test was performed on the calf (posterior muscle belly) of the lower leg. A high distortion indentor test was chosen to deform the tissue. They assumed the tibia and fibula were rigid in comparison with the muscular tissue, and they represented the bones by nodal constraints in their model. A mold-making silicone rubber (HSII®, Dow-Corning Corp., Midland, MI) whose properties are similar to muscular tissue was tested uni-axially to obtain the Jamus-Green-Simpson coefficients for modeling. The results of finite element models and experiments due to 8 and 19 mm indentation were compared. They showed good agreement between models and experiments in load-deformation results.

Deng (6) studied the load-deformation responses of the hamstring regions of human thigh in seated postures. A seating fixture with load cells and electronic gauges was built to support people in variety of seated postures. A flat, rigid plate was pushed into the back of people's thighs. The tests were performed for different knee angles at 90, 120, 150, and 165 degree. Deng found that the thigh tissue became stiffer as the knee angle increasing. A 2D finite element model of hamstring muscle in contact with a rigid surface was developed. Ten males and females human subjects with various thigh sizes were selected and ten corresponding models were created. All the models used the assumption of a single property for the hamstring muscle and the only differences were the hamstring muscle dimensions of the subjects. The diameters of hamstring muscles of the mid-thigh were chosen based on the mid-thigh diameter of the subjects. The polynomial form of strain energy potential was chosen for hamstring muscle modeling. The load-deformation results of the experiments were compared with the models. Besides a good agreement between experiments and models, the results also showed the stiffness of muscle increases as the knee angle increases.

In this paper, the non-linear properties of soft tissues have been included in the model, however, the viscoelastic behavior of the tissues was considered not important for a subject who sits for a long period of time. Moreover, there is no benefit by adding the viscoelastic behavior in the model if the tissue-seat interaction is expected not occur in the short period of time, and it will increase the computation time. The objective of this study was to measure and model the interaction of human soft tissues of the posterior side of the right thigh with various foam densities.

METHODS

The soft tissue and foam were modeled as elastomer materials such that the systems of deformable bodies undergo large deformations; Ogden strain energy functions (8,9) were used to formulate these material properties. For practical purposes, the soft tissues and foam were each represented in the model with single, uniform materials with a rigid femur. To obtain the material parameters of the foam in the model, stress-strain curves were measured by uniaxial compression testing. Poisson's ratio effect is very small and it can be neglected. The Ogden formulation for foam will be used to fit these experimental data. Unlike the foam, the soft tissue's properties cannot be obtained with the testing machine. To get the soft tissue properties, the load-deformation curve of the simulation model (thigh in contact with rigid, flat surface) was compared with experiment until the good fit be obtained. Then the chosen soft tissue properties were verified for subjects with different sizes. These soft-tissue properties were also used for thigh-foam modeling.

EXPERIMENTAL - To measure the load-deformation characteristics of the posterior side of human thighs, a chair was designed to support a person in a variety of seated postures that affect the length and tensions in the hamstring muscles on the back of the thigh (6). This chair provided a backrest, a pelvis support, and a movable footrest. The footrest was designed such a way that the knee angle can be changed from 90 to 180 degrees. The knee angle was defined as the angle between two vectors: one from the femur greater trochanter to the lateral condyle of femur at the knee, and another from the lateral condyle of the femur to the lateral malleolus at the ankle. The backrest had a pivot that allowed rotation about lateral axes to comfortably accommodate the subjects. The chair dimensions were adjustable to correspond to each individual's body size. The backrest was adjustable both vertically and horizontally and supported the subject's torso. A flat plate supported the back of the pelvis, and two flat plates supported the right and left buttocks. A load cell was put under each plate to measure the force under the buttocks.

Seven human subjects were tested for the thigh and seat cushion interactions in seated postures. The heights of the subjects ranged from 65 to 76 inches, the weights ranged from 110 to 280 lbs., ages ranged from 18 to 36 years, and the thigh circumferences ranged from 447 to 695 mm. The physical characteristics of the seven subjects were summarized in Table 1.

Table 1
Subject Physical Data

Subject Number	Gender	Age (years)	Height (in)	Weight (lb.)	Mid-Thigh Circumference (mm)
1	Male	18	76	175	520
2	Female	23	67	135	531
3	Male	19	72	155	510
4	Male	21	71	135	462
5	Male	36	70	205	608
6	Male	26	76	280	695
7	Female	29	65	110	447

Based on UM-TRI (10) study, the subjects were divided into three categories according to the circumference of their mid-thigh such as small female (5th percentile), mid-sized male(50th percentile), and large-male (95th percentile) in erect seated postures. The measurements summarized in Table 2

Table 2
Thigh Circumference for Small Female (5th Percentile), Average-Male (50th Percentile), And Large-Male (95th Percentile)

Subjects	Minimum (mm)	Mean (mm)	Maximum (mm)
small female	370	427	472
average-male	442	504	550
large-male	485	559	635

Figure 1a shows a subject sitting on the testing chair. The subject's right thigh was aligned in such a way that it was level on the surface of a piece of foam (267 x 216 x 100 mm^3) and the mid-thigh was placed in the middle of that foam. The contour measurement system was located underneath the foam and above a load cell. The load cell was used to measure the reaction force under the thigh. The foam, contour measurement system, and load cell were moved up by a hydraulic jack with a six ton capacity and 254 mm stroke such that the back of a right thigh deformed due to the stationary the femur. A digital caliper was fixed from the jack to the contour measurement system and was used to measure the vertical displacement of the system. The thigh lateral deformation was recorded by placing another digital caliper across the middle section of the thigh and adjusted throughout the subject.

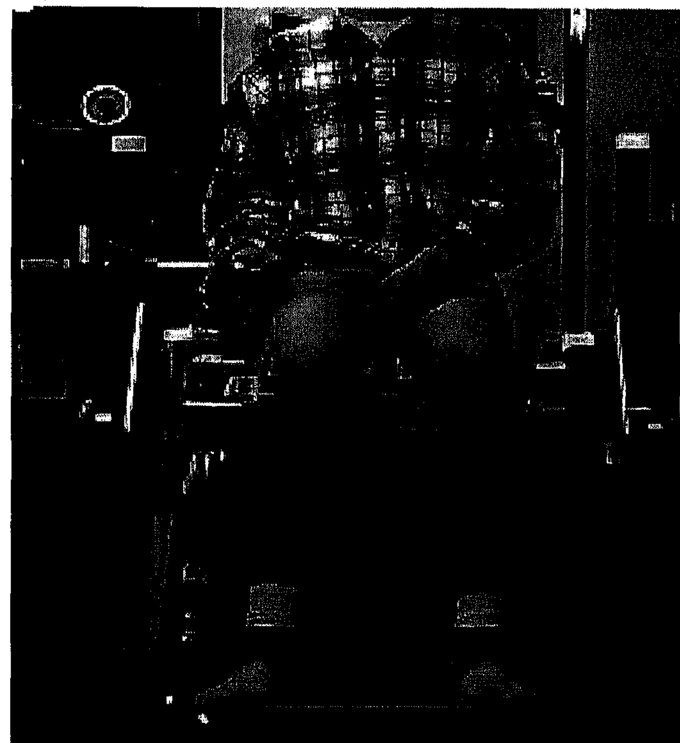


Figure 1a. A subject sitting on a testing chair.

Figure 1b. shows the right thigh of a subject in contact with a piece of flat foam. Under the foam, there was a contour measuring system.

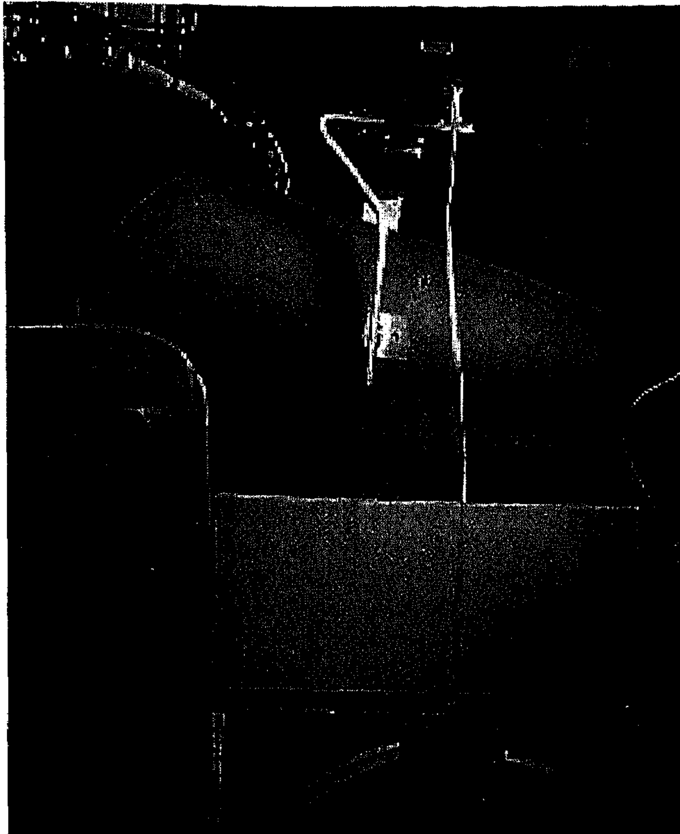


Figure 1b. The right thigh of a subject on a flat foam.

The Contour Measurement System - The contour measurement system uses wires through the seat cushion foam samples. This was used to measure deflections of people's soft tissues with seat cushion foams of different stiffness and thickness. When the foam was deformed, the wires were displaced an equal amount. The video camera captured an image for each wire displacement. These displacement images were digitized using a computer connected to the camera. While raising the contour measuring system into the back of the thigh, measurements of load, contour, and deformation of the foam and thigh were collected.

Thigh-Foam Interaction Tests - The chair was adjusted according to the subject's body size. The subject sat in the seating fixture, feet placed firmly on the footrest plate, head and body in a relaxed posture. All subjects wore shorts during testing, so that clothing was not a variable. The contour measurement system was displaced in a quasi-static manner up into the back of the subject's thigh. Contact was identified when the measurement force exceeded 0.22 N. The digital caliper attached to contour measurement system recorded the total vertical deformation of the thigh and foam for a given position. The foam was raised into the thigh in displacement steps of 5 millimeters. Each displacement step was held for 20 to 30 seconds before the resulting force was recorded. A rest time of 20 to 30 seconds was chosen in order to allow for stress relaxation. After this time, the forces were stable. The contour measurement system was displaced until the reaction force lifted the leg so that the foot was lifted from its rest. The total deformation (δ_{Total}) in the vertical direction was measured by the displacement of the contour measurement system and the foam deformation (δ_{Foam}) was recorded by the displacement of the wires. A video camera captured the wire displacement image for each reaction force and was digitized by the computer. The foam deformation can be obtained by the following formulation:

$$\delta_{Foam} = \delta_a - \delta_b$$

where:

δ_b: The end position of wires before the foam deformation.

δ_a: The end position of wires after the foam deformation.

The vertical deformation of the mid-thigh at the contact area with the foam was calculated by subtracting the foam deformation from the total deformation. This can be formulated as follows:

$$\delta_{Thigh} = \delta_{Total} - \delta_{Foam}$$

This computation was based on two assumptions: (1) the femur did not move during the test (Figure 2); (2) the frictional force between the foam and thigh prevented the thigh's tissue from sliding at the contact surface.

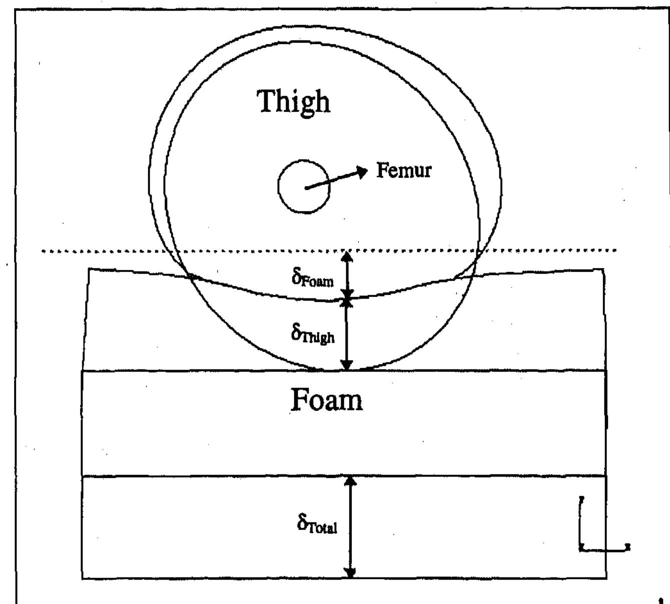


Figure 2. Definitions of total, thigh and foam vertical deformations.

The deformed contour of the foam, which is in contact with the thigh, can be obtained through the wire displacements. Figure 3 shows the measured contour of a piece of foam that was deformed by contact with thigh.

Two foam densities, 51 kg/m^3 (very stiff), and 41 kg/m^3 (softer) with two different thicknesses (100 mm, 25 mm) were used in this experiment.

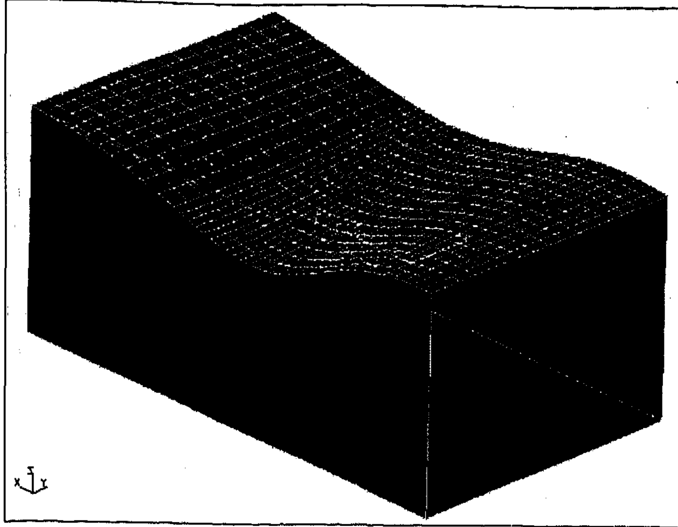


Figure 3. The contour of foam deformation due to a contact with the right human thigh.

FINITE ELEMENT MODEL- A two-dimensional, plane strain finite element model was used to represent the interaction of the mid-thigh and a seat cushion foam. A cross section of the mid-thigh which was used in model was based on sitting position of a subject with 100 degrees knee angle. For simplicity, a single uniform material property was used to represent the composite of the tissue (fat, muscle, skin, etc.) in the model. Large deformations occurred in the tissue when the back of the thigh was in contact with a flat, rigid surface or a piece of foam. Even though the stiffness of the hamstring muscle may be different from one person to another, the models employed the same properties for all subjects. This is very important for practical purposes because the model can be scaled easily to simulate the thigh geometry of each subject without considering different tissue properties for various subjects.

Two commercial finite element software packages, ABAQUS and MARC, were used for modeling. It was very important to run the model with more than one software for verification because there is no analytical solution of these models. The soft-tissue has a low stiffness which increases with deformation and the foam has an initial stiffness which decreases. These differences in material properties of soft tissues and foam make the problem more difficult to solve due to the convergence problem if a large deformation occurs in the model. Small volume change is assumed in the soft tissue with Poisson's ratio close to 0.5. On the other hand, the uniaxial compression test of the seat cushion foam showed that the lateral displacement was very small even though with compression up to 80%. It means the zero Poisson's ratio effect is a practical assumption. To model the soft tissues, the hyperelastic material with Ogden form was chosen rather than polynomial form of the strain energy potential. This strain energy form contains a bulk modulus term which controls the compressibility of the hyperelastic material. Even though there is a slight difference in formulating a generalized Ogden strain energy potential in ABAQUS and MARC, we can still find a similarity between them. The following expressions of the Ogden potential strain energy for ABAQUS and MARC are:

- ABAQUS

$$U=\sum_{n=1}^{N}\frac{2\tilde{\mu}_n}{\tilde{\alpha}_n^2}(\bar{\lambda}_1^{\tilde{\alpha}_n}+\bar{\lambda}_2^{\tilde{\alpha}_n}+\bar{\lambda}_3^{\tilde{\alpha}_n}-3)+\sum_{n=1}^{N}\frac{1}{D_n}(J_{el}-1)^{2n}$$

- MARC

$$U=\sum_{n=1}^{N}\frac{\mu_n}{\alpha_n}J^{-\frac{\alpha_n}{3}}(\lambda_1^{\alpha_n}+\lambda_2^{\alpha_n}+\lambda_3^{\alpha_n}-3)+4.5K(J^{\frac{1}{3}}-1)^2$$

where $\bar{\lambda}_i=J^{-\frac{1}{3}}\lambda_i$ is a principal stretch ratio, and the elastic volume ratio J_{el} is a ratio between the total volume ratio J and the thermal volume ratio J_{th}. The J_{th} is a function of the linear thermal expansion ε_{th} such that $J_{th}=(1+\varepsilon_{th})^3$. Material parameters are $N,\mu_n,\alpha_n,\tilde{\mu}_n,\tilde{\alpha}_n,D_n$, and K.

A simple two dimensional plane strain models of a flat rigid surface in contact with the back of a thigh was used to defined how many terms were needed for Ogden model to represent the behavior of human soft tissues. The results of using seven healthy human subjects with various thigh sizes showed that only one term is sufficient for representing a thigh tissue. Typical material parameters in the plane strain model are as follows:

$N=1$, $\mu_1=1900\ N/m^2$, $\alpha_1=1.75$, $K=15{,}875\ N/m^2$,
$\tilde{\mu}_1=1662.5\ N/m^2$, $\tilde{\alpha}_1=1.75$, $D_1=1.4\times10^{-5}\ m^2/N$, $\varepsilon_{th}=0.0$.

Several flat cushion foams, with dimensions 400 x 400 x 100 mm^3 were tested to obtain their properties by applying standard uniaxial compression test using Instron machine. Unlike the hyperelastic materials which are slightly compressible, the hyperfoam materials are highly compressible. Both commercial finite element software packages, ABAQUS and MARC, provide a similar strain energy potential for foam modeling. The generalized Hill strain energy potential for foam modeling was used in ABAQUS and it can be written as:

$$U=\sum_{n=1}^{N}\frac{2\tilde{\mu}_n}{\tilde{\alpha}_n^2}[(\bar{\lambda}_1^{\tilde{\alpha}_n}+\bar{\lambda}_2^{\tilde{\alpha}_n}+\bar{\lambda}_3^{\tilde{\alpha}_n}-3)+\frac{1}{\tilde{\beta}_n}(J_{el}^{-\tilde{\alpha}_n\tilde{\beta}_n}-1)]$$

The generalized compressible Ogden model was used in the MARC program and the strain energy is defined as:

$$U=\sum_{n=1}^{N}\frac{\mu_n}{\alpha_n}(\lambda_1^{\alpha_n}+\lambda_2^{\alpha_n}+\lambda_3^{\alpha_n}-3)+\sum_{n=1}^{N}\frac{\mu_n}{\beta_n}(1-J^{\beta_n})$$

Here, J_{el} is elastic volume strain like a hyperelastic material. The only difference is the β_n and $\tilde{\beta}_n$ terms in foam modeling and D_n and K terms in tissue modeling. The β_n can be related to Poisson's ratio effect v_n such that

$$\beta_n=\frac{v_n}{1-2v_n}\quad\text{or}\quad v_o=\sum_{n=1}^{N}\frac{-\beta_n}{\alpha_n-2\beta_n}$$

where ν_0 is the initial Poisson's ratio.

Neglecting the Poisson's ratio effect means the β_n and $\tilde{\beta}_n$ terms become zero. The parameters of the hyperfoam model were obtained by using non-linear curve fitting method. The stress-strain curve of uniaxial compression test was fitted using the hyperfoam models. In order to fit the stress-strain curve from uniaxial data, two terms were needed. The parameters of a selected foam which has specific weight 41 kg/m^3 are:

- ABAQUS

  $\tilde{\mu}_1 = 18345.226$ N/m^2, $\tilde{\alpha}_1 = 17.41$,
  $\tilde{\mu}_2 = 210.46$ N/m^2, and $\tilde{\alpha}_2 = -1.984$.

- MARC

  $\mu_1 = 2107.39$ N/m^2, $\alpha_1 = 17.41$,
  $\mu_2 = -212.2$ N/m^2, and $\alpha_2 = -1.984$.

Figure 4 shows an undeformed mesh of 2D model with a piece of flat foam. The tissue model consisted of 900 four-noded quadrilateral elements which accommodate the compressibility behavior and the foam consisted of 102 four-noded quadrilateral elements. This size of mesh is good for deformation up to 50 mm and finer mesh was required in the contact region to prevent element distortions.

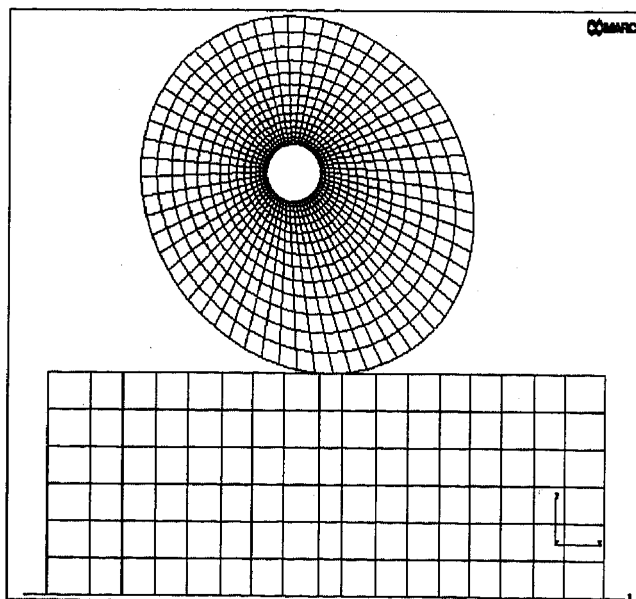


Figure 4. The finite element mesh of mid-thigh-foam cross section.

RESULTS AND DISCUSSION

The thigh-foam interaction responses of seven subjects were analyzed and the results of these subjects showed a similar behavior. For the convenience of this paper, an average male subject was chosen for comparing an experiment and a model. Therefore, all the results reflected to the average male.

Load-deformation responses of the foam and the thigh in seated postures were analyzed to study the mechanical behavior of soft tissue deformation interaction with different foam stiffness and thickness. Two trials of loading and unloading the thigh (Figure 5a) and the foam (Figure 5b) with a rigid plate were conducted to check the test repeatability. The graphs from these tests exhibited hysteresis behavior, and the different between the results of the two test trials was very small, so the tests were repeatable.

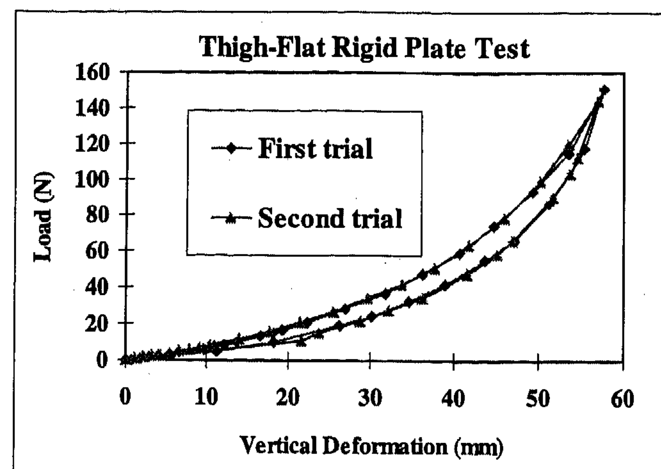


Figure 5a. Two deformation trials of thigh with a flat, rigid surface.

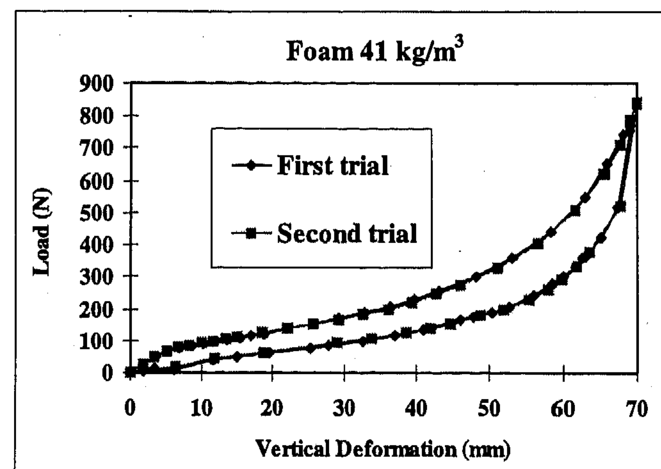


Figure 5b. Two deformation trials of foam with a flat, rigid surface.

The deformations of the two wires in the center of the foam sample were used to study the thigh-foam interaction. These two wires represent the first contact point between the foam and the thigh. A typical nonlinear load-deformation relationship of the thigh and foam for one of the two selected wires was shown in Figure 6. There were three curves which represented the thigh, foam, and total deformation versus load. One interesting thing can be seen in this figure that the stiffness of the thigh tissues is initially much smaller than the foam stiffness. After the initial loading, the stiffness of the

thigh increases and becomes greater than the stiffness of the foam.

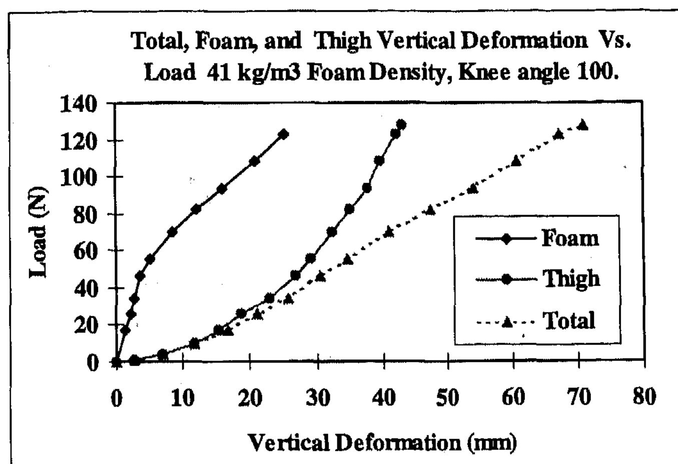


Figure 6. A load-deformation curves (total, thigh, and foam) with 100 degrees knee angle, and 41 kg/m^3 foam density, and 100 mm foam thickness.

To simplify the thigh-foam interaction response, the non-linear load-deformation curve of the thigh and foam was split into two linear parts. The first part was from 0 to 7% of a subject weight, the second part ranged from 7 to 16% of a subject weight. The second part of this curve represented the working region of the thigh stiffness when the thigh interacts with seat cushion in seated posture. This working region was defined based on the contact pressure of the thigh-foam interaction from the results of pressure mat for various people in the seated posture. The straight line for each part represented the experimental data, and a best-fit line was obtained for each part. The lines were highly correlated to the data with correlation coefficients at least 0.92. The slopes of the best-fit line of the thigh deformation in the first and second parts in Figure 7 were 1.3 and 4.4 N/mm, respectively. The stiffness of the second part was higher than the first part, meaning that the thigh became stiffer as the displacement and contact region increased. The stiffness of the foam in the second part was smaller than the first part. The slopes of the best-fit line of the foam deformation in the first and second parts were 11.1 and 3.7 N/mm, respectively.

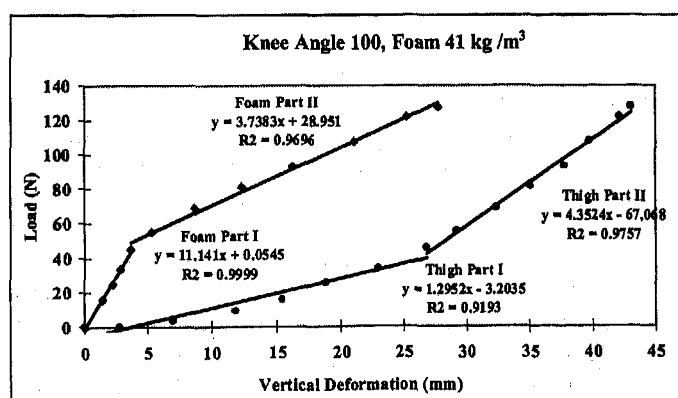
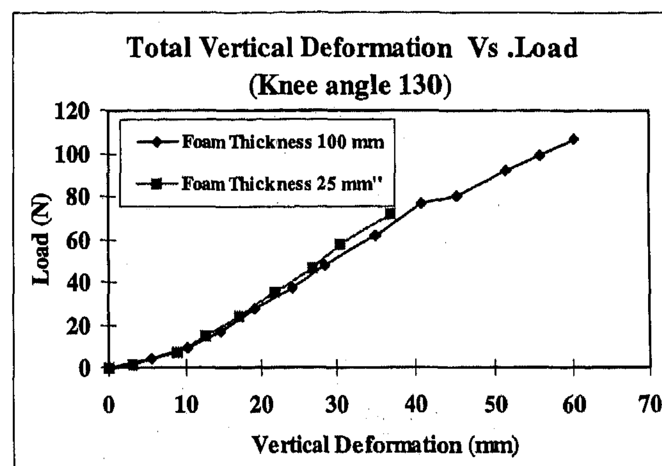


Figure 7. Linearization of thigh and foam deformation curves.

Two foam thicknesses (25 and 100 mm) and two densities (41 and 51 kg/m^3) were selected for the thigh-foam tests. Two different knee angle positions of the subjects were examined. The results of the test showed that the foam thickness did not affect the thigh vertical deformations. (Figure 8). The thigh deformations due to the different foam densities were very close for the same load (Figure 9). Larger knee angle resulted in stiffer thigh response. This change in the thigh stiffness occurred because the tension in the hamstring muscles increased as the knee angles increased. Figure 10 shows the change in the thigh stiffness when the knee angle changed.



Figures 8. Thigh vertical deformation versus load for two types of foam thicknesses.

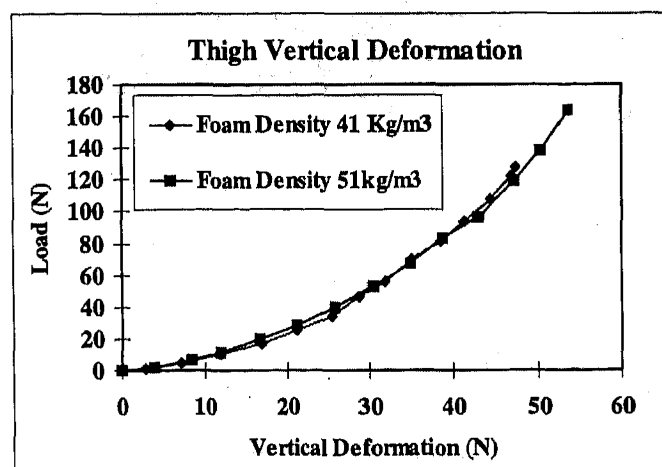


Figure 9. Thigh vertical deformation versus load for two types of foam densities.

The experimental results of load-deformation curves of ten healthy human subjects were used to predict the material properties of the thigh tissues (6). Although it has been reported in the literature that the human tissues are viscoelastic and non-linear, linear elastic material is still in common use. We assumed that soft tissues behave like a rubber material, with the Ogden strain energy potential, which can experience large deformations. Results of ten subjects show that a single

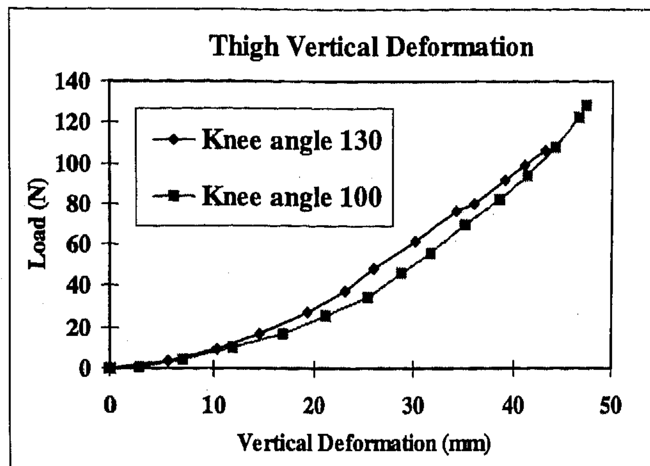


Figure 10. Thigh vertical deformation versus load for two different knee angle positions.

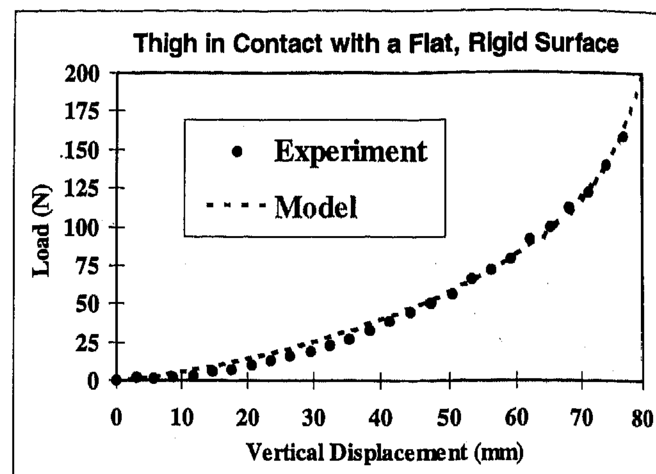


Figure 12. Total deformation of the soft tissues versus load due to a flat, rigid surface.

term of Ogden model is sufficient for representing human thigh tissues. Figure 11 shows a deformation of the thigh tissues due to the displacement of a flat rigid surface.

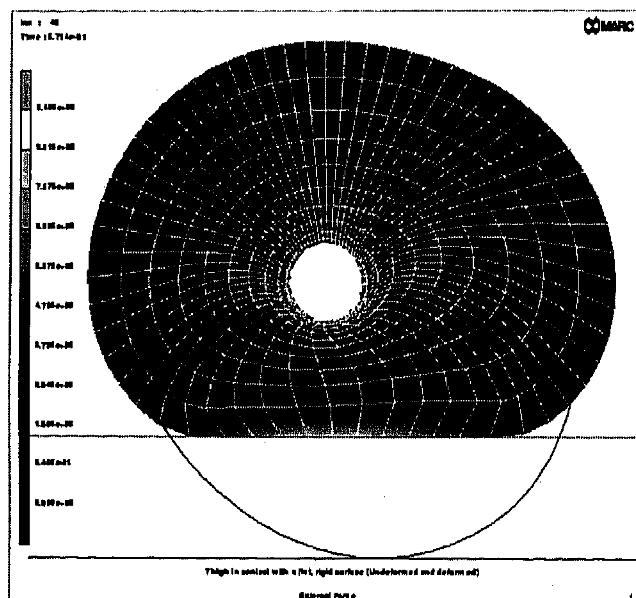


Figure 11. External force distribution of the deformed soft tissues due to a flat, rigid surface

The load-deformation were used as a basis to compare the results of the models and experiments. A load-deformation curve of a thigh deformation due to a flat rigid plate of a model and an experiment for a selected subject is shown in Figure 12. There was a close agreement between a model and an experiment.

The equivalent von Mises stress distribution due to the deformation of the two dimensional model of the mid-thigh in contact with a piece of foam (41 kg/m^3) was shown in Figure 13.

Total deformation of the thigh and a flat rigid surface is actually a deformation of the thigh itself because a flat rigid surface cannot deform. However, the total deformation of a thigh-foam model is a deformation of the thigh and the foam.

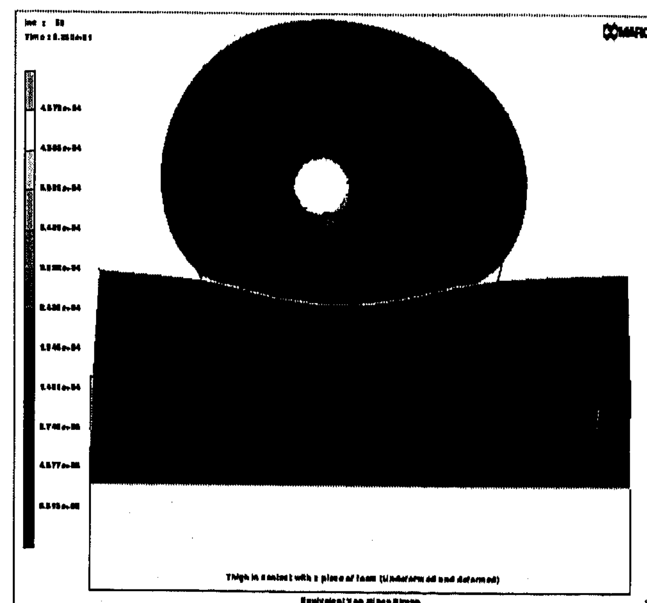


Figure 13. The von Mises stress distribution of the thigh and foam after 50 mm total deformation.

To compare a model and an experiment, a total deformation (thigh and foam deformations) was used. Knowing that the foam thicknesses did not affect the thigh and foam deformations, the 100 mm foam thickness was selected for comparing experiments and models. The results of two different foam densities (41 and 51 kg/m^3) of experiments and models were shown in Figure 14. The results of load-deformation responses of experiments and models for two different foam thicknesses showed good agreements.

The deformations of a flat foam due to the contact with the right thigh were measured through the displacement of the wires. The element sizes of the foam model were chosen in such a way that three element sizes is equal to the distance of the two wires. There were twenty-four wires in the foam, but deformations of two wires in the center of the foam sample were chosen for comparing experiments and models. A contact

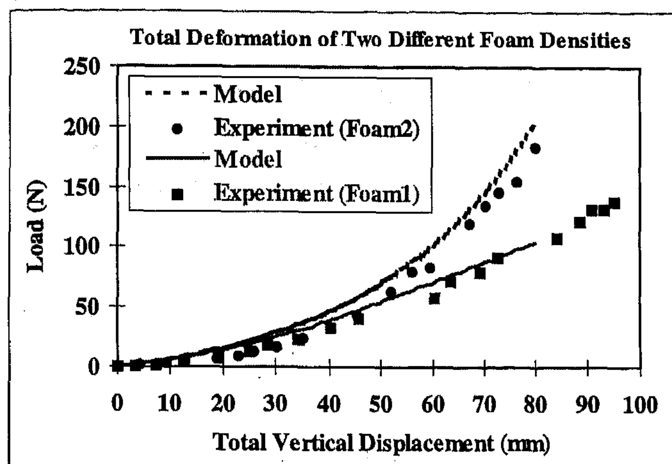


Figure 14. Load-total vertical deformation responses of experiments and models for two types of foam densities, 41 and 51 kg/m^3.

reaction and vertical deformation of the wires were generated in the experiments and models. Figure 15 shows the results of models and experiments for one of the selected wires. There were close agreements between experiments and models for two types of foam densities.

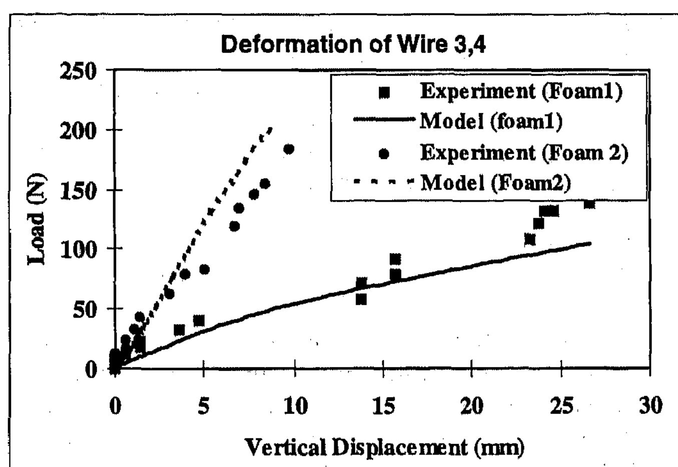


Figure 15. Deformation of one of selected two wires due to two different foam densities, 41 and 51 kg/m^3.

CONCLUSIONS

The cushion materials in a seat and the soft tissues of the human thigh deform when a person sits. These tissue and seat deformations result in the deformed contours between the person and seat. In this study, deformations and contours of human thighs and seat cushion foams were measured and modeled. The measurements were accomplished by transferring the deformed shape of the cushion surface to a video image using wires through the cushion. The contact between the thigh and foam was modeled with finite element methods using large deformation formulations for thigh tissues and foam properties. The results given in the paper are for a single subject, but they are typical of the results for all subjects. As the thigh and foam are loaded, the deformations of the thigh are relatively large and the initial stiffness is low. The initial deformations in the foam are smaller with stiffness that is greater than the thigh. As loading increases, the stiffness of the thigh increases and the stiffness of the foam decreases. The results from measurements and modeling agreed well. This study provides a basis for development of computer models of the entire thigh and buttocks that are in contact with a seat pan cushion. With such a model, seat and tissue deformations could be simulated to predict occupant position. With tissue and seat deformations, contact pressures could be predicted. These results would be useful for correlation to seating pressure patterns that are perceived to be comfortable.

ACKNOWLEDGMENTS

This research was supported in part by the Automotive Systems Group of Johnson Controls Inc.

REFERENCES

1. Fung, Y.C. *Biomechanics: mechanical properties of living tissues.* New York: Springer-Verlag., 1981.
2. Chow, W.W. and Odell, E.I. *Deformations and Stresses in Soft Body Tissues of a Sitting Person.* Journal of Biomechanical Engineering. , Vol. 100, pp 79-87, May 1978.
3. Brunski, J.B., Roth, V., Reddy, N., and Cochran, G.V.B. *Finite element stress analysis of a contact problem pertaining to formation of pressure sores.* ASME: Advances in Bioengineering., 1980.
4. Reddy, N.P., Brunski, J.B., Patel, H., and Cochran, G.V.B. *Stress distributions in a loaded buttock with various seat cushions.* ASME: Advances in Bioengineering., 1980.
5. Todd, B.A. and Thacker, J.G. *Three-dimensional computer model of the human buttocks, in vivo.* Journal of Rehabilitation Research and Development, Vol. 31 No. 2, pp 111-119, 1994.
6. Deng, B. (1994) *Measuring and Modeling Force-Deflection Responses of Human Thighs in Seated Postures.* Master of Science thesis, Michigan State University, Department of Materials Science and Mechanics.
7. Vannah, W.M. and Childress, D.S. (1996). *Indentor tests and finite element modeling of bulk muscular tissue in vivo.* Journal of Rehabilitation Research and Development Vol. 33 No. 3, pp. 239-252.
8. ABAQUS Finite Element Software Packages 5.5, Theory Manual, Hibbitt, Karlsson, and Sorensen Inc.
9. MARC Finite Element Software Packages K6.2, Volume A: User Information, MARC Analysis Research Corporation
10. Schnider, L.W., Pflug, M.A., Snyder, R.G., Anthropometry Of Motor Vehicle Occupant, volume 1.

970597

Redefining Seat Comfort

Wenqi Shen and Alicia M. Vértiz
Delphi Interior and Lighting Systems

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper proposes a new concept towards the understanding of seating comfort. Physical comfort is defined as the physiological and psychological state perceived during the autonomic process of relieving physical discomfort and achieving corporeal homeostasis. In addition to static and dynamic comfort in an automotive seat, the occupant will also experience transient comfort. This is the comfort perceived during posture change or fidgeting on the seat. It may depend most on the material properties if posture requirements are satisfied. Lumbar support is exemplified to illustrate this concept. New seat comfort evaluation tools should be developed to account for the transient properties of the seat supporting system.

INTRODUCTION

Under the restrictions of vehicle packaging, the automotive seat is evaluated through sensory measurements such as seat-occupant interface pressure, comfort and seat feature ratings etc., and through human posture measurements such as lumbar curvature, pelvic rotation, joint angles etc. It is also evaluated through physical measurements of seat systems and/or components such as pressure distribution pattern, force-deflection test, vibration transmissibility etc. Some techniques are only appropriate for research purposes due to either the sophisticated instrumentation or the invasiveness of implementation.

The capability of seating comfort assessment is gained through continuous fundamental and applied research. Basically, two categories of knowledge have been accumulated. One is human sensory aspect in seated posture and it includes psychometric measures of seated comfort/discomfort and psychophysical measures of stimulus induced discomfort. A battery of subjective measurement techniques are available (Corlett, 1989) such as checklists, rating scales, and cross-modality matching. A number of rating scales were evaluated in terms of their validity and reliability for seated comfort assessment (Shen and Parsons, 1997). Research on stimulus-perception relationships such as vibration discomfort (Griffin, 1986) and pressure discomfort (Shen, 1994; Shen, Parsons and Vértiz, 1995; Shen and Vértiz, 1996) provides fundamental knowledge for the understanding of seated comfort and for establishment of industrial testing criteria. The other category of knowledge is the assessment methods and techniques of seated comfort and discomfort. Extensive reviews can be found in Corlett (1989) and Lueder (1983).

For practical evaluation of seat products, more attention has been paid to seat system/component tests than sensory and posture measurements, because posture is more individual and undetermined whereas seat features are more to the interest of product engineers. Indeed, from an engineer's point of view, the physical performance and comfort properties of a seat system is the major concern and they can only be improved through work on seat features and components. Findings from fundamental studies may not directly help engineers solve problems in seat comfort due to at least three reasons as summarized below.

1. Factors such as interface pressure and vibration which above certain thresholds could cause distinctive sensory discomfort, may no longer be significant for automotive seating. Discomfort prediction through psychophysical relationships may become improbable at very low stimulus levels e.g. around discomfort thresholds in automotive seating, owing to the improvement of product quality. With better vehicle dynamic performance and road conditions, the vibration levels that a seated occupant is subject to can be substantially lower than those for a vehicle on the road twenty years ago. Similarly, seat surface contour is becoming more harmonious to promote appropriate sitting posture and conformity with human body shape. This leads to fewer hard spots or regional protrusions which are frequent sources of excessive local pressure. When seats of different design meet basic requirements for comfort and are free of significant sources of discomfort, the sensitivity of comparative evaluation is much reduced.

2. When discomfort due to excessive physical stimuli become minimal, it can be difficult to predict which seat is more comfortable. The most severe component principle (Shen and Vértiz, 1995) states that when there are two or more sources or components of discomfort, overall seated discomfort is determined by the component inducing the highest discomfort. When all discomfort components become less severe, it may be obscure as to which is the true

source or component. Some apparently non-important features may start playing a major role. Moreover, there are other contributors to comfort than the absence of discomfort. As is well known, pressure distribution on seat-occupant interface relates both to occupant seated posture and to seat properties or features, and pressure measurement is one of the readily used tools for seat evaluation. Most of the car seats can provide acceptable support without producing pressure far higher than discomfort threshold. Yet two seats which produce similar pressure distribution may yield quite different comfort ratings even under static conditions. Some intrinsic properties in cushioning materials are essential to comfort perception but yet need profound exploration, apart from better understanding of the meanings of seat comfort.

3. Assessment techniques do not produce immediate solutions to improving seat comfort. Objective techniques, measure sources or covariants of discomfort, and subjective techniques measure levels of discomfort resulting from inappropriate design. In using these techniques, we can only interpret as to what features and properties, and to what direction, we should tackle to possibly improve seats. Take seat cushion material as an example. While we understand that cushion support is provided by the assembly of suspension, padding, trim cover, etc., the properties of padding materials are of great importance due to its unique supporting features. A number of properties of foam material contribute to cushioning comfort as well as to durability, such as density, indentation force/load deflection (IFD or ILD), support factor, resilience, compression set, and dynamic fatigue over time. We do not know for sure what are the relationships between each parameter and seated comfort, although indicative guidelines are available. There are other issues in seating comfort on which controversial opinions exist. Ideally, human factors research on seating comfort should be combined with physical testing in order to identify the problem and produce feasible solutions.

With this paper, we attempt to functionally and operationally redefine seating comfort for better understanding of seating comfort in today's vehicles. A three stage comfort model is proposed to explain subjective judgment of seat comfort as well occupant's sitting behavior in riding. Finally, the properties of lumbar support is discussed as an application of the new comfort concept. We hope this work could inspire further thinking and discussion for better design of automotive seating systems.

PHYSICAL COMFORT: A PROCESS OF RELIEVING DISCOMFORT

Human comfort is not a concept that is precisely defined due to its different attributes in various domains, although it could be considered as "a generic term for a pleasant state of psychological, physiological, and physical harmony between a human being and the environment" (Slater, 1985). We have psychological aspect of comfort of personal well-being and satisfaction, physiological aspect of comfort of being in good health and free of illness, and physical aspect of comfort without negative disturbance from external environments. There are perhaps not many words that are used so often in so many aspects of our daily life as *comfort*. This fact along shows how wide a notion it conveys and how difficult to define it. In the assessment of seats, the lack of agreed definition of seated comfort often causes confusion in both the methods of evaluation and the techniques used for it.

Let's first look at some previous definitions of comfort or seat comfort in literature.

Comfort is some state of well-being or being at ease (Oborne and Clarke, 1973). Similar definitions: Griffin (1995) states that comfort implies a conscious well-being, and Reynolds (1993) describes comfort as an occupant's empirical perception of being at ease. Product and environment design could improve physical aspects to promote to an extent psychological and physiological comfort. Such a definition is very general. We may find a number of words to associate with this broad notion of comfort but none exclusively express it. Obviously, this definition does not denote any means of measuring and/or promoting comfort.

Comfort is the absence of discomfort (Branton, 1969, Corlett, 1973, Hertzberg, 1958). To argue for this definition against positive state of comfort, Branton (1969) stated that he fount it "most difficult to envisage deriving extreme feelings of well-being merely from sitting however good the chair may be. In our view, therefore, the possible continuum extends only from indifference to extreme discomfort." From an operational point of view, Hertzberg (1958) stated that "although we tend to speak of comfort and discomfort as if they were two states of consciousness, for testing purposes it is more realistic to consider that there is only one, discomfort, and that 'comfort' is only the absence of discomfort". According to this definition, one cannot provide comfort in a seat design, one can only eliminate sources of discomfort. Take car seats as an example, as was stated by Verbrugge (1990), "Although automotive engineers and advertisers may hope that a car seat can promote a sense of delight or pleasure, and may want to have research that helps them to enhance this, it is an impossible goal scientifically. A goal of 'indifference' is far more sensible and feasible". This type of definition of "indifference" insinuates that comfort is not measurable or comparable because it is regarded as the starting point of a continuous discomfort scale. We know from experience and from testing that this is not true. One can easily compare comfort sitting in two seats by stating "I feel more comfortable in this seat than being in that one" or "this seat is more comfortable than that one". Of course we agree he/she experiences different levels of comfort. The problem now is: as both seats are not uncomfortable, what is the basis for him/her to compare comfort?

Physical comfort should naturally be related to discomfort, but it is more than absence of discomfort. We propose a definition of physical comfort as follows:-

Physical comfort: the physiological and psychological state perceived during the autonomic process of relieving physical discomfort and achieving bodily homeostasis.

This definition needs some detailed explanation. Comfort and discomfort are not two states which are distinctively separated at two opposite directions. They are not necessarily in the same continuum. But rather, one can instantly suppress the other. As by definition, comfort exists when physical discomfort id reducing. If the removal of severe discomfort is by a leap, comfort can instantly prevail. As an example, think what you may feel when the annoying sound of a power drill suddenly stops.

Comfort is dynamic as is experienced during a process, rather than static. This is a key point to understand comfort. Discomfort can be static and exist for as long as the bodily balance is not assumed, but comfort exists only when some positive change is being made. When no discomfort is present, comfort does not necessarily exist, and it is only "indifference" (Branton, 1969). It denotes a relatively stable state of no awareness of a bad feeling and does not necessarily entail a positive affectation. Therefore, "absence of discomfort" does not define comfort. Comfort only occurs when the environment around you makes or is made a positive change. When you take a shower after an intensive treadmill exercise, the sticky sweat and fatigue is being washed out and you feel a sensation of comfort. Yet soon you will be indifferent and the level of comfort does not sustain even if the thermal condition in the room remains neutral.

This is to say feelings of comfort may gradually saturate or even disappear some time after discomfort is eliminated and homeostasis is reached. Fisher and Connolly (1970) defined physical comfort as a state in which man's physiological needs have been satisfied and homeostasis exists between the internal and external environmental forces. According to the Webster's Dictionary (1984), homeostasis is a state of equilibrium between different but interrelated functions or elements, as in an organism or group. In physiology it concerns the self-regulation of body functions. According to this concept, when there is no discomfort, a man should be always "comfortable". This may not be the case in reality. We know that human beings are stimulus seekers and that human sensory functions work as cycles of excitation and adaptation. When the body is in ease and in homeostasis for a period of time, external excitations are gone and, as excitation induces sensation, the sensation of comfort is also gone. Reporting comfort is an evaluative experience and involves a lot of attention to judge and report the intensity level. When nothing in the environment is being changed, there is no fresh stimulus.

This phenomenon has its psychophysical interpretation. No matter what factors induce stimulus to seated person, the perception of comfort / discomfort always follows similar characteristics to other human sensory modalities. According to the adaptation level theory (Coren *et al.*, 1979), a stimulus is not simply weak or intense, it is weak or intense when judged against the subjective adaptation level, which is established by pooling the effects of three classes of stimuli: focal, background, and residual. If the subject's attention is directed to the background, the identity of the background is shifted to focal stimulus. This context change can affect our perceived intensity. Residual stimuli, the stimuli that occurred in the past and are no longer physically present in the judgment situation, can also affect the perception of sensory magnitude. In the case of seat comfort perception, the focal stimuli are from those seat features that significantly attract the occupants attention for being pleasant or unpleasant. The background stimulus is the general seated environment including seated posture and contact pressure. The identity of focal and background stimuli is not fixed and, as soon as some background features become prominent, the background and the focal shift positions. For example, some background factors, such as sustained medium pressure under the thigh or persistent low back muscle exertion, may become focal stimuli and cause discomfort in a long period of sitting.

When some seat features have already induced discomfort, the elimination process of discomfort will be the focal stimulus to elicit perception of comfort. Towards the diminishing of discomfort, focal stimulus gradually descends to background, and the perception of comfort also falls into a state of unawareness. Comfort as "indifference" expresses only the state of mind when the process has settled. Corlett (1973) once remarked that comfort "can be seen as arising from the summation of all the bodily sensations, each having a random distribution of levels. In any given situation, if the balance is such as not to draw attention to any one sensation, a person can be said to be comfortable". What he denotes may still be the state of having no discomfort, but not the comfort as a capability or a potential of removing discomfort.

To summarize, perceived comfort relates both to the level of discomfort and to the elimination process of discomfort. This approach can be described with hypothetical time courses of changing discomfort and comfort, as shown in Figure 1. At Phase I when discomfort is severe, comfort is at the lowest. At Phase II, discomfort level is being reduced, and comfort increases due to the less severe level and to the reduction rate of discomfort. Phase III is when discomfort is removed but perception of comfort adapts, possibly to an "indifference" saturation level. At Phase IV, some insignificant feature may become the focal stimulus and discomfort creeps up to attract occupation attention, so that perceived comfort drops. Phase V represents another cycle of discomfort removal.

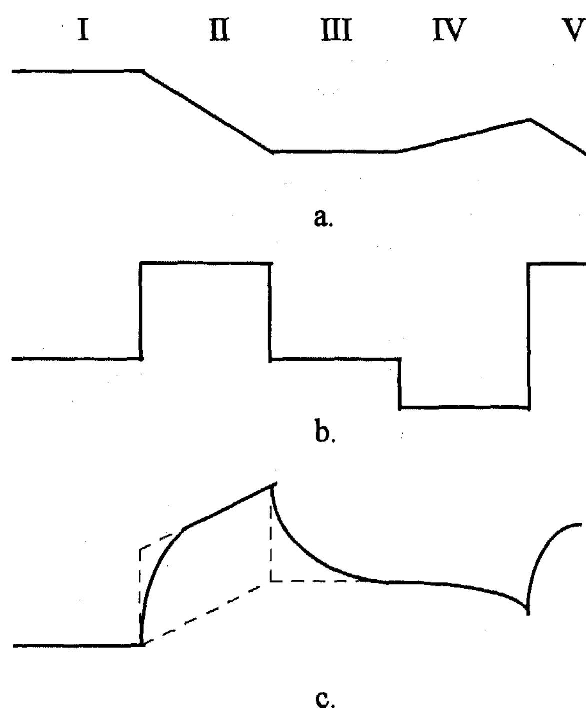


Figure 1. Time courses of comfort. (a) time course of discomfort, which serves as the stimulus to elicit comfort; (b) rate of reduction of discomfort; (c) time course of comfort, as a delayed response function of both discomfort level and the rate of its decrease.

THREE MODES OF SITTING COMFORT VS. CUSHION PROPERTIES

The concept of temporal comfort has its important implication to the design of seats, vehicle interiors and other consumer products. Experience tells us that just as any other tasks, sitting is a naturally fatiguing process. No matter how good a seat or a vehicle is, it cannot maintain an occupant free of discomfort for hours of sitting. And this is a biological mechanism to protect the human tissues from compression damage. If a product is to excel, it has to provide effective ways to minimize the accumulation of discomfort and help remove it. Careful consideration should be given to provide features for soothing and relieving discomfort. With such an active approach towards comfort, we can suppress some of the factors that cause occupant discomfort.

The temporal comfort model also broadens the scope for seat evaluation techniques. In retrospect, current evaluation tools for seat comfort are actually for discomfort assessment. Generally, those tools include interface pressure distribution, postural indicators of muscle activity and spinal shrinkage, and thermal properties for static evaluation, as well as vibration for dynamic evaluation. When anything drifts away from relevant criteria, the occupant may experience discomfort. However, absence of discomfort at a particular rating moment does not guarantee sitting comfort. According to Figure 1, we have to evaluate a seat in terms of its capability of relieving or eliminating discomfort, in addition to testing the absence of discomfort. Cavender (1993) stated that "the very best time to look at properties such as load bearing, resiliency, load ratio, etc. is during the period of loading or the time that work is actually being done to flex the cushioning material". We feel that there are some seat features which may promote comfort and, for lack of proper terminology, we call them transient features. If a seat provides proper postural support and occupant protection, the judgment of comfort is often made at the moments when the occupant is being seated and/or when he/she is seeking change of posture by fidgeting or re-positioning. He/she may sit on the seat, bouncing around and compressing and twisting the body on the seat, a ubiquitous practice when someone is asked to give a quick rating on seat comfort. Unfortunately, no valid tools or techniques are available to measure this kind of phenomenon except some observational or behavioral methods for sitting behavior and postural change (Corlett, 1989). These behavioral techniques have not been proved as practical. Objective tests either on human physiological responses or material properties can be more appropriate to describe such transient stage of comfort.

To better understand the transient features, we can classify the process of ride comfort into three modes: initial comfort, transient comfort, and dynamic comfort. The following explanation will be made in respect to Figure 2, a hypothetical load-deflection curve of a seat cushion assembly.

INITIAL COMFORT. The seat should provide a gradual, gentle, cozy support to the "landing" or indenting of occupant's body upon seat cushion (Pham *et al.*, 1987, and Stone, 1965). The initial softness and sensation of being gently but progressively cushioned is very important to occupant's perceived comfort. To properly support riding postures, the cushion should further produce increasing resilience to the body load indenting the seat, a feature well held in polyurethane foam. This property in foam padding may be expressed by SAG-factor (65% IFD/25% IFD, ASTM Standard D 3574 - 91), and the initial local modulus of the load-deflection curve. In terms of this principle, the local modulus at (D1, L1) of Phase I, Figure 2, should not exceed that of Phase II which is the normal supporting phase. The local modulus is the gradient of the curve at a given deflection under load. Stone's work (1965) concluded that the two seats with low initial local modulus were apparently more comfortable and preferred than the two seats with the highest initial local modulus. ASTM Standard D 3574 - 91 even includes a Modulus Irregularity Factor (MIF), derived from load-deflection curve, and indicates that high, positive MIF value may result in some discomfort in seat applications. As an example, the situation depicted in Figure 2 obviously fails this requirement. Of course, the presumption is there are no significant problems with other seat features such as contouring. On seat-occupant interface, body pressure distribution may examine the supporting quality by identifying hard points and inappropriate posture.

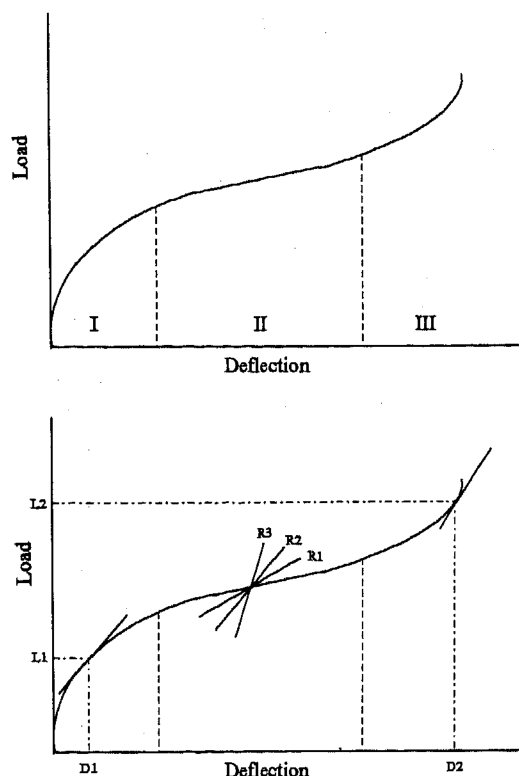
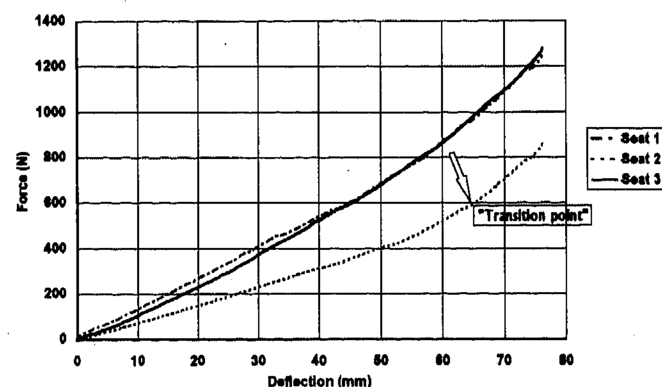


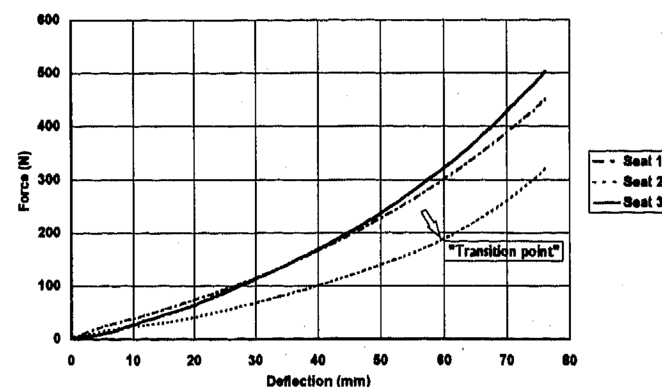
Figure 2. Seat force-deflection curves. Left: three stages of deflection. Right: local modulus at (D1, L1) and (D2, L2), and dynamic modulus at vibration rates $R1<R2<R3$.

TRANSIENT COMFORT. As mentioned earlier, transient comfort is perceived as the subject adjusts posture or fidgets around on the seat. Riding tasks often require frequent change of posture, so do the muscle groups when some relaxation is in need. A comfortable seat should not constitute any local impact or interfere with such changes, but provide gentle damping effect on the transient loading. This again relates to the local modulus on the basis of settlement deflection, the deflection at static body load. The hardness measured as Indentation Force Deflection (IFD) value at 40% or 50% is the mean modulus for up to 40 or 50% of total deflection. Therefore all IFD stands for is the accumulated "dynamic/ local" IFDs. After the occupant sits down and the cushion settles, neither the initial local modulus nor IFD at 40% affects perception, but the local modulus at the settled deflection starts to influence the perception of cushion comfort. The amount of settlement deflection depends mainly on occupant's body weight as well as IFD value. The majority of automotive seats operate at a settlement deflection of 15 - 50% (Cunningham *et al.*, 1994), at the second stage on the load-deflection curve. However, a large and heavy occupant may indent the seat more than 50% close to the Phase III when the cushion hardens up or "bottom-out". A small occupant may indent less than 15%, and in the case shown in Figure 2, the seat would also feel harder. To evaluate seat performance for supporting a range of occupants of different body weight, we may need more comprehensive indices than a single IFD value. Although SAG factor can generally indicate the effectiveness of support, the local modulus for supporting 5% female and 95% male weight should be derived on load-deflection curve to ensure the settlement deflection points lie well in Phase II.

Figure 3 describes results of a force-deflection test on three automotive seats, using (a) a buttock-form indentor and (b) a three-inch-diameter flat indentor, simulating the vicinity of the right ischial tuberosity. Seats 1 & 3 have no apparent differences between stages and were subjectively perceived as being slightly too hard. Seat 2 has low modulus before 65 mm deflection, the "transition point", and then the modulus increases to a level similar to that of the other two seats. This feature provides excellent initial and transient comfort if the cushion deflection is less than 65 mm under the occupant's body weight. Yet the support is effective without bottoming-out as the modulus increases after 65 mm deflection. Figure 3b indicates that, in order to ensure initial and transient comfort, the weight taken by each ischial tuberosity should be less than 200 Newton. Since the ischial tuberosities take the majority of body weight in sitting, this load limit to the region is important to avoid the numbing feeling of the buttocks.



(a) Using buttock-form plate indentor



(b) Using a three-inch-diameter indentor at buttock-pocket

Figure 3. Force-deflection test for three seats.

DYNAMIC COMFORT. Seating comfort under vibration conditions can be considered as the transient comfort when the imposed loading upon static settlement deflection is either cyclic or random. The local modulus or rather, dynamic modulus, depends partially upon the vibration rate, or frequency. The higher the frequency, the higher the dynamic modulus (Figure 2). However, transmissibility measurements of real vehicle seats show high levels of vibration isolation above 20 Hz. Actual vibration sensitive frequency is between 0 and 6 Hz. For this frequency range, reducing seat dynamic modulus is very important for achieving vibration comfort. However, measurement of such modulus may have technical difficulties.

Physical properties of seat supporting system is not as simple as a single IFD value. The three modes of seat comfort suggest for other more potent indices. It is likely these indices would be most related to cushioning material tests. This is an exciting area for comfort enhancement but more knowledge is needed on test methodologies.

MYTH OF LUMBAR SUPPORT

Lumbar support is another seat feature that is of vital importance in seating comfort. A proper amount of lumbar lordosis is fundamental for maintaining the torso upright or reclined and the eyes at the right vision level. Without lumbar support, back muscles have to work to keep the potentially forward-rotating rib cage upright and overcome the weight. If the lumbar support is appropriate, the lumbar spine will shape into natural lordosis, which helps the pelvis rotate forward. Moreover, a secured lumbar support carries a large amount of torso load from above the rib cage, and the lumbar support actually serves as a second pivot point, on which the rib cage is caught safely and the abdomen of the occupant opens up. This action in turn enables the occupant's shoulder to be supported on the upper back and each region on the back can obtain adequate support. A lumbar support also provides stability of the pelvis.

Proper lumbar support can reduce lumbar spine loads as measured by intradiscal pressure. Based on a series of physiological studies, Andersson *et al.* (1979) advocated that lumbar prominence should be 50 mm. However, this amount of lumbar prominence has not been validated as generally achievable and preferable. Porter and Norris (1987) used a wooden test seat to gauge the preferred lumbar prominence and found that 20 mm prominence was preferred to 40 and 50 mm with both reclined and vertical back angles. Incidentally, Dowell (1995) measured 773 seated persons at a vertical back angle and obtained the apex depth of the lumbar curve. Mean lumbar depth was 25 mm for male and 22 mm for female. We also found with our routine seat tests that large amount of lumbar prominence is not actually favored as many people would think. A 20 mm prominence is about the right amount. It seems that the purpose of lumbar support is more for catching the rib cage and maintaining a relatively stable back posture and pelvis than being for a lordosis itself. After all, the lumbar lordosis in a seated posture will be reduced as compared to a standing position by as much as 38° on average(Andersson *et al.*, 1979), due to the tendency of backward rotation of pelvis under hamstring muscle tension.

Preference of less lumbar prominence is related to how the lumbar support is engineered. We often see people test seat back comfort through moving around and swinging the upper trunk fore-aft. If the seat back supports the human back properly along the spine and is still compliable, it may feel comfortable. During actual vehicle riding, it is only too often for the occupant to move slightly adjusting posture and relieving muscle activities. If the lumbar prominence is achieved through a rigid mechanism or hard padding protrusion, the lumbar would stick forward on the occupant's back and cause pressure discomfort. Occupant's back is literally pushed forward so that other back regions cannot maintain close contact with the seat. We experienced many situations when subjects liked the lumbar prominence but the mechanism to achieve it was objectionable.

Therefore a delicate balance should be reached between the prominence and compliance, among other relations in the total seat system. Greater prominence may be preferable if it is achieved through soft padding and flexible lumbar mechanism and suspension. A number of lumbar adjustment systems are commercially available. Unfortunately, when prominence is increased, compliance tends to reduce. A close examination of seat-back load-deflection property should be carried out. This property may change dramatically by using different lumbar adjustment mechanisms. To date, most effort in lumbar support has been on the amount of prominence, and the regional force-deflection property has hardly been investigated.

CONCLUSIONS

Acknowledging that seat comfort is a temporal process rather than a static condition, the authors of this paper proposed a new definition of seat comfort. The approach treats comfort as the reliving process of discomfort instead of being simply the absence of discomfort. The philosophy also leads to a three-modes seat comfort model including initial, transient, and dynamic comfort. Some cushion material properties are analyzed and the lumbar support is illustrated to exemplify the concept. The following aspects can be summarized as conclusions:-

1. As to an automotive seat, comfort is not a static property but a potential of relieving the occupants of any posture and contact associated discomfort. A seat free of discomfort stressors is not necessarily a comfortable one but it must have comfort enhancers.

2. As to an occupant, comfort is not a sustaining feeling accompanying a fixed seated posture, but it is a sensation perceived when discomfort is relieved during transient periods of sitting, e.g. landing on the seat, posture change, and fidgeting.

3. To be comfortable, a seat should provide be capable of providing initial and transient comfort. It should comfortably accommodate posture change and the changed posture. Lumbar support should especially be compliable for adaptation to different lumbar prominence preferences.

4. New seat comfort evaluation tools should be developed to account for the transient properties of the supporting system.

ACKNOWLEDGMENTS

Thanks are due to Raghu Gurram and John Bruce for conducting the force-deflection test.

REFERENCES

Andersson, G. B. J., Murphy, R. W., Örtengren, R., and Nachemson, A. L. (1979) The influence of backrest inclination and lumbar support on the lumbar lordosis in sitting. *Spine*, 4, 52-58.

Branton, P. (1969) Behavior, body mechanics and discomfort. *Ergonomics*, 12, 316 - 327.

Cavender, K. D. (1993) Real time performance testing. *Journal of Cellular Plastics*, 29, 350 - 364.

Corlett, E. N. (1973) Human factors in the design of manufacturing systems. *Human Factors*, 15, 105 -110.

Corlett, E. N. (1989) The evaluation of industrial seating. *Ergonomics*, 32, 257 - 269.

Coren, S., Porac, C. and Ward, W. M. (1979) *Sensation and Perception*. New York: Academic, pp. 56 - 60.

Cunningham, A., Huygens, E. and Leenslag, J. W. (1994) MDI comfort cushioning for automotive applications. *Utech Conference*, Den Haag, paper 5.

Dowell, W. R. (1995) An estimation of lumbar height and depth for the design of seating. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (p. 409-411). October 9-13, San Diego, California.

Fisher, V. G., and Connolly, A. F. (1970) *Promotion of Physical Comfort and Safety*. Foundation of Nursing Series, Brown.

Griffin, M. J. (1986) Evaluation of vibration with respect to human response. SAE Paper No. 860047.

Griffin, M. J. (1995) The ergonomics of vehicle comfort. In *Proceedings of the 3rd International Conference on Vehicle Comfort and Ergonomics.* March 29-31, Bologna, Italy. Paper No. 95A1026, pp. 213 - 221.

Hertzberg, H. T. E. (1958) Seat comfort. In *Annotated Bibliography of Applied Physical Anthropology in Human Engineering* (Edited by H. T. E. Hertzberg). WADC Technical Report 56-30, Wright Air Development Center, Wright-Patterson Air Force Base, Dayton, Ohio. Appendix 1, pp. 297-300.

Lueder, R. K. (1983) Seat comfort: a review of the construct in the office environment. *Human Factors*, 25, 701-711.

Oborne, D. J. and Clarke, M. J. (1973) The development of questionnaire surveys for the investigation of passenger comfort. *Ergonomics*, 16, 855 - 869.

Pham, T., van Assche, J, Watts, A. and Lockwood, R. J. (1987) MDI based flexible foam: A status report on world developments. In *Proceedings of Polyurethanes World Congress 1987*, Aachen, Germany, September 29 - October 2, pp. 532 - 537.

Porter, J. M. and Norris, B. J. (1987) The effects of posture and seat design on lumbar lordosis. In *Contemporary Ergonomics* (Edited by E. D. Megaw), pp. 191-196, London: Taylor & Francis.

Reynolds, H. M. (1993) Automotive seat design for sitting comfort. In *Automotive Ergonomics* (Edited by B. Peacock and W. Karwowski), pp. 99 - 116, London: Taylor & Francis.

Shen, W. (1994) *Surface Pressure and Seated Discomfort*. Unpublished Ph.D. Thesis, Loughborough University of Technology.

Shen, W. and Parsons, K. C. (1997) Reliability and validity of rating scales for pressure sensation and perceived discomfort in sitting. *International Journal of Industrial Ergonomics* (in press).

Shen, W., Parsons, K. C. and Vértiz, A. M. (1995) The effect of stimulus locus on perceived pressure intensity and discomfort in seated thigh regions. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (p. 955). October 9-13, San Diego, California.

Shen, W. and Vértiz, A. M. (1995) Integration model of seating comfort. *Automotive Interiors International*, Summer issue, pp. 64-71. Presented by Vértiz at *95' International "Automotive Interior" Conference*, March 28-29, Paris, France.

Shen, W. and Vértiz, A. M. (1996) Pressure induced discomfort in seated human back. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (p. 1277). September 2-6, Philadelphia, Pennsylvania.

Slater, K. (1985) *Human Comfort.* Springfield, Illinois: Thomas.

Stone, P. T. (1965) An approach to the assessment of the comfort of foam cushioning. *Automotive Body Engineering*, January, 28 - 30.

Verbrugge, L. M. (1990) Older adults and automobile seats. In *Report to General Motors on Biomechanically Compatible Automotive Seating* (Edited by H. M. Reynolds), July, GM Research Lab, Warren, Michigan, pp. 115 - 128.

970598

Limitations in Predicting Human Vibration Response Characteristics From Manikins and Rigid Body Masses

Suzanne D. Smith
Department of the Air Force

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The objective of this study was to use the driving-point impedance and transmissibility techniques to evaluate and compare the resonance behavior of two male humans, the Hybrid III aerospace manikin, and a rigid body mass using a rigid seat and a selected helicopter seat cushion. All occupants represented the 95th percentile or higher of the male population for weight. The results showed that the resonance frequencies associated with the peak impedance, chest, and head transmissibilities were significantly higher in the manikin regardless of the seating configuration or input acceleration level. While the magnitude of the peak impedance was higher in the manikin, differences in the chest and head transmissibilities depended on both the seating configuration and acceleration level. Neither the manikin nor the rigid body were effective in predicting the primary human resonance effects occurring in the sensitive region of 4 to 8 Hz. However, some similarity did exist when comparing the manikin response to human spine transmissibility data. A discussion of an alternative method for predicting human body resonances using mathematical models is included in this paper.

INTRODUCTION

The operation of both civilian and military ground vehicles, including automobiles, tanks, and trucks, can expose the occupants to whole-body vibration which has been shown to affect comfort, performance, and health. Griffin [1] applied the ISO 2631 [2] to seat transmissibility data collected on cars, buses, and trucks operated over several terrain surfaces. For all vehicles, the vertical seat vibration was found to be more severe than vibration measured in the other translational axes. The weighted accelerations were found to be about twice as severe on the poorer surfaces. For the cars, the values reached the reduced comfort boundary [2] within 30 minutes of operation over the poorer surface. The exposure limit, which should not be exceeded according to the ISO 2631 [2], occurred within as little as 4 to 8 hours. Griffin [1] observed that higher weighted acceleration levels and shorter acceptable exposure durations occurred with the operation of buses and trucks. In the military, health hazard assessments of new Army ground vehicles have shown that the exposure limit can be reached within as little as one hour [3]. Pilots of both Army and Air Force Black Hawk helicopters have reported symptoms of low back pain and general discomfort after about four hours of flight [4].

The improvement of comfort, performance, and safety in ground vehicles challenges the commercial and military communities with designing vehicle suspensions and seating systems which minimize vibration effects. Both experimental testing and mathematical modeling provide the methods for evaluating suspension and seating system design concepts for reducing human body vibration, however, both methods must be capable of properly predicting human body resonance effects. For example, vibration testing and analysis methods have used manikins and rigid body masses to evaluate seat and cushion concepts for human use, particularly when human testing was not feasible. Gurram et al. (1995) compared the vibration response of a human and a plastic water-filled dummy of similar weight while sitting in a seat/cushion system. Transmissibility was measured between the output acceleration at the subject/cushion interface and the input acceleration at the shaker table. The seat itself was shown to have a transmissibility near unity. The primary resonance of the human occurred at 4.25 Hz with a secondary peak observed between about 7 and 8 Hz. The single resonance peak observed for the dummy occurred at 7.5 Hz. The magnitude of the response was found to be quite higher in the dummy. Using both the driving-point impedance and transmissibility techniques, Smith (1994, 1996) tested military aircraft cushions using humans weighing between 50 and 105 kgs, and using rigid masses weighing between 39 and 68 kgs. While the primary resonance peak for all of the human subjects occurred between 4 and 7 Hz, the single peak observed for the rigid mass tests occurred between 8 and 15 Hz. The limitations of the manikin and rigid body mass in simulating human body resonances should be well understood before using

these methods to optimize seating systems for human use. The objective of this study was to evaluate and compare the vibration response characteristics of the human, manikin, and rigid body mass of similar weight using a rigid seat and a selected helicopter seat cushion. The driving-point impedance and transmissibility techniques were used to generate frequency response profiles for the whole-body and for specific anatomical structures or regions. The resonance behaviors observed in these profiles were evaluated and compared between the humans, manikin, and rigid body mass. Also included in this paper is a discussion of the mathematical modeling method for predicting human body resonance behaviors and seat cushion effects.

METHODS AND MATERIALS

Four 'occupants' of approximately the same weight were used in the study. Two of the occupants were human males weighing approximately 104 kgs. The third occupant was a Hybrid III aerospace anthropomorphic manikin obtained from First Technology Safety Systems, Inc. and weighing approximately 99 kgs. The fourth occupant was a rigid body mass comprised of bags of metal shot weighing approximately 101 kgs. All occupants represented the 95th percentile or higher of the male population for weight.

An electrodynamic vibration platform (Unholtz-Dickie) was used to provide the vertical vibration exposures. The system was designed to operate in the frequency range of DC to 5000 Hz with a force limit of 6672 N peak and a payload of 113 kgs. A rigid aluminum test seat was mounted on top of the platform and included a seatback, lapbelt, and double shoulder harness. The angle between the seatback and seatpan was 90 degrees. The test seat has been shown to respond as a rigid system for the subject weight and frequency range under investigation. No additional padding was provided for the seated subjects. The humans were loosely restrained, similarly to what would be expected during the operation of civilian ground vehicles. The human subjects used in this study were experienced in vibration testing and fully understood the importance of maintaining the upright posture during data collection. The presence of the seatback aided in this process. It was necessary to restrain the manikin more tightly to maintain the upright posture. Due to the seat payload limitations, both the human and manikin legs were allowed to hang freely. The subjects hands were folded and placed in their lap during testing. The transmitted force of the combined seat and occupant was measured by three load cells (Strainsert FL1U-256 KT) located between the seat and vibration platform. Two accelerometers (Kulite GYD-4-280-20 and Kistler 815A5) were attached to the seat for measuring the input acceleration magnitude and phase, respectively. Vertical accelerations were measured using miniature accelerometers (Entran EGA-125-10D) attached with double-sided carpet tape to specific anatomical regions on the occupant. For the human, accelerometers were attached on the chest (at the level of the manubrium) and on a bitebar (head) molded with dental acrylic. For the manikin, an accelerometer was similarly attached at the chest (on the surface). Manikin head acceleration was measured using an accelerometer mounted within the head. The location was slightly anterior to and above the attachment site for the head and neck. For the rigid body mass, an accelerometer was attached on top of the mass.

In order to evaluate the effects of a cushion on the vibration response of the occupants, tests were also conducted using a seatpan cushion obtained from a Black Hawk helicopter and placed between the rigid seating surface and the occupant. The cushion was fabricated of three layers of foam with different densities. The bottom layer was made of high density plastic foam and varied in thickness from about 1 cm at the back to 6 cm at the front, providing a contour or "bucket" seating surface. Air vents ran from the front to the back along the inside surface. The top layer consisted of polyurethane foam about 2.5 cm in thickness. Sandwiched between these two materials was a layer of polyurethane foam of greater density than the top layer and approximately 1.5 cm in thickness. The cushion was covered with black lambswool and weighed 910.5 gm. For the cushion tests, a ride quality meter, consisting of three orthogonal accelerometers imbedded within a rubber disk, was placed between the occupant and cushion for calculating the interface transmissibility. Only the vertical acceleration was used for the evaluation and comparison.

The three input vibration profiles at the seat included discrete sinusoidal frequencies and two sum-of-sines profiles generated by combining the discrete sinusoidal frequencies. The frequency components used for all three input profiles ranged from 3 to 21 Hz in 1 Hz increments. Two seat acceleration levels were used; 0.59 m/s^2 rms and 2.35 m/s^2 rms (0.06 g_{rms} and 0.24 g_{rms}). While the frequency content and rms acceleration levels were identical for the two profiles, the crest factor (CF= 2.9 and 4.2) was varied by altering the phase relations between frequency components. The crest factor, defined as the ratio of the peak acceleration to the rms level, is an indicator of the impulsiveness of the motion. Figure 1 illustrates the two sum-of-sines profiles (P1 and P2) at 0.59 m/s^2 rms. A computer program was used to generate the vibration profiles at the seat and for simultaneously collecting all transducer data. Data were collected for two seconds at a sampling rate of 1024 Hz. A Fast Fourier Transform algorithm was used to determine the transducer magnitude and phase difference between the sum of the three load cell forces and the input velocity calculated from the input acceleration at the seat. The impedance frequency response profiles were calculated as the magnitude ratio and phase difference between the transmitted force and input velocity at each discrete frequency component between 3 and 21 Hz. The impedance of the rigid seat (collected separately) was subtracted from the calculated impedance to obtain the impedance frequency response profile of the occupant. The magnitudes of the vertical transmissibility frequency response profiles were calculated as the ratios between the accelerations measured at the chest, head, ride quality meter, and rigid body mass, and the input acceleration at the rigid seat at each discrete frequency component between 3 and 21

SUM-OF-SINES VIBRATION SIGNALS

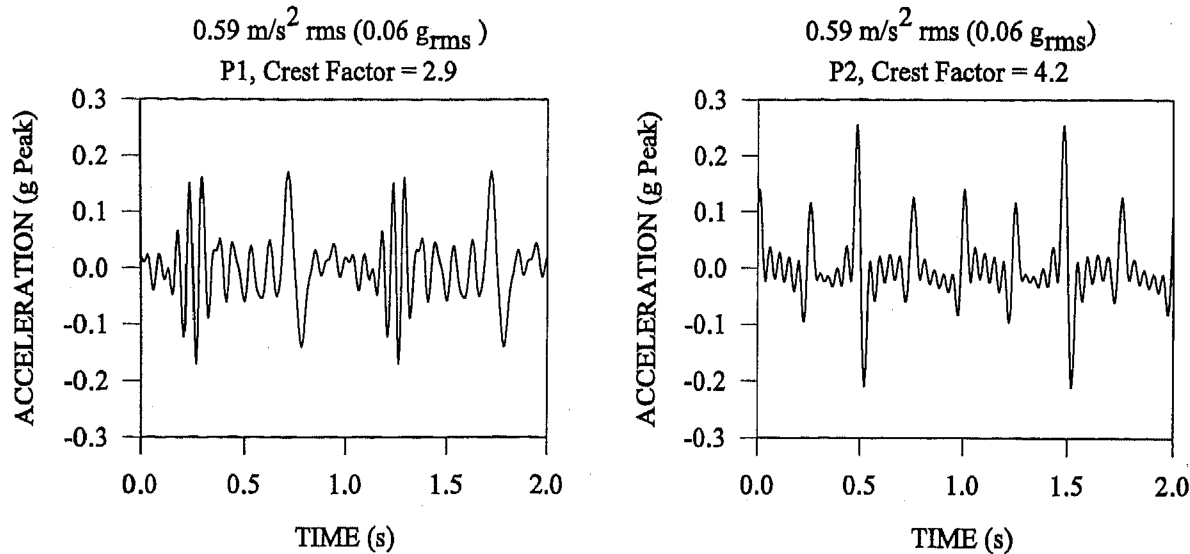


Figure 1 Sum-of-sines vibration profiles.

Hz. The impedance magnitude and phase, and the transmissibility magnitude frequency response profiles were generated at each of the two acceleration levels. The three magnitude frequency response profiles (sinusoidal and sum-of-sines) were used to compare and evaluate the frequency location and magnitude of the peak or resonance responses. The means and standard deviations were calculated for the resonance frequencies and peak magnitudes extracted from these data and represented the subject response variability occurring for different types of vibration exposures. In addition, the mean impedance and transmissibility magnitude frequency responses were calculated from the three profiles.

RESULTS

RIGID SEAT EFFECTS - Figure 2 illustrates the mean impedance magnitude and phase frequency response profiles for the occupants using the rigid seat at the lower acceleration level (0.59 m/s² rms). Figure 3 shows the means ± one standard deviation for the primary impedance resonance frequencies and peak magnitudes occurring at the two acceleration levels for the humans and manikin using the rigid seat (labeled R in the figure) and seat cushion (labeled C in the figure). The seat cushion effects are discussed later in the paper. The impedance response of the rigid body mass was calculated using a mass of 101 kg and the known relationship for the impedance magnitude and phase of a rigid mass [8]. Both humans showed a primary or highest resonance peak between 6 and 7 Hz. Additional regions of resonance were observed at higher frequencies, particularly for the first human subject between 11 and 13 Hz, where a smaller peak was observed in all three magnitude response profiles. The impedance magnitudes for the humans remained relatively constant at frequencies

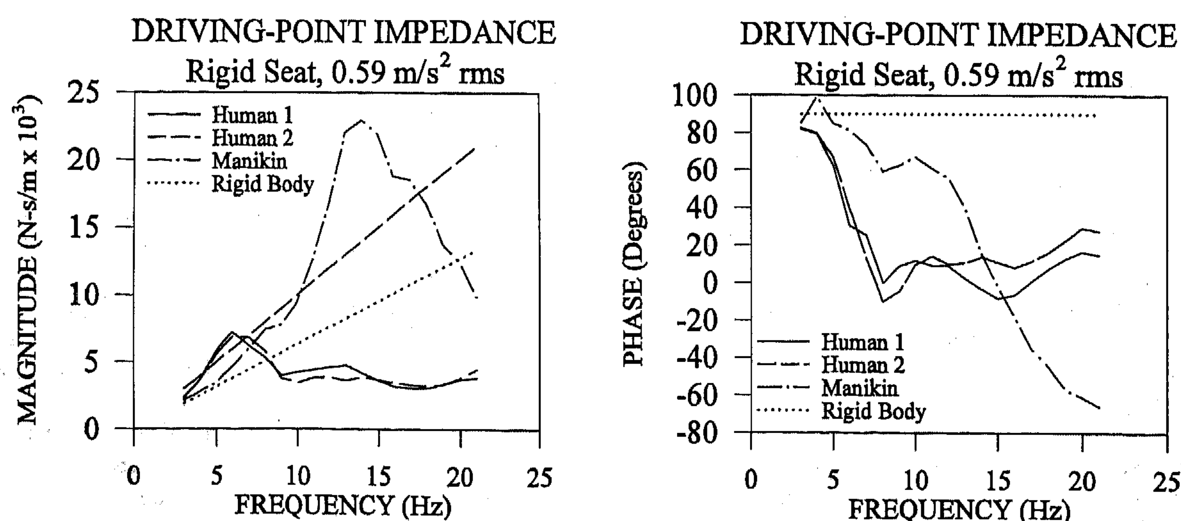


Figure 2 Driving-point impedance magnitude and phase frequency response profiles using the rigid seat.

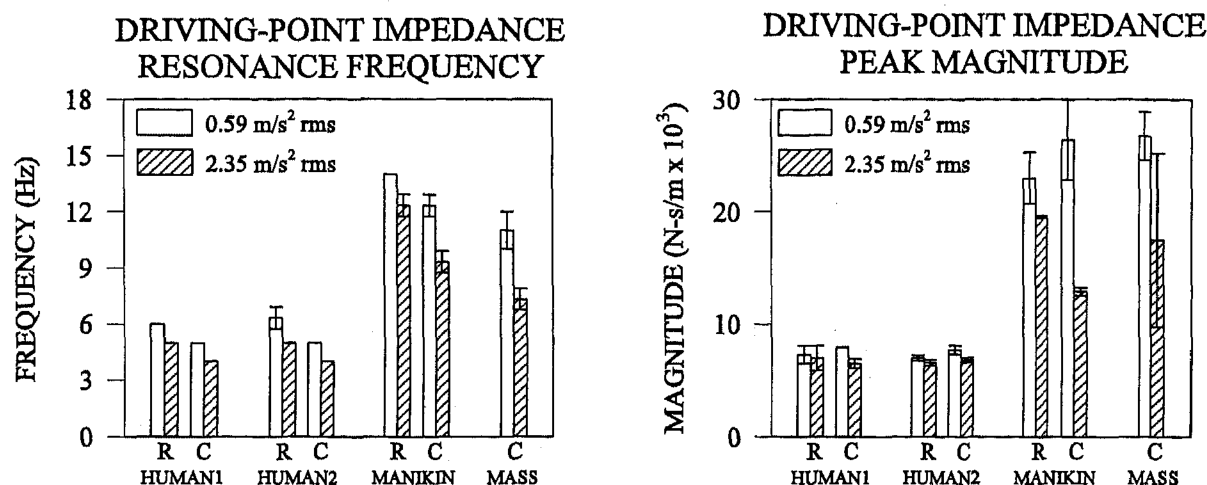


Figure 3 Driving-point impedance resonance frequencies and peak magnitudes (R = Rigid Seat; C = Cushion).

beyond the regions of resonance. For the humans, the phase declined rapidly through the primary region of resonance from about 90 degrees to around 0 degrees. There was a phase change through the second region of resonance observed for the first human subject as illustrated in Figure 2. The phases for the humans tended to remain within ± 20 degrees at frequencies beyond the primary peak. In contrast, the primary or highest impedance peak observed for the manikin occurred at a higher frequency around 14 Hz and was approximately four times the magnitudes of the human peaks. In addition, the majority of the manikin data showed smaller magnitude peaks at 8 and 17 Hz. As shown in Figure 2, the impedance magnitude decreased with increasing frequency beyond the primary region of resonance. The phase changed rapidly through the peak regions. For the manikin, the phase continued to decline at frequencies beyond the primary resonance (14 Hz) to between -60 and -70 degrees at 21 Hz. Figure 2 also shows the calculated magnitude and phase responses for the rigid body mass. Figure 3 shows that the primary impedance resonance frequency observed for both the humans and manikin decreased by 1 to 2 Hz at the higher acceleration level with the use of the rigid seat (labeled R in the figure). The two human subjects showed a peak at 5 Hz, while the manikin showed a peak around 12 Hz. The humans showed no significant difference in the magnitude of the primary peak at the higher acceleration level. The manikin did show a decrease in the primary peak at the higher acceleration level although less variability was observed in the data at this level. In contrast to the effects observed in the humans and manikin, the impedance of a rigid body mass is not affected by the input acceleration level. For both the humans and the manikin, the additional peaks observed in the impedance frequency response profiles appeared more attenuated or were shifted by several Hz at the higher acceleration level.

Figure 4 illustrates the mean magnitudes of the chest and head transmissibility frequency response profiles for the humans and manikin at the lower acceleration level using the rigid seat. Figure 5 shows the means ± one standard deviation for the primary resonance frequencies and peak magnitudes associated with the chest transmissibility, while Figure 6 shows the means ± one standard deviation for the primary resonance frequencies and peak magnitudes associated with the head transmissibility at both acceleration levels using the rigid seat (labeled R in the figures) and seat cushion (labeled C in the figures). The seat cushion effects are discussed later in the paper. The frequency location of the primary or highest chest and head transmissibility peaks coincided closely with the primary resonance frequency observed in the impedance data for both the humans (peaks occurring between 4 and 6 Hz) and the manikin (peaks occurring between 12 and 15 Hz) at the respective acceleration level. As shown in Figures 4 and 5, the magnitude of the peak chest transmissibility observed for the manikin was significantly higher as compared to the human at the lower acceleration level. Figures 4 and 6 show that the magnitudes of the peak head transmissibilities were more similar between the humans and manikin at the lower acceleration level. As depicted in Figure 4, the first human subject showed a relatively pronounced secondary peak between 12 and 15 Hz in both the chest and head transmissibility frequency response profiles, while the manikin showed a consistent peak at about 8 to 9 Hz. The secondary peaks occurred at frequencies which coincided closely with the secondary resonance peaks observed in the impedance frequency response profiles for these occupants. For the rigid seat, increasing the acceleration level had variable effects on the frequency location and magnitude of the primary resonance peaks for the chest and head transmissibilities. As shown in Figure 5, both the humans and the manikin showed a tendency for a decrease in the primary chest resonance frequency with the increase in acceleration level. The second human showed an increase in the magnitude of the peak chest transmissibility, while the manikin showed a decrease in the peak chest response. The first human showed no significant change in the magnitude of the peak chest response given the variability in the data. Figure 6 shows that both humans showed a slight tendency for a decrease in the head transmissibility resonance frequency while the manikin showed a larger decrease in the head transmissibility resonance frequency with the increase in acceleration level. Both humans showed a significant

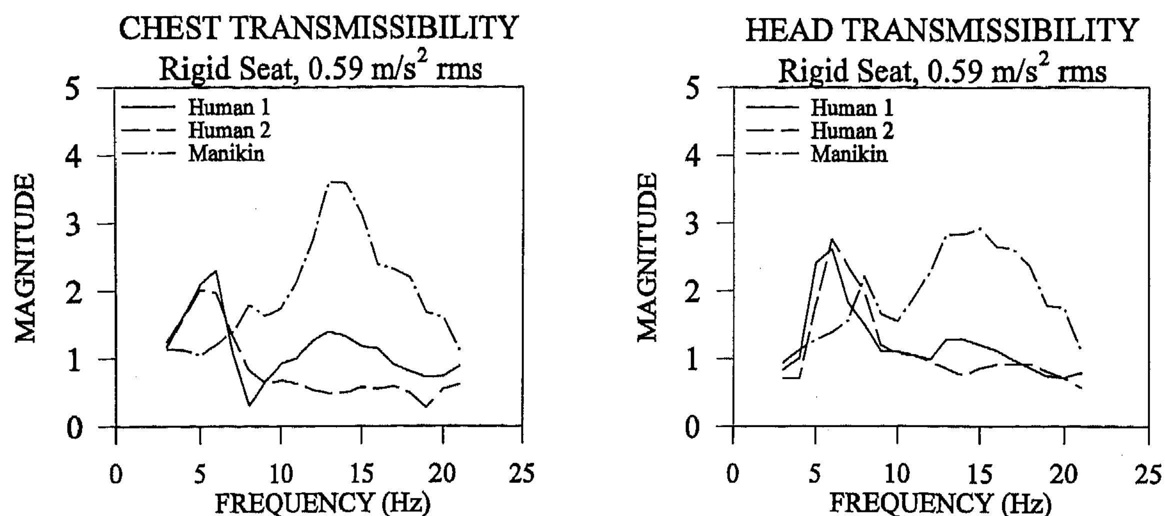


Figure 4 Chest and head transmissibility frequency response profiles using the rigid seat.

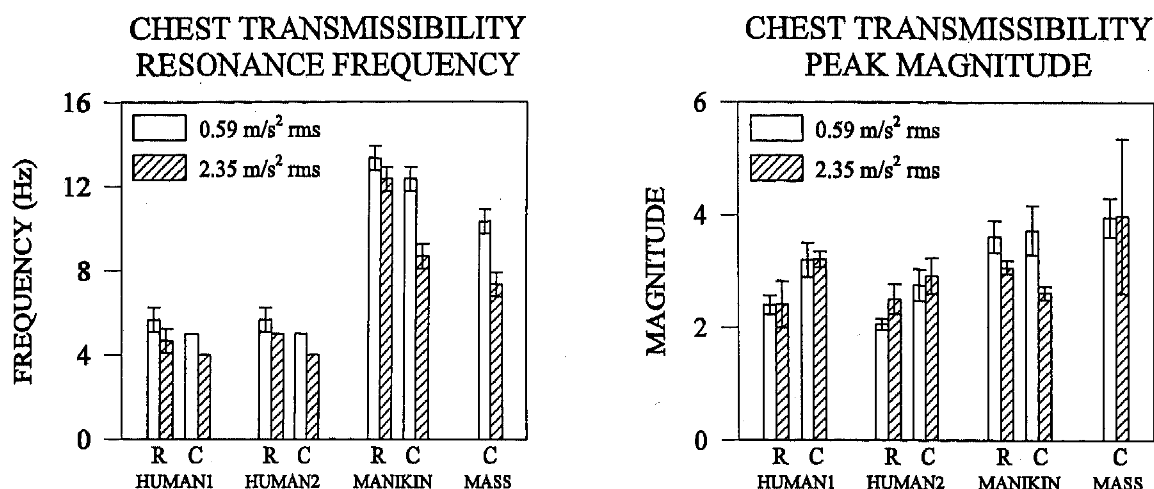


Figure 5 Chest transmissibility resonance frequencies and peak magnitudes (R = Rigid Seat; C = Cushion).

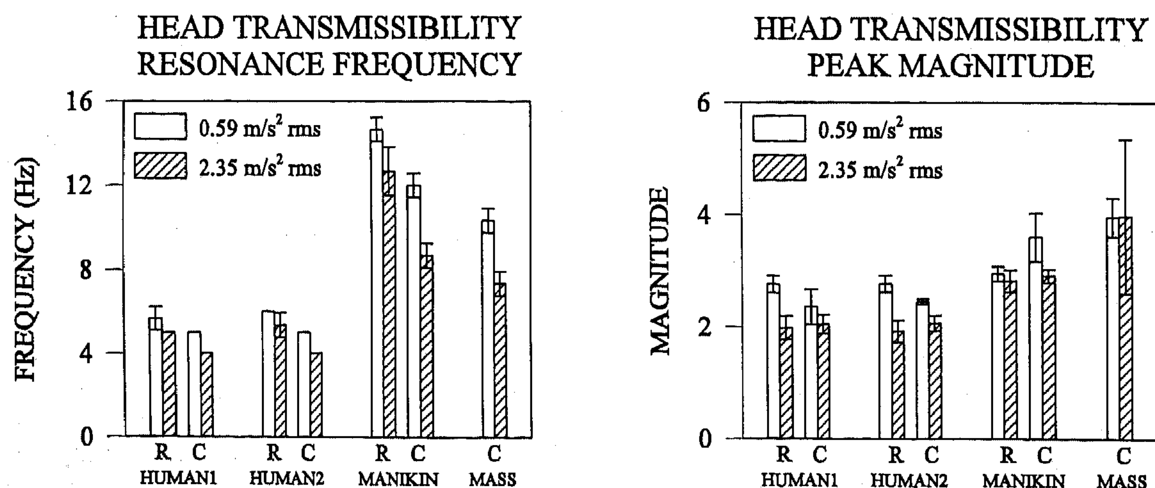


Figure 6 Head transmissibility resonance frequencies and peak magnitudes (R = Rigid Seat; C = Cushion).

decrease in the magnitude of the peak head transmissibility at the higher acceleration level, while the manikin showed similar peak responses at both acceleration levels.

For the rigid seat in general, the peak chest transmissibilities were higher in the manikin as compared to the human at the lower acceleration level, but more similar between the humans and manikin at the higher acceleration level. In contrast, the peak head transmissibilities were similar

between the humans and manikin at the lower acceleration level, but higher in the manikin at the higher acceleration level. The transmissibilities measured at the rigid seat interface were also unity for all occupants, i.e., the interface acceleration was equal to the input acceleration with the use of a rigid seat and no intervening cushion.

Hz). The associated magnitude peaks were also decreased for all occupants at the higher acceleration level with the most pronounced differences occurring for the manikin and rigid body mass. Large variability was observed in the rigid body data at the higher acceleration level. This may have been due, in part, to the relatively high acceleration peaks inherent in the sum-of-sines profiles at the higher level in combination with the cushion effects.

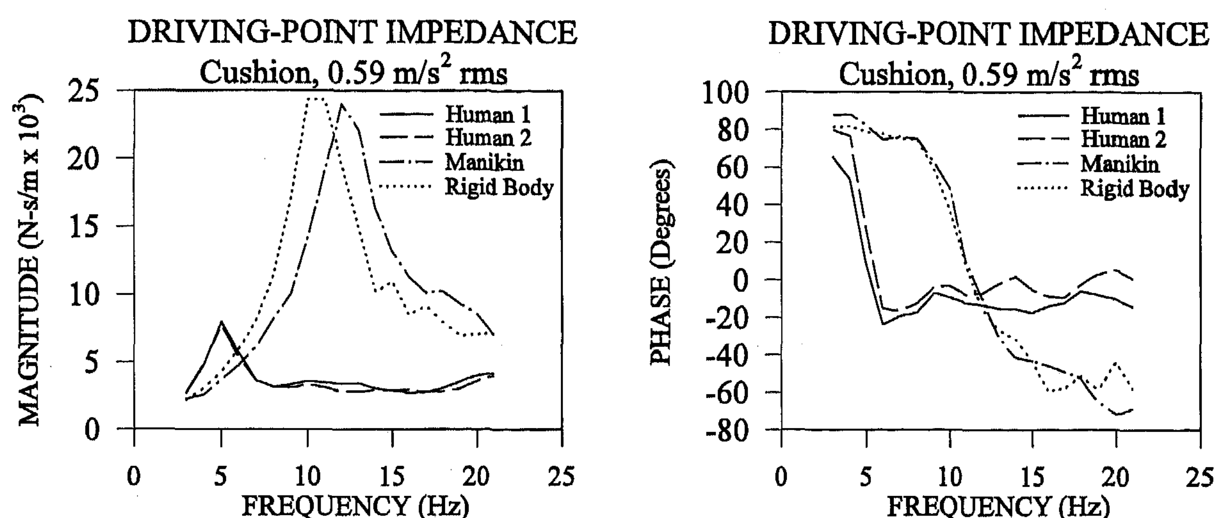


Figure 7 Driving-point impedance magnitude and phase frequency response profiles using the cushion.

SEAT CUSHION EFFECTS - Figure 7 illustrates the mean impedance magnitude frequency response profiles for the humans, manikin, and rigid body mass with the use of the cushion at the lower acceleration level. Figure 3 includes the mean resonance responses of the rigid body mass using the cushion. Figure 7 shows that, with the addition of the cushion located between the rigid body mass and the rigid seat, a prevalent impedance peak is observed at about 10 to 11 Hz at the lower acceleration level. Both the resonance frequency and peak impedance magnitude for the rigid body mass coincided more closely to the manikin data, with both showing a significantly higher resonance frequency and peak magnitude as compared to the human subjects, similar to the results observed using the rigid seat. Any additional resonance peaks appeared to be more attenuated with the use of the cushion for both the humans and manikin. As depicted in Figure 3, the primary impedance resonance frequency was decreased in both the humans and manikin with the addition of the seat cushion (labeled C in the figure) regardless of the acceleration level. The maximum decrease was 3 Hz. The addition of the seat cushion appeared to have a minimal effect on the peak impedance magnitude for the humans regardless of the acceleration level. Given the relatively large variability in the data, the addition of the cushion also had a minimal effect on the peak response of the manikin at the lower acceleration, however, the peak response was significantly reduced with the use of the cushion at the higher acceleration level. The acceleration effects observed with the use of the cushion were similar to the effects occurring for the rigid seat. The peak resonance frequencies were reduced by 1 Hz in the human with larger decreases being observed for the manikin and rigid body mass at the higher acceleration level (up to 4

Figure 8 illustrates the mean magnitudes of the chest and head transmissibility frequency response profiles for the humans and manikin using the cushion at the lower acceleration level. The figure also includes the mean transmissibility frequency response measured for the rigid body mass with the addition of the cushion. Figures 5 and 6 include the mean resonance responses of the rigid body mass with the use of the cushion. Figure 8 shows that the chest and head resonance frequencies and peak transmissibility magnitudes for the manikin were more similar to the responses observed for the rigid body mass. At both acceleration levels, the primary chest resonance frequency tended to decrease for the humans while the associated magnitude peak showed an increase with the use of the cushion as compared to the rigid seat. The manikin also showed a decrease in the resonance frequency with the use of the cushion. The decrease was more pronounced at the higher acceleration level. For the manikin, the peak chest transmissibility did not change significantly with the use of the cushion at the lower acceleration level given the variability in the data. At the higher acceleration level, the peak chest response tended to decrease with the addition of the cushion, in contrast to the results observed for the humans. The primary head resonance frequency showed a tendency to decrease with the use of the cushion for both the humans and manikin at both acceleration levels. The decrease was more pronounced in the manikin. Given the variability in the data, the magnitude of the peak head transmissibility tended to decrease for the humans and increase for the manikin at the lower acceleration level with no significant changes observed at the higher acceleration level with the addition of the cushion. For all occupants, the

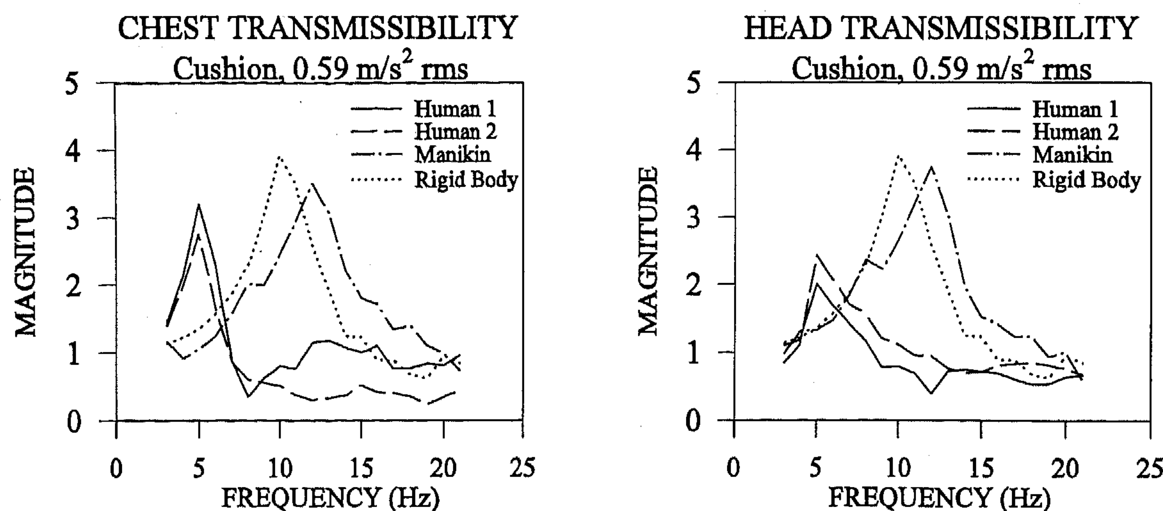


Figure 8 Chest and head transmissibility frequency response profiles using the cushion. Both graphs include the rigid body transmissibility frequency response profiles using the cushion.

effects of acceleration on the chest and head resonance frequencies were similar for the rigid seat and cushion showing a tendency to decrease at the higher level. The effects of the acceleration level on the magnitude of the peak chest transmissibility were minimal for both the rigid seat and cushion for the first human. The second human showed no significant effect of acceleration with the use of the cushion as compared to the tendency observed for the rigid seat. The manikin showed a more pronounced decrease in both the chest resonance frequency (up to 4 Hz) and peak magnitude response at the higher acceleration level with the use of the cushion as compared to the rigid seat. For the humans, the tendency for a decrease in the magnitude of the peak head response at the higher acceleration level was less pronounced using the cushion as compared to the results for the rigid seat. In contrast, the peak head magnitude response appeared to decrease in the manikin with the use of the cushion at the higher acceleration level as compared to the minimal effects observed for the rigid seat. The effects of acceleration on the magnitude of the peak head transmissibility were less pronounced in the humans as compared to the manikin with the use of the cushion as compared to the rigid seat. The peak resonance frequency was also decreased for the rigid body mass at the higher acceleration level. However, the magnitude of the peak responses were similar at both acceleration levels given the large variability in the data collected at the higher acceleration level. In general, the manikin resonance frequencies remained significantly higher than the humans with the use of the cushion. However, the peak chest transmissibilities were more similar between the humans and manikin, while the peak head transmissibilities showed higher responses in the manikin at both acceleration levels with the use of the cushion. The contrasting effects can be observed by comparing Figures 4 and 8 for the lower acceleration level.

For the transmissibility measured at the interface between the occupant and cushion, all occupants showed resonance frequencies which coincided closely with the resonance frequencies observed in the other measured responses. The peak magnitude responses at the interface for the humans and manikin were less than 2.0. The second human did show a slight increase at the higher acceleration level. As with the majority of responses, the manikin and rigid body mass showed a higher transmissibility peak at the interface as compared to the humans. For the manikin and the rigid body mass, the acceleration level did not appear to have a significant influence on the response. The rigid body mass did show a lower transmissibility at the interface than at the top of the mass. The placement of the metal shot bags on the cushion and the location of the accelerometer on top of the mass may have caused some off-axis motion in the mass.

DISCUSSION

HUMAN RESONANCE BEHAVIOR USING MANIKINS AND RIGID BODY MASSES - The results of this study showed that the whole-body vibration response characteristics of the human were different as compared to the response characteristics of the Hybrid III aerospace manikin and rigid body mass representation. The differences were specifically reflected in the dramatically higher resonance frequencies observed in the driving-point impedance and transmissibilities for the manikin as well as the rigid body mass regardless of the seating configuration (rigid seat vs cushion). Both the manikin and rigid body mass produced significantly higher impedance magnitude peaks as compared to the humans regardless of the seating configuration. Although the two human subjects were specifically selected to avoid response differences occurring due to body weight, the results strongly suggest that the distribution of the mass, stiffness, and damping characteristics are different between these humans and the manikin. In a recent study [7], in which the vibration response characteristics of a wide range of female and male subjects were compared, including the two males used in this study, the same conclusion was drawn about the distribution of structural properties. However, the differences between

the humans were primarily observed in the magnitudes of the peak responses and not on the location of the resonance frequencies. A comparison of these data shows that, in general, the manikin tested in this study produces significantly higher primary resonance frequencies as compared to humans, regardless of any structural differences occurring among the humans. Of particular concern for predicting cushion effects were the contrasting trends observed in the peak impedance and peak chest and head transmissibilities between the humans and manikin with the addition of the seat cushion.

There does appear to be some similarity between the primary resonance behavior observed in the manikin and the secondary resonance observed in the humans at higher frequencies. In the human, the primary impedance resonance peak occurring between 4 and 8 Hz has been considered to be strongly influenced by the dynamic behavior of the upper torso, including the shoulders and the soft tissues contained within the chest cavity [9]. The coincidence of the peak chest transmissibility and peak impedance in the frequency range of 4 to 8 Hz further supports this influence [7, 10]. In addition, these observations indicate that the resonance behavior observed for the head at the low frequencies was strongly influenced by the motions of the upper torso. The contrasting effects of the acceleration level and cushion between the head and chest peak responses requires further investigation of coupling effects in the human body. The majority of subjects tested in this laboratory have also shown a prominent transmissibility peak at C_7 at higher frequencies between 12 and 18 Hz [7]. These findings have strongly indicated that the spinal column structures may be a major contributor to the additional impedance resonance peak which has been observed at higher frequencies between 12 and 19 Hz [7, 10, 11]. The additional peak located at about 13 Hz for the first human coincided with the peak transmissibility measured at the seventh cervical vertebra (C_7) in this subject. A peak was also observed at C_7 for the second human used in this study but was lower as compared to the first human [7]. The second human did not show an identifiable impedance peak at the higher frequencies. While measurements were not made in the spinal column of the manikin, it appeared that the resonance frequency observed in the impedance and chest transmissibility frequency responses for the manikin coincided more closely with the spinal column resonance observed at the higher frequencies in the human. The magnitude of the peak chest transmissibility in the manikin did fall within the wide range of the peak spine (C_7) transmissibility magnitudes observed for humans and coincided closely with the magnitude peak observed at C_7 for the first human. The similarity between the resonance behavior of the human spine and the primary impedance and chest resonances observed in the manikin becomes even more apparent when considering the effects of the seat cushion on vibration response. The majority of current seat cushion designs have been shown to increase the transmission of vertical vibration to the human in the vicinity of the primary resonance (4 to 8 Hz) as demonstrated in this study and in previous studies [7, 10, 12, 13, 14]. In contrast, these cushions tended to attenuate the vibration at higher frequencies as shown in the tendency for a decrease in the peak spine (C_7) transmissibility occurring at higher frequencies for humans [7, 10]. Interestingly, the results of this study showed that the addition of the cushion decreased both the primary impedance and chest transmissibility peaks in the manikin, similar to the tendencies observed in the human spine transmissibility at higher frequencies.

Current biodynamic manikins are primarily designed for simulating the effects of short duration, high accelerative loading of the human body. These designs do not consider the low frequency, low amplitude human body sensitivity observed during vibration exposure. The manikin used in this study did appear to show the attenuating effects of cushions at higher frequencies which coincide with human spinal motion, however, the ability to simulate the resonance response of humans in the region of greatest sensitivity (4 to 8 Hz) is critical for developing improved seating systems.

PREDICTING HUMAN RESONANCE BEHAVIOR USING MATHEMATICAL MODELS - Lumped-parameter mathematical models have been used to predict the primary resonance behavior of the human body occurring between 4 and 8 Hz. These models use mass, spring, and damper elements to represent the dynamic response of 'lumped' anatomical structures or regions. A five degree-of-freedom (DOF) model was developed to simulate the multiple regions of resonance identified in the driving-point impedance response [11]. The model elements were selected to represent the dynamic responses of the upper torso (chest), spine, legs, and lower torso. The current model does not include the model elements for the head. The five DOF model was found to be effective in simulating human impedance response. Based on transmissibility data collected at specific anatomical sites, the model was also found to be quite effective in predicting the general resonance behaviors of the anatomical structures represented in the model [10]. The model was expanded to include seat cushion stiffness and damping properties estimated from rigid body mass tests. On comparing the predictions for two military cushions tested in this laboratory, the model was found to be effective in simulating the relative increases in the impedance and chest transmissibility peaks occurring at low frequencies (4 to 8 Hz), and the relative decreases in the impedance and spine transmissibility peaks occurring at higher frequencies (12 - 19 Hz) with the addition of the cushions [10]. A simplified three degree-of-freedom model, which lumped the torso and spine structures, was also shown to predict the relative effects of the seat cushions, but only in the low frequency region [10]. It appears that human models which are capable of predicting the low frequency response effects would be more appropriate as compared to manikins for evaluating the effects of seat cushions on humans, particularly if human testing is not feasible. The cushion stiffness and damping properties required by the model could be estimated from rigid body testing of the cushion or from known cushion material properties. Caution should be taken in using these models since differences in

vibration response have been observed between females and males of varying sizes [7]. It is important to consider the ability of a model to predict these differences. Specifically, the five DOF model developed in this laboratory requires further modifications and validation of its predictive capabilities before being used as a tool for optimizing seat cushion designs for a variety of occupants.

ACKNOWLEDGMENT

The author acknowledges the assistance of Mr. Raymond Newman, DynCorp, Inc., in the operation of the vibration testing facility.

REFERENCES

[1] Griffin, M. J., "The evaluation of vehicle vibration and seats," Applied Ergonomics, Vol. 9, No. 1, pp. 15-21, 1978.

[2] International Organization for Standardization, "Evaluation of human exposure to whole-body vibration - Part 1: General requirements," ISO 2631/1-1985, International Organization for Standardization, Geneva, 1985.

[3] Moran, A. W. and Butler, B. P., "Whole-body vibration assessment of the M916A1 truck trailer," USAARL Report No 93-34, US Army Aeromedical Research Laboratory, Fort Rucker AL, 1993.

[4] VanIngen-Dunn, C. and Richards, M. K., "Feasibility of reducing incidence of low back pain in helicopter pilots using improved crewseat cushions," SR-1991-0009, Armstrong Laboratory, Wright-Patterson AFB OH, 1991.

[5] Gurram, R. and Vértiz, A., "Nonlinear vibration response of automotive seat systems,"IBEC 1995 Proceedings, Interior and Safety Systems, Vol. 16, November, 1995.

[6] Smith, S. D., "The effects of military aircraft seat cushions on human vibration response", United Kingdom Informal Group Meeting on Human Response to Vibration, Institute of Naval Medicine, Alverstoke, Gosport, Hants, PO12 2DL, 19-21 September, 1994.

[7] Smith, S. D., "Comparison of the effects of whole-body vibration exposure between females and males," ADA309219, Final Report, US Army Medical Research and Materiel Command, Fort Detrick, MD, May, 1996.

[8] Hixson, E. L., "Mechanical impedance and mobility," In: Shock and Vibration Handbook, Harris, C. M. And Crede, C. E., Editors, New York, McGraw-Hill Book Company, 1961.

[9] Guignard, J. C. and Irving, A., "Effects of low frequency vibration on man," Engineering (Lond), Vol 190, pp. 364-367, 1960.

[10] Smith, S. D., "Cushions and suspensions: Predicting their effects on the biodynamic responses of humans exposed to vertical vibration," Accepted for publication in the International Journal of Vehicle Design, Driver-Vehicle Interactions, 1997.

[11] Smith, S. D., "Nonlinear resonance behavior in the human exposed to whole-body vibration," Shock and Vibration, Vol. 1, No. 5, pp 439-450, 1994.

[12] Fairley, T. E. and Griffin, M. J., "A test method for the prediction of seat transmissibility," SAE Paper 860046, Society of Automotive Engineers, International Congress and Exposition, Detroit, MI, 24-28 February, 1986.

[13] Pope, M. H., Broman, H., and Hansson, T., "The dynamic response of a subject seated on various cushions," Ergonomics, Vol. 32, No. 10, pp 1155-1166, 1989.

[14] Mansfield, N. J., "Prediction of the transmissibility of a vehicle seat in the vertical direction," UK Informal Group Meeting on Human Response to Vibration, APRE, Ministry of Defence, Farnborough, GU14, 6TD, England, 20-22 September, 1993.

970601

High Accuracy Semiconductive Magnetoresistive Rotational Position Sensor

Yasushi Ishiai, Noriyuki Jitousho, and Tetsuhiro Korechika
Matsushita Electronic Components Co., Ltd.

Joe LeGare and Sumitake Yoshida
Panasonic Industrial Co.

Copyright 1997 Society of Automotive Engineers, Inc.

Recently there is demand for rotation sensors capable of high-accuracy detection and very low-speed detection of rotation at high temperature for automobile use. To meet this requirement, a rotation sensor using an InSb thin-film magneto-resistors with good thermal stability has been developed.

This sensor transduces magnetic flux change due to gear rotation to resistance change. It is composed of InSb thin-film magneto-resistors fabricated by a newly developed process and signal shaping circuits where resistor signals are converted to digital signals using no amplifier. Accordingly, the signals are independent of the measured frequencies, making possible very low speed (0 to 20 Hz) detection. The sensor stably operates in the temperature range from -40 to 150 degree C for thousands hours. There is no need for a shielded harness due to the digital output signal. It can be used for high temperature environments like in the engine compartments for rotation detection of ignition systems, electronically controlled AT systems, antiskid braking systems, etc.

INTRODUCTION

Trends, such as regulations for automobiles, like CAFE, OBD II, and LEV, along with energy and safety consciousness, and the need high-accuracy in navigation systems, etc., have increased demand for improved rotation sensors. These sensors need the following characteristics: (1) capability of very-low-speed detection at 0 - 20 Hz and (2) capability of high-accuracy detection at high temperature. In order to meet these requirements, we have been studying the "semiconductor magneto-resistor" that enables high-sensitivity sensor output and thus greatly-reduced peripheral circuits.

This semiconductor magneto-resistor (=semiconductor MR) was already reported in 1950's[1], especially in the period when the deposition process of InSb thin-film, as III-V family semiconductor thin-film, first became the subject of research[2]. The semiconductor MR, whose magnetic sensitivity is in proportion to electron mobility of semiconductor, limits its material to InSb and InAs. Therefore, it is an important point regarding how the crystalline semiconductor thin-film material is formed. Traditionally, slices of a bulk single-crystal wafer or those formed by the growth of the high-quality InSb thin-film on Mica substrate are transferred onto another substrate [3]. In either construction, because of the use of an adhesive layer, the problems, mainly caused by the difference in expansion coefficients, made it inadequate for the high-temperature specifications as in the electronic package applications. Therefore, we have developed the InSb thin-film MR based on the following basic technologies:

(1) The technology for the growth of InSb thin-films with excellent crystallinity directly on the Si substrate.

(2) The electrode material & process technology for ensuring reliable connections over a wide range of temperature.

(3) The heatproof packaging technology.

These technologies as well as the characteristics of a complete rotation sensor for automotive use are described below.

INSB THIN-FILM MAGNETO-RESISTORS

The basic structure of InSb thin-film magneto-resistors is shown in Fig. 1. It is basically constructed, after the growth of InSb thin-film on the substrate, by depositing a number of ladder-structure electrodes called, "shorting bar electrodes". In this construction, whose principle of operation is shown in Fig.2 (a), a phenomenon is used that the current path is elongated by the Lorentz

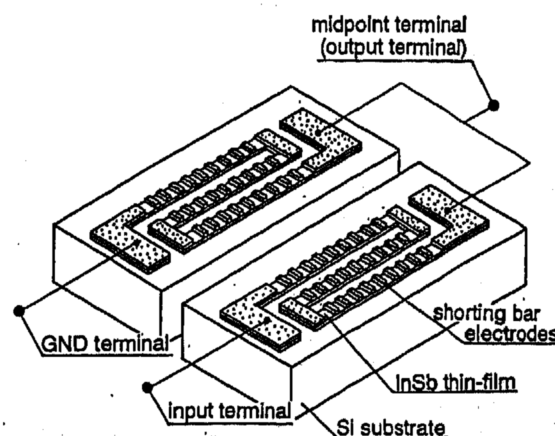


Fig.1 Basic structure of InSb thin-film magneto-resistors

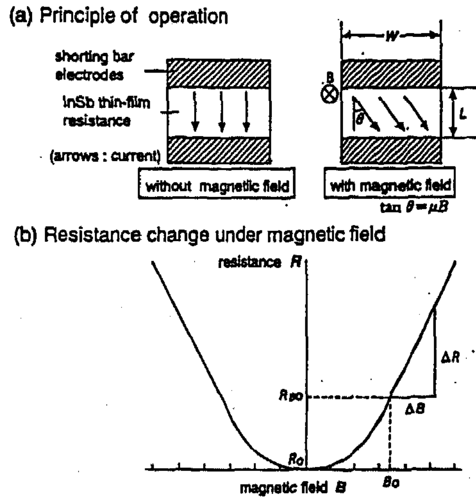


Fig.2 Principle and basic characteristics of InSb thin-film magneto-resistors
(a) Principle of operation
(b) Resistance change under magnetic field

force applied to the carrier when the magnetic field is applied in the transverse direction relative to the film surface of the InSb thin-film between the shorting bar electrodes. Here, as shown in Fig. 2 (b), the resistance change against the magnetic field differs between the low magnetic field region and the high magnetic field region. With adequate accuracy it can be expressed by the following equations:

(low-magnetic-field characteristics)

$$R_B / R_0 = \frac{\rho_B}{\rho_0}(1+g(\mu B)^2) \qquad (1)$$

(high-magnetic-field characteristics)

$$R_B / R_0 = \frac{\rho_B}{\rho_0}(G+\frac{W}{L}\mu B) \qquad (2)$$

where, R_B: resistance with magnetic field, R_0: resistance without magnetic field, μ: electron mobility of the semiconductor, B: applied magnetic field, g and G: functions of element's shape (L/W) respectively, L: element length, W: element width, ρ_B / ρ_0: magneto resistive effect (nearly equal to 1).

Generally, the change over point from square-law characteristics to linear characteristics is given by the following empirical equation:

$$\mu B = 0.65 \quad \text{(SI units)} \qquad (3)$$

As is obvious in Fig. 2 (b), the resistance change of this element is greater in the high-magnetic-field region, and accordingly it is necessary, in practical use, to set the point of operation in the linear (high field) region by placing a magnet on the backside of the element.

Furthermore, as is obvious in Equation 2, in this semiconductor magneto-resistor, the magnetic sensitivity or the magnitude of resistance change is in proportion to the element's shape (*W/L* ratio) and the electron mobility of semiconductor used. Since the magnitude of electron mobility, in particular, is a key factor of the element's characteristics, InSb, which has the highest value of μ among semiconductors, is used. In addition, as regards the merits of thin-films, the following points are indicated: (1) capability of high impedance, (2) improvement in heat resistance ,and (3) great reduction in the material (wafer) cost.

We have developed the technology for the growth of InSb thin-film with excellent crystallinity and thus realized the InSb thin-film magneto-resistors with high sensitivity and high heat resistance. The fabrication process is described below.

FABRICATION PROCESS OF INSB THIN-FILM MAGNETO-RESISTORS - The fabrication process of InSb thin-film magneto-resistors is shown in Fig. 3. As for the fabrication technology for the InSb thin-film with high electron mobility or excellent crystallinity, studies have been reported by several groups. Studies about heteroepitaxy on the InSb on GaAs substrate are the most popular, and as for growth technology, MBE [4,5], MOCVD [6,7], and sputtering [8] have been reported. In addition, examples of growth on CdTe [9] and sapphire [10] have also been reported. As a fabrication technology of InSb thin-film in practice, the method that, after the deposition on the Mica substrate, transfers it onto another substrate via an adhesive layer is being used in manufacturing hall elements [11]. Among them, taking into account the substrate cost, heat resistance, and the priority of proper orientation of the surface of the InSb thin-film, we have decided to use the Si(111) substrate.

Prior to the growth of the InSb thin-film, the surface washing treatment of Si substrate is an important point. Traditionally, in heteroepitaxy of InSb thin-film on Si substrate, a method that obtains an atomically clean surface by heating the substrate at high temperature above 900°C in a vacuum chamber and sublimating SiO existing on the surface is used [12]. In this method, the high-temperature and the ultra-high vacuum that are required, results in poor mass-productivity. Since the optimum temperature in the growth of the InSb thin-film is relatively low (430°C), we adopted the hydrogen termination treatment where the Si surface structure can be maintained sufficiently stable up to this temperature.

In the fabrication process, we use the 4-inch Si(111) substrate (resistivity ≧ 1 kΩ·cm). After treatment of the substrate surface by RCA washing, it is immersed in dilute HF solution. The hydrogen termination treatment is applied removing the oxidized layer on the substrate surface and terminating the unoccupied valence on the Si surface with hydrogen atom. After that, the substrate is rinsed with ultra pure water and introduced into the vacuum chamber.

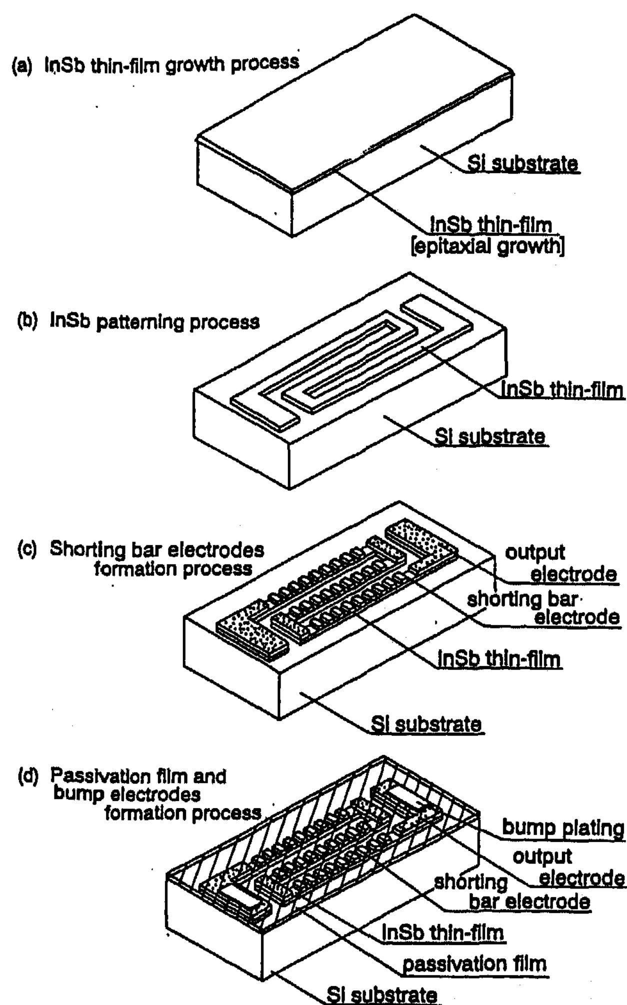


Fig.3 Fabrication process of InSb thin-film magneto-resistor

Then, the growth process of the InSb thin-film starts. In this process, since the lattice mismatch between InSb and Si is as great as about 19%, a simple one-step growth technology will result in a polycrystalline thin-film, the two-step growth technology is adopted where in (1) after the preliminary growth layer is deposited at low temperature, (2) the main layer is grown at normal growth temperature. In addition, in this growth technology, the vacuum evaporation method (three temperature method) is used where In and Sb are separately evaporated by

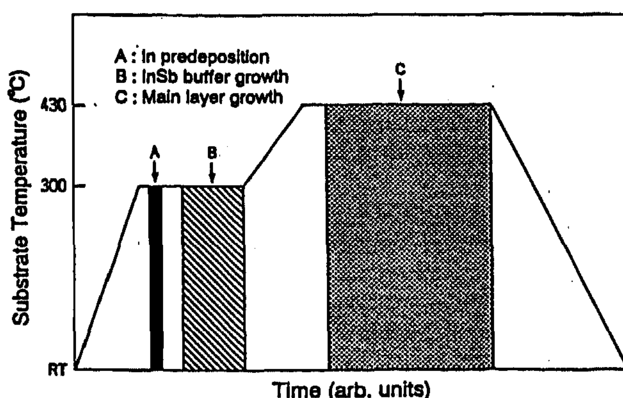


Fig.4 Typical substrate temperature profile for growth of InSb film on Si(111)

resistance heating. At the first stage of study, the typical substrate temperature profile for growth was a simple two-temperature type with low temperature and high temperature, as shown in Fig. 4.

<u>(1) Deposition of the preliminary growth layer</u> - The hydrogen-terminated Si(111) substrate described above is introduced into the vacuum chamber and, after lowering the vacuum pressure below 5×10^{-5} Pa, the substrate temperature is raised to 300℃. Then, as shown in Fig. 4 A, about 0.5 nm thick of In film is predeposited under the pressure of less than 1×10^{-4} Pa at the deposition rate of about 0.1 nm/s. Here, if the substrate temperature is too low, judging from the observation of the surface morphology of the InSb thin-film after the following main growth, a discontinuous mass crystal will be formed. Therefore, it is necessary to maintain the appropriate temperature. If the predeposited In film is too thin, it will not play a sufficient role as the core formation material, and if it is too thick, flocculation will be caused. Therefore, it is necessary to control the appropriate thickness.

After pre deposition of In, maintaining the substrate temperature at 300℃, In and Sb are co-evaporated as shown in Fig. 4 B. Here, with the atomic ratio Sb/In = 3 and the In deposition rate of approximately 0.1 nm/s, the approximately 25 nm-thick InSb thin-film is deposited. The change of reflection high-energy electron diffraction (RHEED) pattern of a series of low-temperature preliminary growth layers from the early stage of growth is as follows:

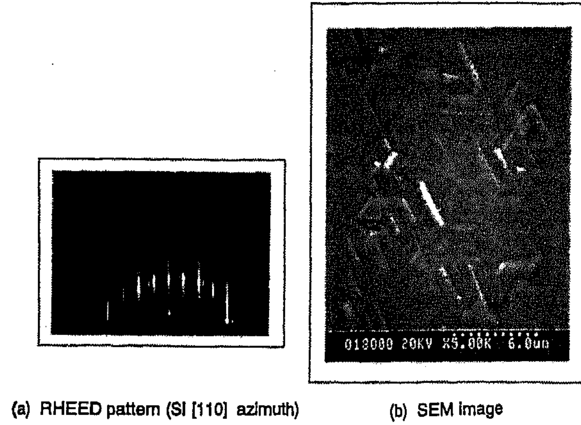
As In is fed from the early stage of growth, the pattern intensity of the substrate Si(111) gradually decreases and, on the coevaporation of In and Sb, the intensity of Si(111) rapidly decreases and finally disappears. On the other hand, InSb(111) gradually emerges and finally the complete orientation pattern of InSb(111) is observed.

Here, the epitaxial growth of InSb(111)//Si(111), InSb<110>//Si<110> is confirmed. After that, as shown in Fig.4 B → C, the substrate temperature is elevated to 430℃ for the main growth.

<u>(2) Deposition of the main growth layer</u> - After maintaining the substrate temperature at 430℃, as shown in Fig. 4 C, an approximately 3 μm-thick InSb thin-film is deposited with the atomic ratio Sb/In = approximately 2.5 and the In deposition rate of approximately 0.3 nm/s (the InSb growth rate 0.75 nm/s).

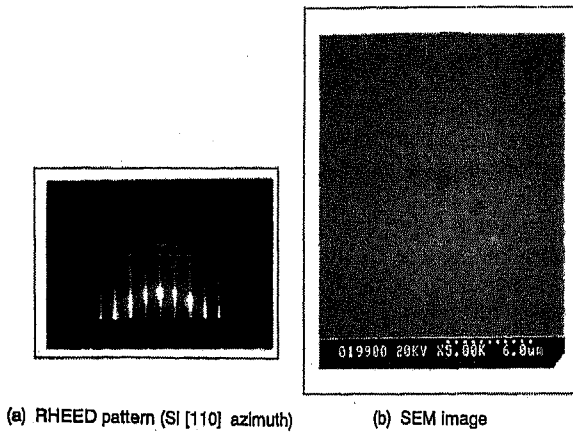
The RHEED pattern and SEM image after completion of the main growth are shown in Fig. 5 (a)(b). Fig 5 clearly shows that, in the film after the two-step growth with the two-temperature profile with low temperature and high temperature as shown in Fig. 4, although strong and clear streaks are observed, the boundary exists between the domains of epitaxial growth. That is, although the epitaxial growth is achieved satisfying InSb(111)//Si(111), InSb<110>//Si(110), domains with two direction exist: one is the exactly same direction as Si and the other is the direction rotated by 180deg around the Si(111) plane. In measurement of electron mobility (μ) of the film in this state using the van

der Pauw technique, μ was found to be $\mu = 3.63\mathrm{m}^2/\mathrm{V \cdot s}$, which is not fully satisfactory.



(a) RHEED pattern (Si [110] azimuth) (b) SEM image

Fig.5 Structure of InSb film after typical growth process



(a) RHEED pattern (Si [110] azimuth) (b) SEM image

Fig.6 Structure of InSb film after optimized growth process

In this regard, we particularly focused on the temperature profile during the predeposition to clarify the relation between this profile and the crystallinity of the growing film and thereby optimized the conditions of the main growth. Thus, as shown in Fig.6 (a) (RHEED pattern) and (b) (SEM image), we realized the epitaxial growth of InSb thin-film on the Si substrate where clear streaks of InSb(111) with Kikuchi lines are observed and the domain boundary shown in Fig. 5 disappears to form the nearly single domain with excellent surface smoothness.

For the completed InSb thin-film, μ was measured and the value nearly equal to the theoretical value[13] was obtained (maximum value $\mu = 5.99\mathrm{m}^2/\mathrm{V \cdot s}$). In addition, this growth process of the InSb thin-film is well reproducible and enables the consistent fabrication of thin-films with excellent crystallinity and high electron mobility.

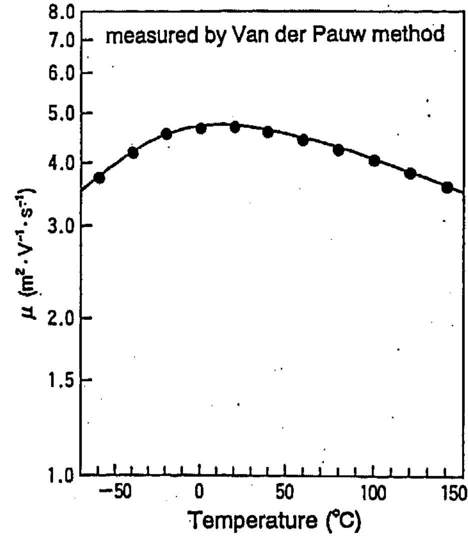


Fig.7 Temperature dependence of mobility of typical InSb thin-film

Fig. 7 shows an example of the temperature dependence of electron mobility of typical InSb thin-films (thickness: 3 μm) fabricated by this process. Electron mobility has its maximum value near room temperature and decreases in both high temperature and low temperature regions. The temperature dependence of electron mobility, with carrier concentration at room temperature of approximately $1\times10^{22}\mathrm{m}^{-3}$, is shown by the following theoretical equations[14,15], where electron mobility under scattering by ionized impurity is μ_I, electron mobility under scattering by lattice is μ_L, electron mobility under scattering by dislocation is μ_D, resultant electron mobility under these scattering mechanisms is μ_H, electron mobility considering surface scattering on the thin-film is μ_{TH}:

$$\mu_I = 1.10T^{0.73} \qquad (4)$$

$$\mu_L = 1.09\times105T^{-1.68} \qquad (5)$$

$$\mu_D = 8.165\times10^{11}\frac{T}{N_D} \qquad (6)$$

$$\mu_H = \left(\frac{1}{\mu_I}+\frac{1}{\mu_L}+\frac{1}{\mu_D}\right)^{-1} \qquad (7)$$

$$\mu_{TH} = \frac{\mu_H}{1+\dfrac{2.06\times10^{-3}\cdot\mu_H T^{0.5}}{d}} \qquad (8)$$

N_D: dislocation density, d: thickness of the InSb thin-film.

Based on Equations 4-8, the theoretical values of the temperature dependence of electron mobility are obtained and, by comparison between the theoretical values and our measurement values, the dislocation

density is considered to be approximately $10^{13}m^{-2}$. This is because dislocations generated by interfacial mismatch with the Si substrate still remain in the growing thin-film. In addition, it is found that scattering by dislocations in the low-temperature region and scattering by the lattice in the high-temperature region are dominant factors of electron mobility respectively. The differences in electron mobility due to thin-film quality is great mainly in the low-temperature region, and the better quality of thin-film results in smaller decreases in the low-temperature region. In the high-temperature region, since scattering by the lattice is dominant and nearly uniform, the differences are not very great.

After completion of growth of the InSb thin-film, through the photo lithograph process shown in Fig. 3 (b), the InSb thin-film is processed in the longitudinal direction of the element by the wet process, then the formation process of shorting bar electrodes with the ladder-structure shown in Fig. 3 (c) starts. After the uniform coating of electrode material, the desired ladder shape of electrodes is obtained by the photo lithograph process. In this process, since a number of ladder-structure electrodes are formed on the InSb thin-film, the important points are as follows:

(1) to minimize contact resistance between the electrodes and the InSb thin-film,

(2) to minimize stress by the electrode material on the InSb thin-film, and

(3) to prevent erosion of InSb thin-film in processing the electrodes (wet process).

We formed the ohmic contact with low contact resistance on the InSb thin-film and, from the viewpoint of chemical selectivity (on InSb) during the wet process, adopted the Cu/Cr laminated electrode. As for the above-mentioned point (1), in addition to HCl solution treatment used in the removal treatment of the InSb oxidized layer prior to the electrode formation, $(NH_4)_2S_x$ (ammonium polysulfide) treatment was introduced to keep the surface of the InSb thin-film in an unoxidized condition and to realize low contact resistance and, as for (2), for the evaporation formation, total stress was reduced mainly by the electrode thickness control and, as for (3), the process was stabilized by an appropriate selection of treatment solution.

After the formation of shorting bar electrodes, as shown in Fig. 3 (d), a passivation film electrodes was formed over the electrodes. For the passivation film, low-expansion polyimide ($\alpha: 5\times10^{-6}$ /deg) was used.By using the low-expansion material, as previously described, we minimized the stress on the InSb thin-film and suppressed the change of the forbidden band width shown in the following equation. In particular, expansion coefficient of the upper layer material is great to increase the forbidden band width when tensile stress is applied and to increase effective mass of electron and to deteriorate the magnetic characteristics of the element.

$$\frac{dE_g}{dP} = 1.58\times10^{-10}\,eV\cdot m^2\,/\,N$$

(Left side: change of forbidden band width per unit stress, stress dependence: tensile (+), compressive (-))

Stress applied to the InSb thin-film in burning this polyimide is at most approximately (tensile stress) and, in the wide range of temperature, has no adverse effect on the element's characteristics. After spin coating of this polyimide material, plating windows are made through the conventional photo lithograph process and it is burnt in vacuum. By the polyimide passivation film thus formed, under specifications for electronic packaging reliability, sufficient reliability is assured. Then, by electrolytic plating, Au/Ni is plated to complete the element.

Fig. 8 shows the characteristics of resistance change with magnetic field of the completed element (μ:4.5m^2 /V·s, room temperature value). Under 0.4 T, the resistance is three times as high as the resistance without a magnetic field.

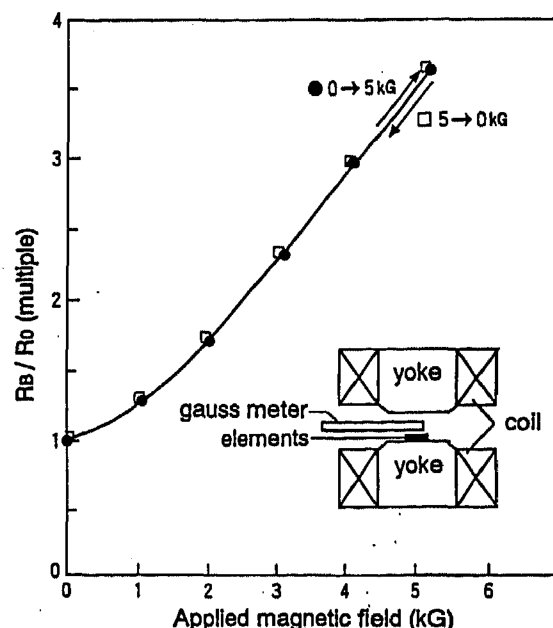


Fig.8 Characteristics of resistance change under magnetic field of InSb thin-film magneto-resistors

Fig. 9 shows the relation between the output voltage and the air gap (between the detection gear and the element surface) at room temperature in the differential element. A differential wheatstone bridge configuration is used to compensate for the large temperature dependence of the InSb element's impedance characteristic. It is understood that output voltage of SMR is sufficiently high under the usual condition (air gap is about 1.5mm).

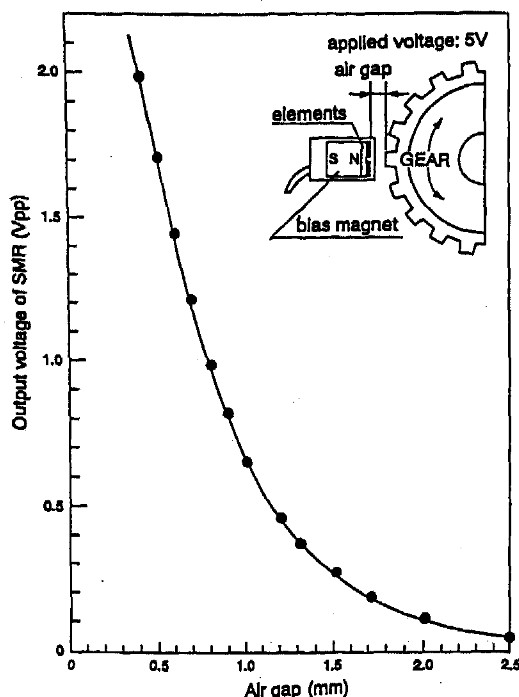


Fig.9 Characteristics of output voltage vs. air gap (at 5V applied)

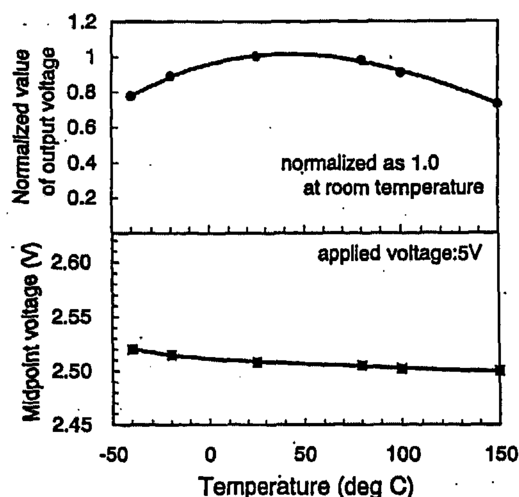


Fig.10 Characteristics of dependence of output voltage (normalized at 1 at 25 degree C value) on temperature

In addition, Fig. 10 shows the typical characteristics of temperature dependence of output voltage (normalized at room temperature value) and midpoint voltage in the differential element. Output voltage has a similar temperature dependence to that of electron mobility and has its maximum value near room temperature. It also tends to decrease in both the low-temperature and high-temperature regions. Midpoint voltage tends to slightly diverge in the low-temperature region and converge in the high-temperature region. This is caused by heterogeneity of the element including differences in the thin-film quality between the two elements. This needs to be further reduced in the future.

HEAT-PROOF PACKAGING TECHNOLOGY - When the InSb thin-film magneto-resistors are used as a rotation sensor for automobiles, the packaging technology is important factor.

The construction of inner lead bonding parts is shown in Fig. 11. In packaging, important points are as follows:

(1) to assure the stable connection in the temperature range of -40 ~150℃ as required for automotive use, and

(2) to reduce the package thickness (including mold) of the upper surface of the element (in order to allow the distance between the element and the detection gear to be as small as possible).

Usually, in sensor elements for automobile use, wire-bonding→transfer products such as hole IC are popularly used. However, we require the upper surface package thickness to be less than 200 μm as described in (2) and transfer products do not necessarily satisfy this requirement. In this regard, we adopted the TAB packaging that enables a thinner package.

As shown in Fig.11, since the InSb thin-film magneto-resistors form the detection part with two-chip differential type, two device holes are made to keep the two chips at a certain distance (to fit the detection gear pitch) and, as for the ILB (Inner Lead Bonding) connection, a Au-Sn eutectic is formed between Sn plating on the Cu foil lead of TAB and Au electrodes on the element connection parts. After that, a resin coating is applied in order to protect the element surface and the ILB parts. This completes the detection element.

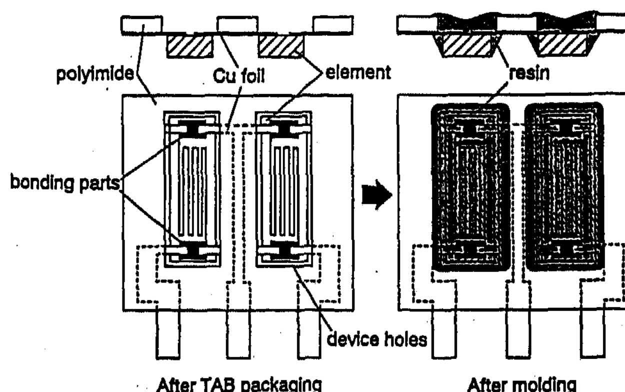


Fig.11 Construction of inner lead bonding parts to InSb thin-film MR

Since TAB packaging has not been applied to automotive area so far, durability verification was needed. In order to assure the stability of the connections, as previously described in (1), the design assurance of the structure particularly in terms of its consistency was an important point. Regarding the design assurance, we performed theoretical analyses in terms of thermal factor (thermal stress), vibration factor (resonance resistance), and connection factor (lead disconnection). Fig. 12 shows the evaluation models for analyzing thermal stress. In this figure, (a) is the face structure model of TAB and (b) is the cross-sectional structure model. Using this

model, the stress distribution was obtained. According to the analytical results, the maximum stress occurs on the Cu foil in the OLB (Outer Lead Bonding) parts indicated with the dotted line in this figure, but breakage will not occur because it is below the breaking strength at the use temperature limit of 150°C. At this point, in terms of equivalent safety ratio at 150°C, as shown in Table 1, a safety ratio greater than 1.5 will be assured by the forming height more than 0.1 mm. As for the vibration factor, a resonance frequency more than 20 kHz shall be assured in design and, as for the connection factor, the connection conditions shall be defined to avoid breakage of the lead. Thus, the suitability of the design was theoretically evaluated and it was found that it is sufficiently reliable in specifications for automobile use. This was confirmed by reliability tests.

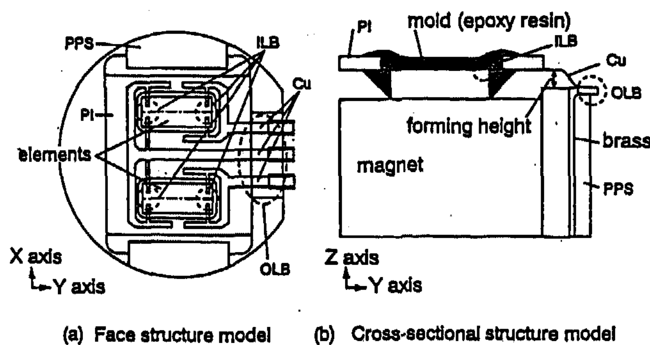


Fig.12 Model for analyzing thermal stress to TAB structure

Table 1 Safety ratio thermal stress analysis for TAB structure

forming height (mm)	safety ratio by thermal stress analysis	
(a) TAB face structure model		
	X-axis direction	Y-axis direction
0	2.50	2.50
0.1	2.50	2.50
0.5	2.50	2.50
(b) TAB cross-section structure model		
	Y-axis direction	Z-axis direction
0	1.09	1.85
0.1	1.60	2.50
0.5	2.01	2.50

ROTATION SENSOR WITH INSB THIN-FILM MAGNETO-RESISTORS

CONSTRUCTION OF ROTATION SENSOR - The construction of rotation sensor with the above-mentioned InSb thin-film MR as the detection part is shown in Fig. 13.

The InSb thin-film magneto-resistors are placed on the Sm-Co bias magnet. The peripheral circuit parts are constructed on the flexible printed circuit (FPC), and the OLB connection is made by welding at the junction terminal connecting the element and the peripheral circuit. After that, by ultrasonic bonding, the resin cap is integrated with the functional parts to complete the sensor

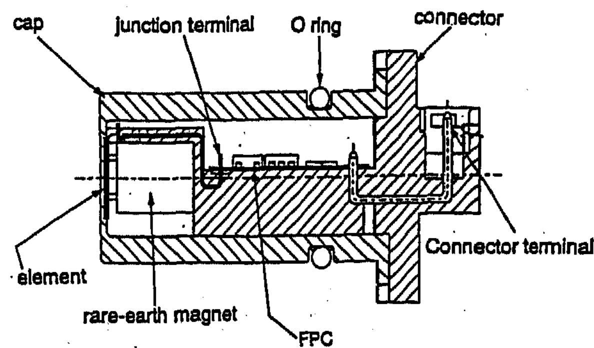


Fig.13 Construction of rotation sensor with InSb thin-film MR

unit. In this rotation sensor, construction of the peripheral driving circuit is shown in Fig. 14. Since output sensitivity of the InSb thin-film MR is high, the waveform shaping circuit (for converting into a rectangular wave) is constructed without an amplifier, using a simple circuit consisting only of a single comparator. For this peripheral circuit, each component is chosen to sufficiently meet the required heat resistance, pressure resistance, etc.

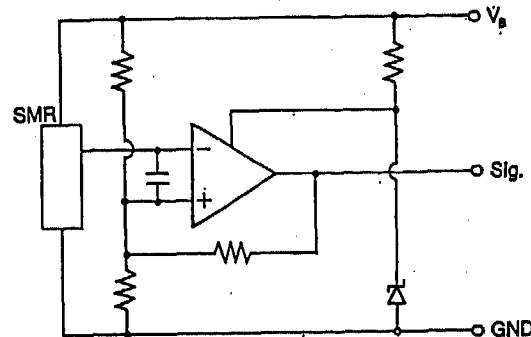


Fig.14 Construction of peripheral driving circuit

Since an amplifier is not used, the following important points must be considered.

(1) to optimize the magnetic circuit part,

(2) to reduce the gap between the element surface and the unit surface within the sensor unit,

and in order to assure reliability such as heat resistance,

(3) to assure structural reliability.

As for (1), optimization of the pitch between the two elements and the shape of magnet and clarification of optimum operating magnetic field of the element (resistance change per unit magnetic flux is greatest in the range of 0.3~0.4 T) were performed to increase the difference in resistance between the two parts of the differential element as much as possible. As for (2) and (3), as previously described about TAB, it was designed to assure theoretically sufficient reliability.

ROTATION SENSOR CHRACTERISTICS - Fig. 15 shows the output signal of the InSb thin-film MR and the sensor output signal after signal shaping at

temperatures of -40, 25, and 150°C. In addition, Fig. 16 shows the resistance change of the element after the heat cycle test between -40°C and 150°C under 5V drive. Then Fig. 17 shows the resistance change of the element after the high-temperature voltage bias test at 150°C, 5V. The InSb thin-film magneto-resistors have sufficient temperature stability (and achieve stable operation) for use as a rotation sensor.

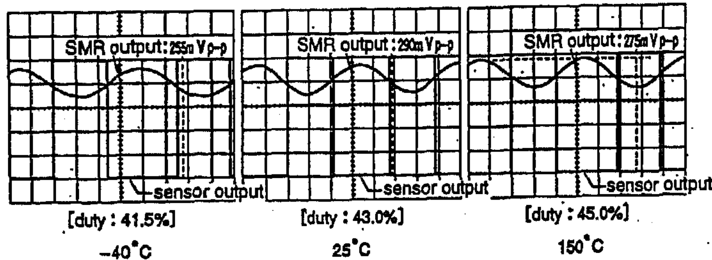


Fig.15 Output signal of InSb MR and sensor output signal after signal shaping

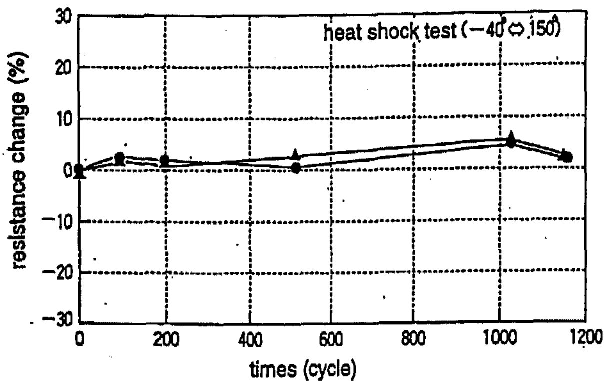


Fig.16 Resistance change of InSb MR after heat shock test

Table 2 shows the reliability test results as a rotation sensor for automobile use. Table 3 shows the basic specifications of this InSb thin-film MR type rotation sensor. Considering the absence of fatigue deterioration, it clearly has sufficient performance for use as a rotation sensor for automotive applications.

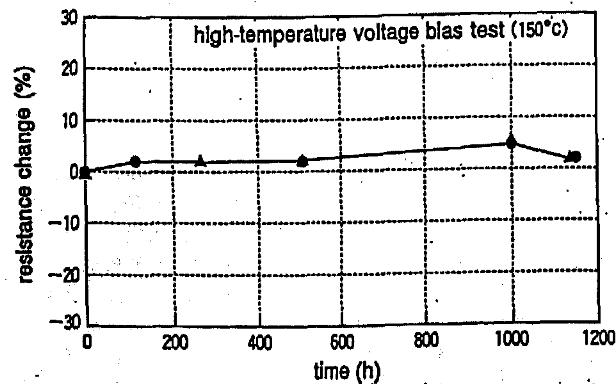


Fig.17 Resistance change of InSb MR after high-temperature voltage bias test

Table 2 Reliability test results of rotation sensor

test items	conditions	results
high temperature	150°C, 500h	1000h OK
low temperature	-40°C, 500h	500h OK
high temperature and high humidity	85°C, 90~95%, 500h	500h OK
heat shock	-40°C, 150°C, 1h each, 1000cycle	2400cycle OK
high temperature voltage bias	150°C, 5V, 1000h	1000h OK
electrostatic charge	±1kv, 10 times for each terminal	OK
surge durability	70V, 200ms	OK
EMI test (strong electric field)	2~200 MHz, 100 V/m	OK

Table 3 Specification of InSb thin-film MR type rotation sensor

[Basic Specifications]

operating temperature	-40~150°C
operating voltage	5±0.25V
consumption current	10 mA or less
output voltage	High:4.8 V, Low:0.8 V
output current	10 mA or less
duty	50±20%
air gap	1.5 mm or less
detection frequency	15kHz or less
insulation resistance	500M ohm or more

CONCLUSION

We have developed the technology for the epitaxial growth of InSb thin-films with high electron mobility directly on Si substrates and , using this technology, have realized InSb thin-film magneto-resistors with high heat resistance and high sensitivity and incorporated these into a rotation sensor with sufficient reliability for automobile use. In the future, we plan to establish the technology for mass production for automotive applications such as electronic control AT system, engine ignition system, and ABS.

ACKNOWLEDGMENTS

Finally, we greatly appreciate the staff of Ion Engineering Center K.K., whose facilities were used in developing this InSb thin-film MR rotation sensor, particularly in the basic study on InSb/Si epitaxial growth technology.

REFERENCES

[1] W.Thomson Phil. Trans. 146, p.736 (1956).

[2] H.Lippmann and F.Kuhrt Z.Naturforsch 13a, p.462 (1958).

[3] Toshiaki Fukunaka, et al. : The application and properties of thin InSb films for highly sensitive magneto-resistance elements. Toyo-tsushinnki Technical Report. No.40 (1987).

[4] J.L.Davis and P.E.Thompson : Molecular beam epitaxy growth of InSb films on GaAs. Appl. Phys. Lett. 54, p.2235 (1989).

[5] J.R. Soderstrom, et al. : Molecular beam epitxy growth and characterization of InSb layers on GaAs substrates. Semicond. Sci. Technol. 7, p.337 (1992).

[6] C. Besikci, et al. : Anomalous Hall effect in InSb layers grown by metalorganic chemical vapor deposition on GaAs substrates. J. Appl. Phys. 73, p.5009 (1993).

[7] D.L. Partin, et al. : Growth of high mobility InSb by metalorganic chemical vapor deposition. J. Electronic Ater. 23, p.75 (1994).

[8] T.S. Rao, et al.:Effect of substrate temperature on the growth rate and surface morphology of heteroepitaxial indium antimonide layers growth on (100) GaAs by metalorganic magnetron sputtering. J. Appl. Phys. 65, p.585 (1989).

[9] J.C. Chen, et al. : Low-temperature heteroepitaxial growth of InSb on CdTe by metalorganic chemical vapor deposition. Appl. Phys. Lett. 53, p.773 (1988).

[10] Toshinori Takagi, et al. : Preperation of InSb and GaAs thin films by ionized-cluster beam deposition. Vacuum. 22, No.7 p.15 (1979).

[11] Ichiro Shibasaki Nikkakyo-Geppou (in Japanese). May, p.12 (1988)

[12] J.I. Chyi, et al. : Molecular beam epitxy growth and characterization of InSb on Si. Appl. Phys. Lett. 54, No11, p.1016 (1989).

[13] A.Many, et al. : Semiconductor surfaces. (North Holland, Amsterdom), p.307 (1965)

[14] P.K. Chiang and S.M. Bedair : Growth of InSb and InAs1-xSbx by OM-CVD. J.Electrochem. Soc.131, p.2422 (1986).

[15] D.L. Dexter and F.Seitz : Effect of dislocations on mobilities in semiconductors. Phys.Rev. 86, p.964 (1952).

[16] Tetsuo Kawasaki, et al. : Epitaxial growth of InSb on hydrogen-terminated Si (111). Extended Abstract. J. Appl. Phys. 29a-ZR-1 (1993).

970602

Applications of Magnetoresistive Sensors in Navigation Systems

Michael J. Caruso
Honeywell Inc.

ABSTRACT

Most navigation systems today use some type of compass to determine heading direction. Using the earth's magnetic field, electronic compasses based on magnetoresistive (MR) sensors can electrically resolve better than 0.1 degree rotation. Discussion of a simple 8-point compass will be described using MR sensors. Methods for building a one degree compass using MR sensors will also be discussed. Compensation techniques are shown to correct for compass tilt angles and nearby ferrous material disturbances.

INTRODUCTION

The magnetic compass has been used in navigation for centuries. The inventor of the compass is not known, though evidence suggests that the Chinese were using lodestone—a magnetic iron ore—over 2000 years ago to indicate horizontal directions. It appears that Mediterranean seamen of the 12th century were the first to use a magnetic compass at sea [1]. Today, the balanced needle compass is only a slight variation of this early discovery. Advances in technology have led to the solid state electronic compass based on MR magnetic sensors and acceleration based tilt sensors. Electronic compasses offer many advantages over conventional "needle" type or gimbaled compasses such as: shock and vibration resistance, electronic compensation for stray field effects, and direct interface to electronic navigation systems. Two types of compasses will be discussed in this paper—a basic eight-point compass and a one-degree compass.

EARTH'S MAGNETIC FIELD

The earth's magnetic field intensity is about 0.5 to 0.6 gauss and has a component parallel to the earth's surface that always point toward magnetic north. This is the basis for all magnetic compasses. The key words here are "parallel to the earth's surface" and "magnetic north".

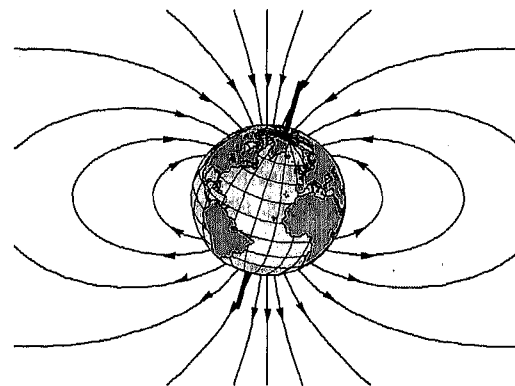
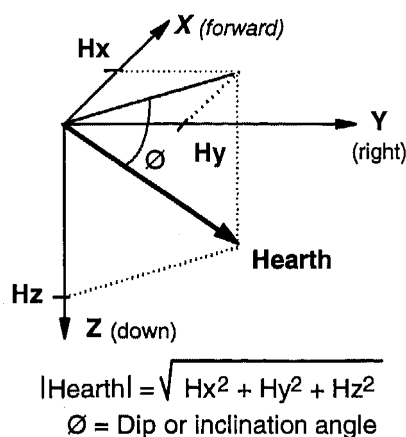


Figure 1 - Earth's Magnetic Field vs. True North

The earth's magnetic field can be approximated with the dipole model shown in Figure 1. This figure illustrates that the earth's field points down toward north in the northern hemisphere, is horizontal and pointing north at the equator, and point up toward north in the southern hemisphere. In all cases, the direction of the earth's field is always pointing to magnetic north. It is the components of this field that are parallel to the earth's surface that are used to determine compass direction. The angle of the magnetic field to the surface of the earth is called the dip, or inclination, angle (see Figure 2). In the northern hemisphere, the dip angle is roughly 70° down toward the north. Only the X and Y components of the earth's field is used when determining the azimuth, or compass direction. The vertical portion of the earth's magnetic field is ignored.

The term magnetic north refers to the earth's magnetic pole position and differs from true, or geographic, north by about 11.5 degrees. True north is at the earth's rotational axis and is referenced by the meridian lines found on maps. At different locations around the globe magnetic north and true north can differ by ±25 degrees, or more as shown in Figure 3. This difference is called the declination angle and can be determined from a lookup table based on the geographic location.



$$|Hearth| = \sqrt{Hx^2 + Hy^2 + Hz^2}$$

Ø = Dip or inclination angle

Figure 2 - Earth's Field in X, Y, Z Coordinates

The key to accurately finding a compass heading, or azimuth, is a two step process: 1) determine the Hx and Hy horizontal components of the earth's magnetic field and 2) add or subtract the proper declination angle to correct for true north.

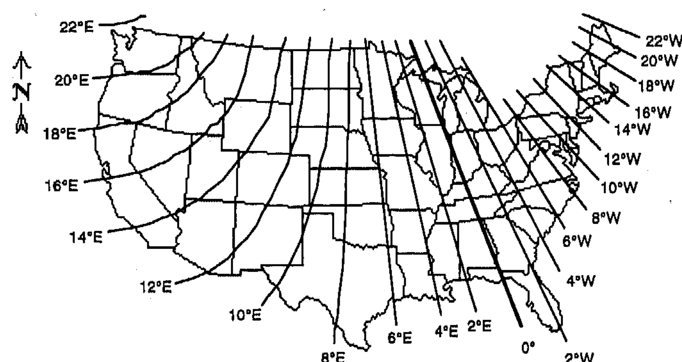


Figure 3 - Declination Angle To Correct For True North

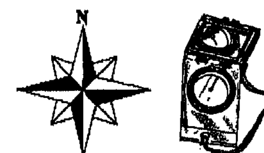
BASICS OF MAGNETIC SENSING

Today, there are several types of electronic compasses to chose from: fluxgate, magnetoresistive, magnetoinductive, and others. A common type of magnetic compass for navigation systems is the fluxgate sensor. The fluxgate sensor consists of a set of coils around a core with excitation circuitry that is capable of measuring magnetic fields with less than 1 milligauss resolution. These sensors provide a low cost means of magnetic field detection; they also tend to be bulky, somewhat fragile, and have a slow response time. Sometimes, fluxgate sensors in motion can have a reading response time of 2-3 seconds. This reading delay may be unacceptable when navigating a high speed vehicle or an unmanned plane.

Another type of magnetic sensor is the magnetoresistive (MR) sensor. This sensor is made up of thin strips of permalloy (NiFe magnetic film) whose electrical resistance varies with a change in applied magnetic field. These sensors have a well defined axis of sensitivity and are mass produced as an integrated circuit. Recent MR sensors show sensitivities below 0.1 milligauss, come in small solid state packages, and have a response time less than 1 microsecond. These MR sensors allow reliable magnetic readings in moving vehicles at rates up to 1,000 times a second.

A magnetoresistive sensor will be used in the remainder of this paper to detect both the sign and magnitude of the earth's field as a voltage output. The sensor will also detect any stray field or field anomaly superimposed onto the earth's field. The magnetic sensor output will have an X, Y, and Z component referenced to the magnetic sensor, or compass, package. For our navigational reference: the X component will be in the forward looking direction, the Y component to the right, and the Z component will be down as shown in Figure 2.

COMPASS DESIGN



There are many forms of compasses used in navigation systems. Two forms will be discussed here that use magnetoresistive magnetic sensors—the eight point compass and the one-degree compass.

EIGHT-POINT COMPASS—A simple eight point compass depicts the cardinal points (N, S, E, W) and the midway points (NE, NW, SE, SW). This type of compass may be used for basic automotive use where the driver needs to know the general direction of travel. For this application, the magnetic sensor can be reduced to a two-axis sensor using only the X and Y axis. An automobile usually travels on a level surface, barring any hills or potholes, so that the X and Y sensors will directly measure the earth's Hx and Hy magnetic fields. The compass can be mounted on the dashboard with the X axis pointing straight ahead and the Y axis to the left. For now, ignore the magnetic effect of the car on the earth's field.

The compass design can be broken into eight regions to indicate the cardinal directions. To analyze the magnetoresistive sensor response, plot the X and Y outputs while the car travels in a circle as indicated in Figure 4. Knowing that the earth's magnetic field is always pointing toward the north, start the analysis with the X axis (and the car) directly pointing north. The X output will be at its maximum value while the Y output will be zero—since no part of the earth's field is pointing to the left, or west. As the car turns clockwise toward the

east, the X axis will diminish to zero while the Y axis will decrease to its maximum negative value. With the car continuing its clockwise turn to point due south, the X axis will decrease to its most negative value while the Y axis will return to zero. This effect is illustrated in Figure 4 and shows the complete circular cycle for the X and Y axis outputs. The X and Y outputs of the magnetometer can be modeled by the cos(Ø) and sin(Ø) functions where Ø is the azimuth, referenced to magnetic north.

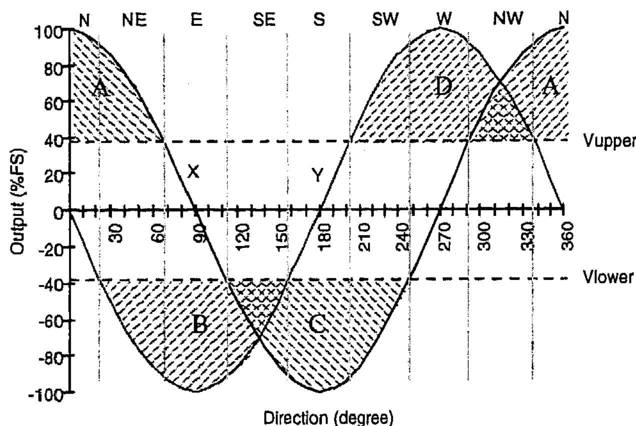


Figure 4 - Magnetic Outputs X And Y For 360° Rotation

The X and Y curves in Figure 4 can be split into eight regions representing the four cardinal and four midway points. A combination of these curves can be formed to represent each region. Two crossover points, Vupper and Vlower, are necessary to distinguish the boundaries of the eight compass direction headings. The crossover points can be determined by knowing the full scale (FS) values for X and Y as:

Vupper = 100*sin(22.5°) (%FS) = 38 %FS (1)
Vlower = -100*sin(22.5°) (%FS) = -38 %FS

Voltage comparators can be used to detect Vupper and Vlower levels to divide the X and Y curves into four regions: A, B, C, and D. The eight points of the compass can be determined by combining the A, B, C, and D using Boolean logic gates, four comparators, and a two axis magnetic sensor as shown in Figure 5. This circuit requires a two axis magnetometer with a signal sensitivity of 1-2 milligauss. Magnetic hysteresis and linearity must be less than 1-2%FS with good repeatability. There are three limitations to consider while using this design: 1) there is no tilt compensation so the compass must be held level, 2) there should be no nearby ferrous material to create magnetic distortions, and 3) the declination angles are difficult to add to this design. These limitations will be addressed in the one-degree compass discussion below.

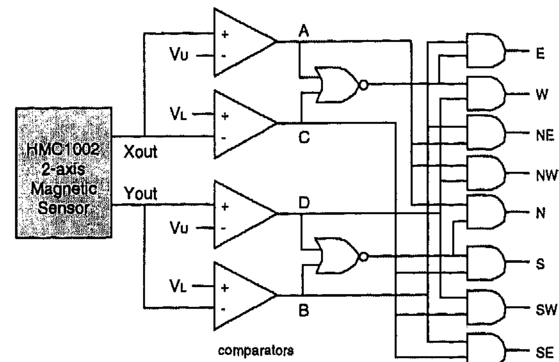


Figure 5 - Eight Point Compass Circuit

ONE-DEGREE COMPASS—Some navigation systems require more than just an eight point compass. For instance, the Global Positioning System (GPS) has led to a sophisticated tracking of vehicle position on video maps with accuracy better than 10 meters. These systems rely on telemetry contact from four satellites, sometimes aided by a system radio tower. It is essential to keep a line of sight with these satellites for position determination. Backup systems are required in cities and tunnels to maintain a course of direction during short blackouts. This is where a more accurate compass can help in GPS based navigation systems. During the loss of GPS signals, knowing the vehicle's speed and heading direction can maintain proper vehicle tracking. Gyros can be used to maintain direction but a lower cost MR based compass is preferred. For these systems, a compass accuracy of one degree is desirable.

To achieve a one degree accurate compass requires a magnetic sensor that can reliably resolve angular changes to 0.1 degrees. The sensors must also exhibit low hysteresis (<0.05%FS), a high degree of linearity (< 0.5%FS error) and be repeatable. The magnetic fields in the X and Y plane will typically be in the 200 to 300 milligauss range—more at the equator, less at the poles. Using the relationship:

Azimuth = arcTan (y/x) (2)

the required magnetometer resolution can be estimated. To resolve a 0.1° change in a 200milligauss field would require a magnetic sensitivity of better than 0.35 milligauss. Solid state MR sensors are available today that reliably resolve 0.07 milligauss signals giving a five times margin of detection sensitivity.

Using the simple magnetic sensor shown in Figure 6, the azimuth can be calculated by using the X and Y outputs in a horizontal plane. To account for the tangent function being valid over 180° and not allowing the y=0 division calculation, the following equations can be used:

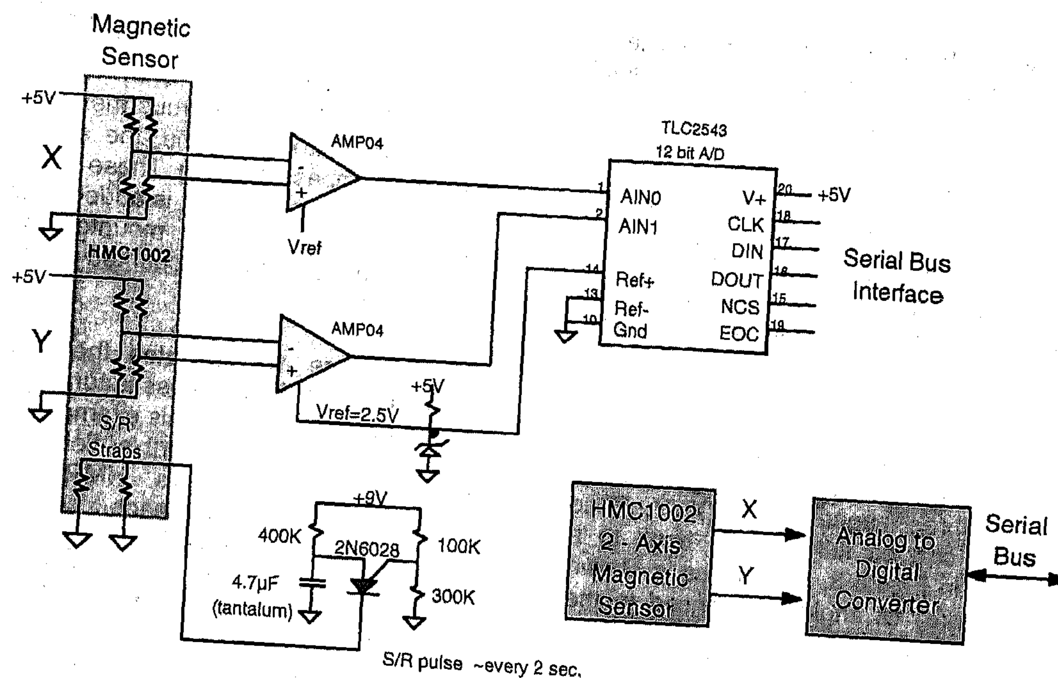


Figure 6 - One Degree Compass System without Tilt Compensation

Azimuth (x=0, y<0) = 90.0 (3)
Azimuth (x=0, y>0) = 270.0
Azimuth (x<0) = 180 - [arcTan(y/x)]*180/•
Azimuth (x>0, y<0) = - [arcTan(y/x)]*180/•
Azimuth (x>0, y>0) = 360 - [arcTan(y/x)]*180/•

The set/reset (S/R) circuit shown in Figure 6 is a current pulse generator used to eliminate the effects of past magnetic effects and temperature drift [4]. The serial bus output can readily interface to a low cost microprocessor for azimuth computation. Equations (3) provide continuous azimuth angles from 0° to 360° in the forward direction relative to magnetic north (H_{North}), see Figure 7. In this example, there is no compensation for tilt and nearby ferrous distortion effects on the azimuth.

COMPENSATING FOR TILT—Most often compasses are not confined to a flat and level plane. They are often hand held, attached to an aircraft, or on a vehicle in an uneven terrain. This makes it more difficult to determine the azimuth, or heading direction, since the compass is not always horizontal to the earth's surface. Errors introduced by tilt angles can be quite large depending on the amount of the Dip angle. A typical method for correcting the compass tilt is to use an inclinometer, or tilt sensor, to determine the roll and pitch angles. The terms roll and pitch are commonly used in aviation: ROLL refers to the rotation around the X, or forward direction, and PITCH refers to the rotation around the y, or left-right, direction (see Figure 8).

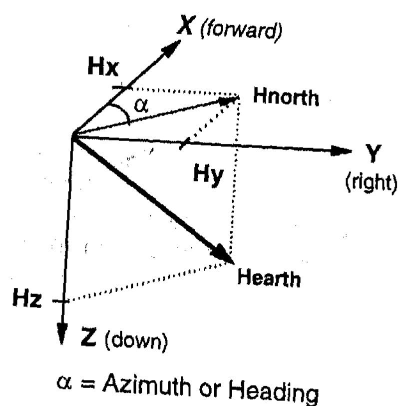


Figure 7 - Azimuth Defined In The X-Y Plane

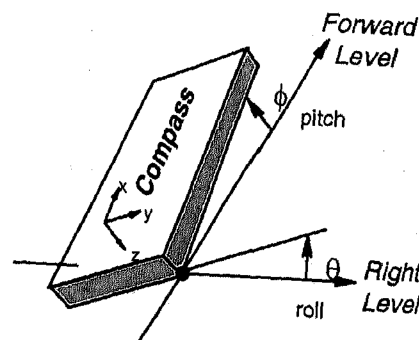


Figure 8 - Compass Tilt Referenced To The Earth's Horizontal Plane

Common liquid filled tilt sensors resemble a glass "thimble" that uses electrodes to monitor the fluid movement as the sensor changes angles. Newer solid state accelerometer tilt sensors are available that measure the earth's gravitational field by means of an electromechanical circuit [5]. The output of these devices are an electrical signal equivalent to the angle of tilt. During compass assembly, the tilt sensor directions must be carefully aligned with the X,Y,Z magnetic axis. Several manufacturers offer these tilt sensors as stand alone circuit boards that provide the roll and pitch angles as outputs.

To compensate a compass for tilt, knowing the roll and pitch is only half the battle. The magnetometer must now rely on all three magnetic axes (X, Y, Z) so that the earth's field can be fully rotated back to a horizontal orientation. In Figure 8, a compass is shown with roll (θ) and pitch (ϕ) tilt angles referenced to the right and forward level directions of the observer or vehicle. The X, Y, and Z magnetic readings can be transformed back to the horizontal plane (X_H, Y_H) by applying the rotational equations shown below:

$$X_H = X^*\cos(\phi) + Y^*\sin(\theta)^*\sin(\phi) - Z^*\cos(\theta)^*\sin(\phi)$$

$$Y_H = Y^*\cos(\theta) + Z^*\sin(\theta) \qquad (4)$$

$$\text{Azimuth} = \arctan(Y_H / X_H)$$

Once the X and Y magnetic readings are in the horizontal plane, equations (3) can be used to determine the azimuth. For speed in processing the rotational operations, a sine and cosine lookup table can be stored in program memory to minimized computation time.

A block diagram for a tilt compensated compass is shown in Figure 9 with a serial bus interface. After the azimuth is determined, the declination correction can be applied to find true north according to the geographic region of operation.

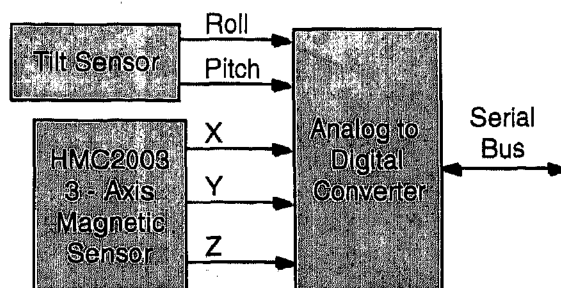


Figure 9- Tilt Compensated Compass System

COMPENSATING FOR NEARBY FERROUS EFFECTS—When a compass is operating in a open area in the absence of any ferrous metals there is no distortion effects on the earth's magnetic field. In reality, though, compasses are mounted in vehicles, aircraft, and platforms that most likely have ferrous materials nearby. The effects of ferrous metals (iron, nickel, steel, cobalt) will distort, or bend, the earth's field which will alter the compass heading. These effects can be thought of as a magnetic field that is added to the earth's field. If the compass is securely mounted in the vehicle, the ferrous effects can be accounted for and removed from the magnetic readings.

Figure 10 illustrates the X and Y magnetic readings when the compass is turning around in a circle in a horizontal plane. In this example, there is no ferrous interference with the earth's field. The readings are taken from Honeywell's HMR2300 Smart Digital Magnetometer where each count represents 67 microgauss. The earth's field magnitude in the X and Y plane reads 2800 counts which is approximately 190 milligauss. When the X and Y readings are plotted with each other they form a circle centered about the 0,0 point. An azimuth can be determined for each reading using equations (3) as show in Figure 10. This plot shows a sine and cosine output response for the X and Y directions during rotation.

If the magnetometer is mounted in a car, the effect of the engine and car body would distort the earth's magnetic field. Driving the car in a circle would produce the curves shown in Figure 11. Note here that the X,Y plot is not a circle (slightly ellipsoid) and that it is offset from the 0,0 point by -480 and -795 counts. This offset and ellipsoid effect are a result of the fixed distortion of the car on the earth's magnetic field. This distortion can be determined systematically and applied to subsequent X,Y readings to eliminated the effects of the car.

To compensate for the vehicle's distortion, two scale factors Xsf and Ysf can be determined to change the ellipsoid response to a circle. Offset values Xoff and Yoff can then be calculated to center the circle around the 0,0 origin. The X,Y values used to compute the azimuth when compensating for the vehicle's distortion are:

$$\text{Xvalue} = \text{Xsf} * \text{Xreading} + \text{Xoff} \qquad (5)$$

$$\text{Yvalue} = \text{Ysf} * \text{Yreading} + \text{Yoff}$$

Here, the scale factors Xsf and Ysf scale each reading to change the ellipsoid to a circle and Xoff and Yoff values shift the center back to the 0,0 origin. The result of this compensation is shown in Figure 12 and should be compared to the 'no interference' curves in Figure 10.

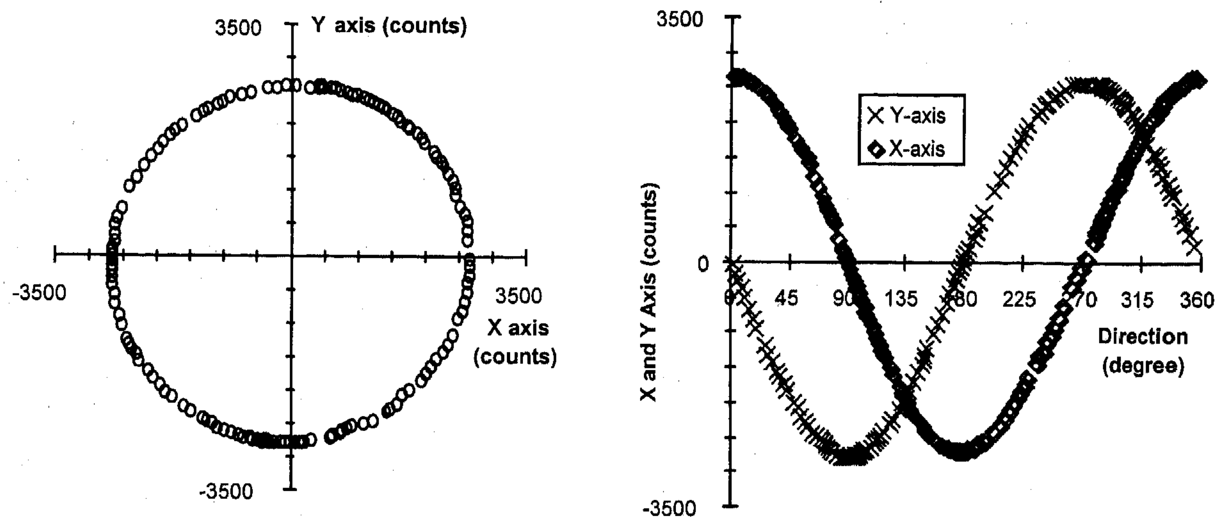


Figure 10 - No Interference Of Magnetometer Readings For 360° Rotation In Level Plane

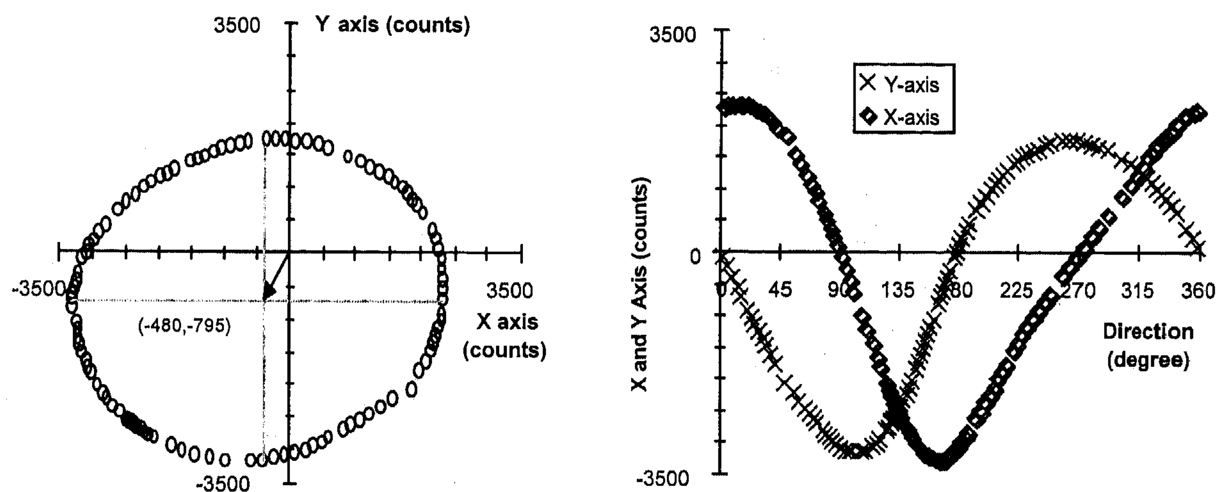


Figure 11 - Car Engine/Body Interference Of Magnetometer Readings For 360° Rotation In Level Plane

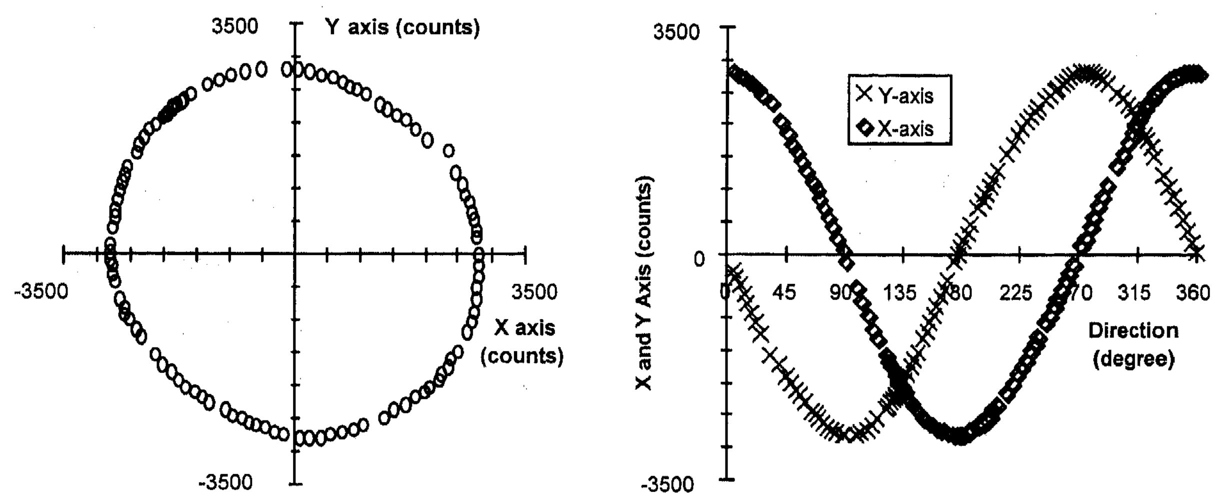


Figure 12 - Car Engine/Body Compensation Of Magnetometer Readings For 360° Rotation In Level Plane

A simple calibration method can be used to determine the offset and scale factor values:

- Mount the compass in the car and drive the car in a circle on a horizontal surface.
- Find the maximum and minimum values of the X and Y magnetic readings.
- Using these four values determine the X and Y scale factors (Xsf, Ysf) and the zero offset values (Xoff, Yoff).

Xsf = 1 or (Ymax - Ymin) / (Xmax - Xmin) (6)
whichever is greater

Ysf = 1 or (Xmax - Xmin) / (Ymax - Ymin)
whichever is greater

Xoff = [(Xmax - Xmin)/2 - Xmax] * Xsf (7)

Yoff = [(Ymax - Ymin)/2 - Ymax] * Ysf

The following example will show how the compensation values are determined. A compass is mounted in a car that has traveled a circle in a vacant parking lot. The magnetic X and Y counts (15,000 counts=1 gauss) from the magnetometer are scanned and the minimum and maximum readings are:

Xmin = -3298 Xmax = 2338

Ymin = -3147 Ymax = 1763

Set the X scale factor (Xsf) to one since (Ymax - Ymin)/(Xmax - Xmin)< 1, according to equation (5). Next, determine the Y scale factor (Ysf) by dividing the X reading span by the Y reading span.

Xsf = 1

Ysf = (Xmax - Xmin) / (Ymax - Ymin) = 1.15

Calculate the offset correction values by taking one-half the difference of the max. minus min. readings and apply the scale factors, Xsf and Ysf.

Xoff = [(Xmax - Xmin)/2 - Xmax] * Xsf = 480

Yoff = [(Ymax - Ymin)/2 - Ymax] * Ysf = 795

Store these values and apply them to every tilt compensated reading—X_H and Y_H. The Xvalue and Yvalue numbers used in the azimuth calculations, equations (3), to determine compass heading are:

Xvalue = X_H + 480

Yvalue = 1.15 * Y_H + 795

CONCLUSION

Compass headings are determined from the earth's magnetic fields in a horizontal plane. In a compass system each magnetometer reading must first be corrected for tilt. Then each reading must be compensated for the effects of nearby ferrous materials. Once the compass reading is tilt compensated and nearby ferrous material corrected, the declination angle should be applied to adjust magnetic north to true north.

Magnetoresistive sensors provide a solid state solution for building compass navigation systems. Their high sensitivity and good repeatability, along with small size, results in a high accuracy and easy to integrate magnetic sensor. There are many other techniques than the ones shown here for tilt and ferrous distortion compensation. The method for compensation depends on the application requirements: accuracy, resolution, speed, size, and cost.

REFERENCES

[1] Grant, George A., and Klinkert, John, The Ship's Compass, 2d ed. (1970).

[2] Barber, G.W., and Aarott,A.S., History and Magnetics of Compass Adjusting, IEEE Trans. Magn. Nov. 1988.

[3] Olson, Gregory J., et al, Nongimbaled Solid-State Compass, Solid-State Sensor and Actuator Workshop, June, 1994.

[4] Set/Reset Pulse Circuits for Magnetic Sensors, Honeywell Applications Note, AN-201.

[5] A Dual Axis Tilt Sensor Based on Micromachined Accelerometers, Mike Horton, Charles Kitchin, Sensors Magazine, April, 1996.

Unit conversion from SI to Gaussian:

1 Tesla= 10^4 gauss = 10^9 gamma = 7.96×10^5 A/m,

1 nTesla = 10 μgauss = 1 gamma = 7.96×10^{-4} A/m

rev.A

970603

Giant Magnetoresistance Materials and Integrated Magnetic Sensors

Carl H. Smith and Jay L. Brown
Nonvolatile Electronics, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Large changes in resistance with magnetic field exhibited by recently discovered thin-film multilayer structures offer new opportunities in the sensors field. Compatibility of these Giant Magnetoresistive (GMR) materials with conventional semiconductor processing technology and semiconductor underlayers allows the production of inexpensive, ultra-small, smart sensors with enhanced magnetic sensitivity. The action of GMR materials is described along with their application to sensor configurations for detecting magnetic fields, magnetic field gradients, and magnetic fields from currents. Characteristics of these sensors are discussed and compared to conventional magnetic field sensors. Finally, the exciting combination of GMR materials with semiconductor underlayers in smart sensors with linear or digital (switched) outputs is explored.

INTRODUCTION

The discovery of any new physical phenomenon which relates some condition imposed on a material to a measurable property of that material often leads to a new class of sensors. Such is the case with Giant Magnetoresistive (GMR) materials. Initial observations of large magnetic field dependent changes in resistance of thin-film ferromagnet/nonmagnet metallic multilayers in 1988, were followed by the commercial introduction of magnetic sensors based on GMR materials in 1994[1, 2]. Intense research continues on a wide variety of new materials which utilize the phenomenon of spin-dependent conduction electron scattering as well as research understanding the phenomenon itself. Efforts to develop applications using GMR materials include not only magnetic field sensors, but also read heads for high density magnetic data recording and memory elements for magnetic random access memory (MRAM).

This paper first describes the properties of some of the GMR materials which have already proved useful for sensors as well as a brief discussion of how GMR works. Various sensor configurations using GMR materials are examined as well as the integration of GMR elements with semiconductor materials. Finally, utilization of GMR sensors in specific applications is discussed and GMR sensors are compared with conventional magnetic field sensors.

GMR MATERIALS

BASIC PRINCIPLES - Resistance of metals depends on the mean free path of their conduction electrons. The shorter the mean free path, the higher the resistance of the metal. A voltage applied to a conductor results in an electrical field in the conductor which accelerates the conduction electrons. Current density in the conductor depends upon the average velocity of conduction electrons and on their number density. The more frequently they are scattered and lose their momentum, the lower their average velocity. Scattering can occur in the bulk or at surfaces. The resistivity of thin films can be considerably larger than the bulk resistivity if the film thickness is less than the mean free path of the conduction electrons.

Ferromagnetic materials introduce an additional complication to the behavior of conduction electrons. Electrons can be either *spin up* if their spin is parallel to the magnetic moment of the ferromagnet or *spin down* if they are antiparallel. In nonmagnetic conductors there are equal numbers of *spin up* and *spin down* electrons in all energy bands. However, due to the ferromagnetic exchange interaction, there is a difference between the number of *spin up* and *spin down* electrons in the conduction sub bands of ferromagnetic materials. Therefore, the probability of an electron being scattered when it passes into a ferromagnetic conductor depends upon the direction of its spin. If two thin ferromagnetic layers are separated by a thin nonmagnetic conducting layer, we can change the resistance by simply changing whether the moments of the ferromagnetic layers are parallel or antiparallel. Layers with parallel magnetic moments will have less scattering at the interfaces and, therefore, longer mean free paths, and lower resistance. Layers with antiparallel magnetic moments will have more scattering at the interfaces and, therefore, shorter mean free paths, and higher resistance. In order for this

spin dependent scattering to be a significant part of the total resistance, the layers must be thinner than the mean free path of electrons in the bulk material. For many ferromagnets the mean free path is tens of nanometers, so the layers themselves must each be typically less than 10 nm (100 Å). For a more complete treatment of spin-polarized transport see the recent article by G. Prinz in *Physics Today*[3]. Figure 1 illustrates the concept of spin dependent scattering at interfaces.

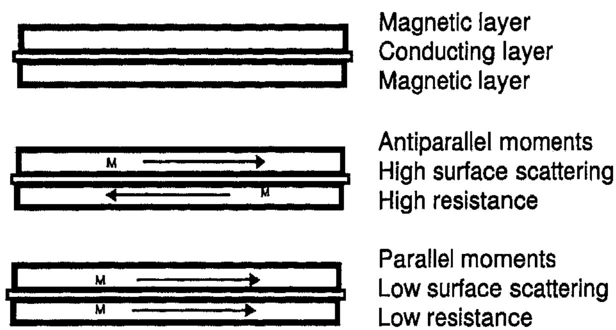


Fig. 1. Scattering from two different alignment of magnetic moments in a two magnetic layer GMR sandwich.

GMR STRUCTURES - There are several methods of obtaining antiparallel alignment of the magnetic layers. In this paper we will discuss the two which are currently used in GMR sensors: *sandwich* GMR materials sometimes called unpinned spin valves, and antiferromagnetically coupled *multilayer* GMR materials. Both of these materials are typically RF diode sputtered onto atomically flat substrates such as silicon nitride on single crystal silicon wafers common in the integrated circuit industry.

Unpinned sandwich GMR materials - consist of two soft magnetic layers of iron, nickel and cobalt alloys separated by a layer of a nonmagnetic conductor like copper. With magnetic layers 4 to 6 nm (40 to 60 Å) thick separated by a conductor layer typically 3 to 5 nm thick there is relatively little magnetic coupling between the layers. For use in sensors sandwich material is usually patterned into narrow stripes. The magnetic field caused by a current of a few mA per µm of stripe width flowing along the stripe is sufficient to rotate the magnetic layers into antiparallel or high resistance alignment. The right hand rule for magnetic fields from currents provides antiparallel alignment of the moments of the magnetic layers with the moments perpendicular to the current direction. An external magnetic field of 3 to 4 kA/m (35 to 50 Oe) applied along the length of the stripe is sufficient to rotate both the magnetic moments parallel to the external field lowering the resistance. A positive or negative external field parallel to the stripe will produce the same change in resistance. An external field applied perpendicular to the stripe will have little effect due to the demagnetizing fields associated with the narrow dimensions of a magnetic object. The value associated with the GMR effect is the percent change in resistance normalized by the saturated or minimum resistance. Sandwich materials have values of GMR typically 4 to 9 %. Figure 2 shows a typical resistance *vs.* field plot for sandwich GMR material.

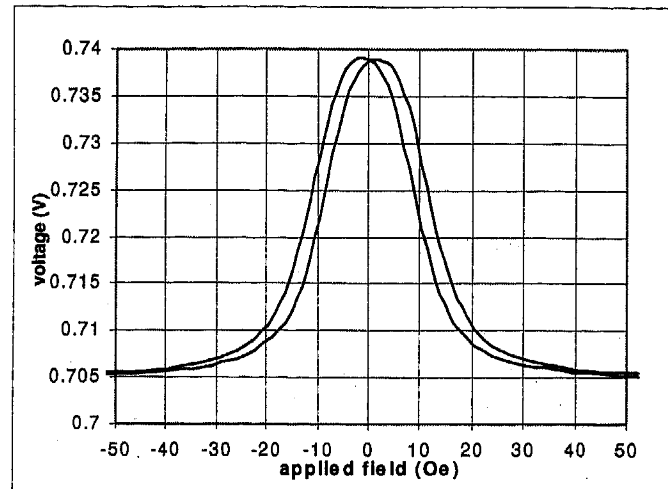


Fig. 2. Voltage *vs.* applied field for a 2 µm wide stripe of unpinned sandwich GMR material with 1.5 mA current. GMR = 5 %.

Antiferromagnetic multilayers - consist of multiple repetitions of alternating conducting magnetic layers and nonmagnetic layers. A typical multilayer structure used in magnetic sensors is shown in Figure 3.[4 5]

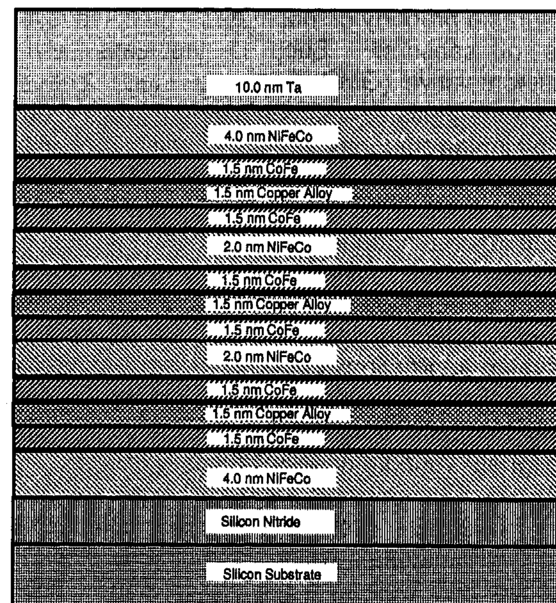


Fig 3. Layer structure of a typical commercial GMR multilayer.

The subdivision of the magnetic layers into two different alloy compositions improves the performance. With more interfaces than sandwiches, the size of the GMR effect is larger. The thickness of the nonmagnetic layers is less than that for sandwich material (typically 1.5 to 2.0 nm) and is quite critical. At the proper thickness, the polarized conduction electrons cause antiferromagnetic coupling between the magnetic layers. In other words, each magnetic layer has its magnetic moment antiparallel to the moments of the magnetic layers on each side – exactly the condition needed for maximum spin dependent scattering. Again, an external field can overcome the coupling which causes this alignment and can align the moments in all the layers parallel reducing the resistance. If the conducting layer is not the proper thickness, the same coupling mechanism can cause ferromagnetic coupling between the magnetic layers resulting in no GMR effect.

A plot of resistance *vs.* applied field for a multilayer GMR material is shown in Figure 4. Note the higher GMR value, typically 12 to 16 % for structures shown in Figure 3, and the much higher external field required to saturate the effect. Multilayer GMR materials have better linearity and lower hysteresis than typical sandwich GMR material.

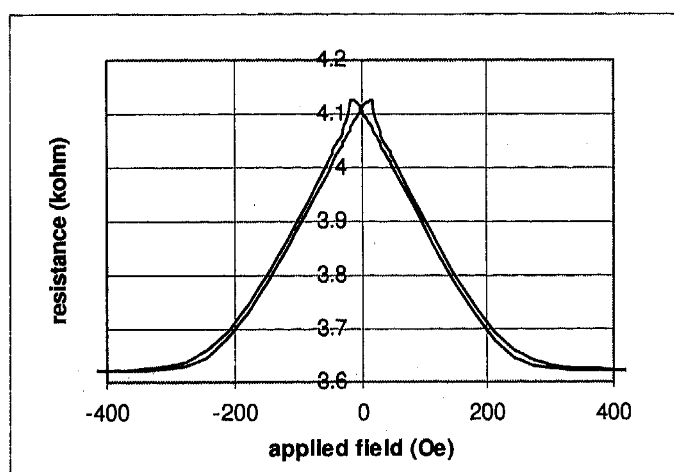


Fig. 4. Resistance *vs.* applied field for a 2 μm wide stripe of antiferromagnetically coupled multilayer GMR material. GMR = 14 %.

MATERIAL CHARACTERISTICS - Besides the basic characteristic of resistance *vs.* applied field, there are several other characteristics of GMR materials which are of importance when applying them to sensors.

Resistivity - Many sensor applications are power limited. This limitation is most severe in battery powered applications. Therefore, a high sensor resistance is desirable. GMR materials are thin film and, therefore have relatively high resistance measured in ohms per square. Typical values are between 12 and 16 ohms per square. Resistors can be patterned in very small areas. A serpentine 10 kΩ resistor can easily be fit into a square 100 μm space using GMR stripes with a 2 μm linewidth. In wafer-scale sensor fabrication, die size has a large influence on sensor cost.

Temperature coefficient of resistivity - Both sandwich and multilayer GMR materials have temperature coefficients of resistivity (TCR) of 0.14 to 0.16 %/°C (1400 to 1600 ppm/°Cz0 which are relatively linear with temperature. This value is about one-half that of nickel-iron permalloy used in other magnetoresistive sensors. The surface scattering which dominates the resistance in thin film GMR materials is less temperature dependent that is bulk scattering. The GMR effect itself is temperature dependent. The field at which the effect saturates decreases slightly with temperature as does the change in resistance. However, as will be shown in a later section, constant current drive can be used to produce a sensor which is quite temperature independent.

High temperature endurance - Sensors materials must not only endure use temperatures which are on the move upward above 150 °C in automotive applications, but also processing temperatures during packaging. The addition of semiconductor processing temperatures place an even greater short term high temperature limit on the sensor materials.

Unpinned sandwich GMR materials have relatively thick copper interlayers and have demonstrated stable properties to about 250 °C. Antiferromagnetically coupled multilayers with much thinner copper alloy interlayers have shown decrease in their GMR above 175 °C and degrade rapidly above 200 °C. Packaged multilayer sensors have survived 1000 hour dry storage tests. Figure 5 shows changes in GMR of multilayer and sandwich materials after 1 hour anneals at various temperatures.

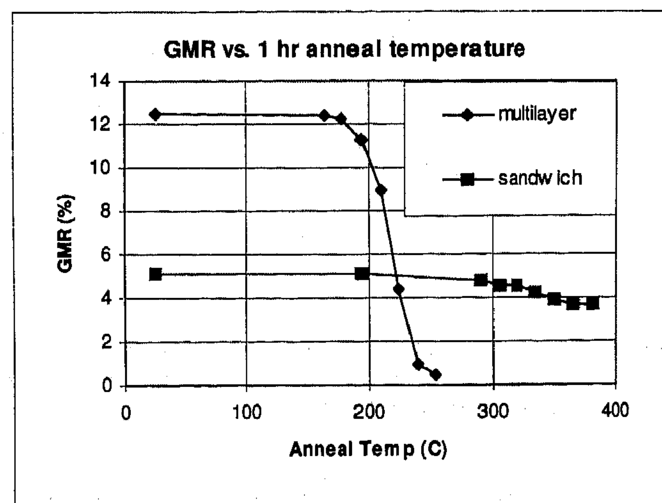


Fig. 5. GMR of sandwich and multilayer materials after 1 hour anneals at various temperatures.

SENSOR CONFIGURATIONS

FIELD SENSORS - Although simple GMR resistors and GMR half bridges can be fabricated from GMR materials, the best utilization of these materials appears to be in a traditional Wheatstone bridge configuration. Electroplated soft magnetic permalloy structures on top of many GMR sensors accomplish two objectives – shielding and flux concentration. Small magnetic shields plated over two of the four equal resistors in a Wheatstone bridge, say the resistor connected to power on one side and the resistor connected to ground on the other side, protect these resistors from the applied field and allow them to act as reference resistors. Since they are fabricated from the same material, they have the same temperature coefficient as the active resistors. The two remaining GMR resistors are both exposed to the external field. The bridge output is therefore twice the output from a bridge with only one active resistor. The bridge output for a 10% change in these resistors is approximately 5% of the voltage applied to the bridge.

Additional permalloy structures plated onto the substrate act as flux concentrators. The active resistors are placed in the gap between two flux concentrators as is shown in Figure 6. These resistors experience a field which is larger than the applied field by approximately the ratio of the gap between the flux concentrators, D1, to the length of one of the flux concentrators, D2. In some sensors the flux concentrators are also used as shields by placing two resistors beneath them as is shown for R3 and R4 in the figure below.

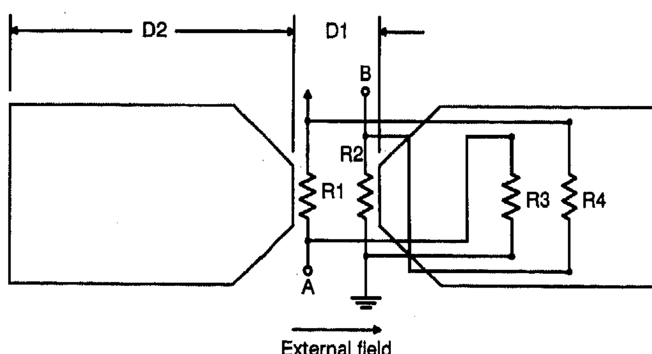


Fig. 6. Configuration of GMR resistors in a Wheatstone bridge sensor. Flux concentrators are shown as well as output between A and B.

The sensitivity of a GMR bridge sensor can be adjusted in design by changing the lengths of the flux concentrators and the gap between them. In this way, a GMR material which saturates at approximately 300 Oe can be used to build different sensors which saturate at 15, 50, and 100 Oe.

An example of the output from a GMR bridge sensor is shown is Figure 7. The curves on the left side represent the output of the bridge at various temperatures with a constant voltage applied to the bridge. Only negative excursions of the external field are shown without crossing zero to avoid hysteresis. The curves on the right side of Figure 7 represent the output of the same bridge when a constant current supply is used. The maximum output still decreases with temperature, but not by as large an amount since the voltage across the bridge automatically increases as the bridge resistance increases with temperature. The output at any constant field which is less than saturation remains almost constant.

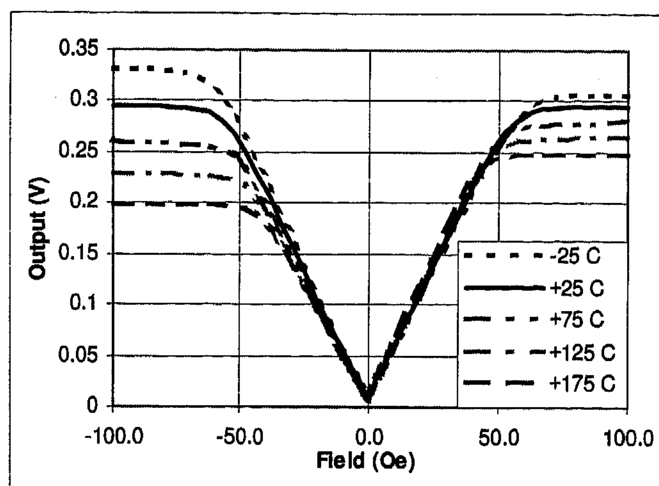


Fig. 7. Output of a GMR Wheatstone bridge *vs.* external field at various temperatures. The left side represents a bridge driven with a constant voltage, the right side a bridge driven with a constant current.

BIASING - Since the resistance of GMR material decreases when an external field is applied and the magnetic moments of the layers can be aligned in either direction for minimum resistance, a bridge circuit produces a symmetrical or omnipolar output. If a knowledge of the direction of a bipolar magnetic field is required, methods of biasing the sensor away from zero are necessary. This biasing can be done either with a small permanent magnet affixed to the top of the sensor package or by electromagnetic means using a coil or a current strap to apply a constant magnetic field to the sensor. Biasing the GMR material away from zero also decreases the hysteresis and non-linearity observed close to the zero field point. The offset caused by a bias can be handled either by ac coupling of the output if there is no dc signal of interest, or by building a negative offset (output at zero field) into the bridge by using

unequal resistors. An additional bias then brings the output back to zero.

GRADIENT FIELD SENSORS - A Wheatstone bridge can be made with GMR resistors without shields or flux concentrators. If the upper left and lower right resistors are located at one end of the chip and the other resistors at the other end of the chip the device becomes a gradiometer. The output of the bridge is due to the *difference* between the resistances of the two sets of resistors. A constant field will change all four resistors by the same amount leaving the bridge in balance. A non-uniform field, however, has a gradient and the two pairs of resistors will experience different fields resulting in an imbalance in the bridge; the bridge is sensitive to the field gradient. Such sensors respond to fields which change very rapidly with position such as those found near magnetized gear teeth. They are insensitive to uniform fields as long as the fields are not large enough to saturate all of the resistors. Gradient field sensors will be discussed under angular position sensing

Half bridges consisting of two GMR resistors separated in position and connected in series with a center tap can also be used as gradiometers. The zero field output at the center tap is one half of the applied voltage. In the presence of a non-uniform magnetic field, the output will vary above and below its zero field value by an amount proportional to the difference in field between the two resistors. In a uniform field both resistors will change by the same amount and the output will remain at the zero field value.

SENSOR APPLICATIONS

GENERAL COMMENTS - A magnetic field sensor can directly sense a magnetic field from a permanent magnetic, an electromagnet, or a current. In sensing the presence of a ferrous object, a biasing magnet is often used. The biasing magnet magnetizes the ferromagnetic object such as a gear tooth, and the sensor detects the combined magnetic fields from the magnetized object and the biasing magnet. A biasing magnet is affixed to the sensor in a position such that its direct influence on the sensor is minimal. Usually the biasing magnet is mounted on the top of the sensor with its magnetic axis perpendicular to the sensitive axis of the sensor. The biasing magnet can be centered such that there is little or no field in the sensitive direction of the sensor. In this way a reasonable large biasing magnet can be used. Occasionally a spacer is used between the sensor and the magnet to reduce the field at the sensor and, therefore, reduce how critically the magnet must be positioned. Figure 8 shows the relative positions of a sensor and a biasing magnet. The magnetic field lines are shown both in the absence and presence of a ferromagnetic object. Note the induced magnetic moment in the ferromagnetic object. The technique of using a biasing magnet can only be used if the ferrous object is in close proximity. In some applications such as detection of motor vehicles, the Earth field acts as a biasing magnet resulting in a magnetic dipole moment along the length of the motor vehicle.

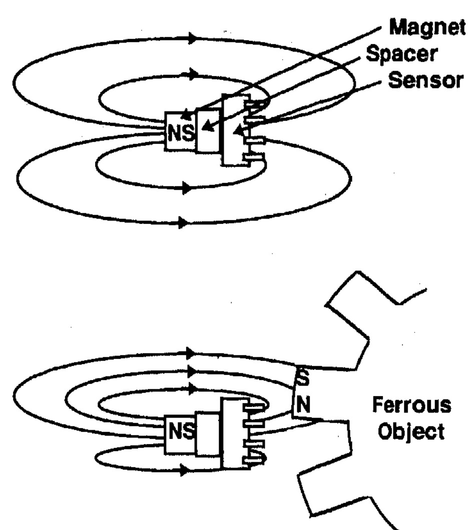


Fig. 8. Side view of biasing magnet and sensor in 8 pin package shown with and without a ferrous object present.

FIELD SENSING - GMR sensors can be used to measure or sense magnetic fields. As was pointed out in a section above, provisions for biasing the sensor away from zero field are necessary if bipolar fields are to be sensed. The high sensitivity of GMR sensors with high flux concentration ratios allow them to detect the fields from small magnetized particles such as those included in the ink on various countries' currencies and the magnetic layer on Microtaggant particles.[6]

DISPLACEMENT SENSING- GMR bridge sensors can be effectively employed to provide position information from small displacements associated with actuating components in machinery, proximity detectors, and linear position transducers. Due to the nonlinear characteristic of dipole magnetic fields produced by permanent magnets, the range of linear output may be limited. A dipole field decreases as the reciprocal of the distance cubed. Figure 9 shows the position and motion of a sensor relative to a cylindrical permanent magnet. In both cases the sensitive axis of the sensor is in the X direction. Sensors A and B can both detect motion in either the X or Y direction but not both at the same time. The sensitive axis of the sensor is indicated by the double headed arrow on each sensor. The rate of change of the X component of the magnetic field (the sensitive axis of both sensors) with position can be

approximated from the formula for the magnetic field from a dipole if the sensor is at least one or two diameters away from the permanent magnet.

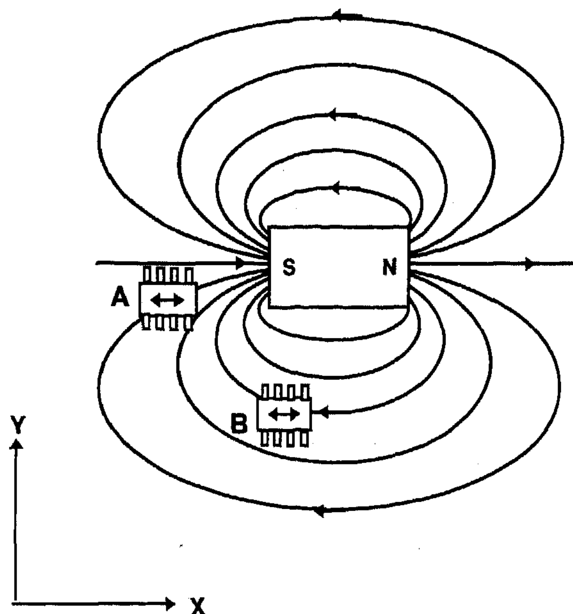


Fig. 9. Sensors positioned to measure displacement relative to a permanent dipole magnet. Sensitive axis indicated on sensors **A** and **B**.

ANGULAR POSITION SENSING - GMR sensors offer a rugged, low cost solution to rotational reference detection. High sensitivity and dc operation afford the GMR bridge sensor an advantage over inductive sensors which have very low outputs at low frequencies and can generate large noise signals when subjected to high frequency vibrations. GMR sensors are field sensors and do not measure the induced signal from the time rate of change of fields as do variable reluctance sensors. The output from a GMR bridge sensor will have a minimum when the sensor is centered over a tooth or a gap and a maximum when a tooth approaches or recedes. Figure 8 shows a bridge sensor in position for angular position sensing.

Higher accuracy sensing of the edge of a gear tooth can be achieved by using a gradiometer. A properly biased gradiometer will produce a maximum and a minimum at each tooth edge and minima over the center of each tooth and gap. Figure 10 shows the output from a gradiometer biased with a permanent magnet.

CURRENT SENSING - Currents in wires create magnetic fields surrounding the wires or traces on printed wiring boards. The field decreases as the reciprocal of the distance from the wire. GMR bridge sensors can be effectively employed to sense this magnetic field. Both dc and ac currents can be detected in this manner. Bipolar ac current will be rectified by the sensors omnipolar sensitivity unless a method is used to bias the sensor away from zero. Unipolar and pulsed currents can be measured with good reproduction of fast rise time components due to the excellent high frequency response of the sensors. Since the films are extremely thin, response to frequencies above a megahertz is possible. Figure 11 shows the relative position of a GMR bridge sensor and a current carrying wire to detect the current in the wire. A wire placed immediately over or under the sensor will produce a field of approximately 0.080 A/m (one mOe) per mA of current. The sensor can also be mounted immediately over a current carrying trace on a circuit board. High currents may require more separation between the sensor and the wire. Low currents may be best detected with the current being carried by a trace immediately over the GMR resistors.

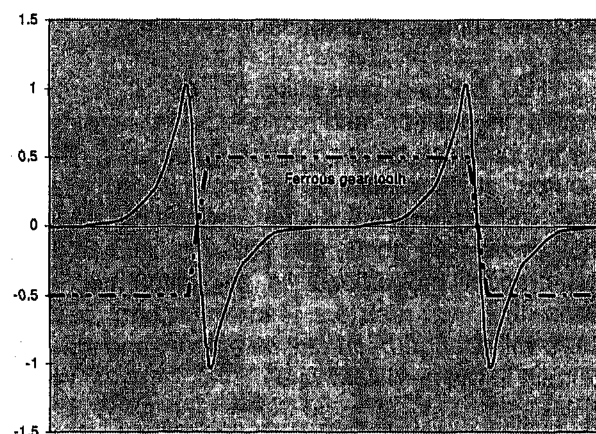


Fig. 10. Output from a biased gradiometer used for gear tooth detection. The relative position of the gear tooth is shown superimposed on the output.

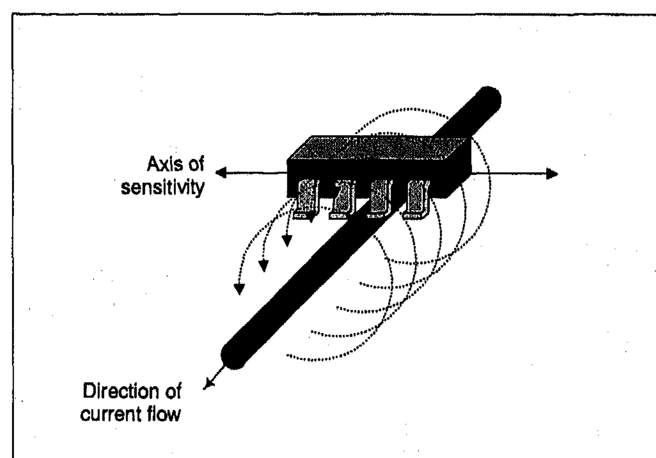


Fig. 11. Proper position of a GMR bridge sensor to detect the magnetic field created by a current carrying wire.

SENSOR COMPARISONS

Various methods of magnetic field sensing which have been developed over the years have their advantages and disadvantages. The oldest technique, variable reluctance (VR) sensors, are basically coils of wire. A change in magnetic field induces a voltage in the coil by Lenz's law. They are relatively large, sensitive to vibration, and do not sense dc fields. Hall effect sensors have been available since shortly after the second world war. A current driven through a slab of semiconductor material develops a voltage perpendicular to the current when a magnetic field is applied perpendicular to the slab due to the Lorentz force on the electrons. These voltages are very small, about 5 mV/100 Oe. Hall effect sensors can be very small and can be integrated with semiconductor circuitry to produce inexpensive sensors which can measure magnetic fields up to 1000 Oe. Anisotropic magneto-resistive (AMR) sensors utilize the change in resistance in metals when the current is parallel to the magnetic moment compared to the resistance when the current is perpendicular to the magnetic moment. The maximum change in resistance in the nickel-iron alloys typically used is about 2.5% compared to 4 to 16% for current commercial GMR materials. The relatively wide stripes of material necessary in AMR resistors made from NiFe films require much larger chip areas than do the resistors from GMR materials to produce high resistance sensors. High sensitivity sensors can be made but the maximum field that they can detect is only tens of Oe. In addition, high magnetic fields can reverse the polarity of the output of the sensor unless some means such as onboard flipping coils are used to periodically saturate the sensor in a given direction. Because of the relatively large areas occupied by the resistors used in AMR sensors, significant integration is impractical.

The best way to compare the relative merits of these magnetic field sensors is to list relevant properties in a table. Table I compares some properties of GMR, AMR, Hall, and VR sensors.

Table I. Comparison matrix for magnetic field sensors.

	GMR	AMR	Hall	VR
DC operation	Yes	Yes	Yes	No
Size	Small	Large	Small	Very large
Signal size	Large	Medium	Small	Varies with size
Sensitivity	Very high	High	Low	Varies
Relative cost	Low to Medium	High	Low	Medium to High
Stability over Temperature	Good	Good	Fair	Good

INTEGRATED SENSORS

GENERAL CONSIDERATIONS - Smart sensors with both sensing elements and associated electronics such as amplification and signal conditioning on the same die are the latest trend in modern sensors. GMR materials are deposited on wafers with sputtering systems and can, therefore, be directly integrated with semiconductor processes. The small size sensing elements fit well with the other semiconductor structures and are applied after most of the semiconductor fabrication operations are complete. Due to the topography introduced by the many layers of polysilicon, metal, and oxides over the transistors, areas must be reserved with no underlying transistors or connections onto which the GMR materials can be deposited. In actuality the GMR materials are deposited over the entire wafer, but the etched sensor elements remain only on these reserved, smooth areas on the wafers[7].

The typical functions which can be included in an integrated sensor include regulated voltage or current supplies to the sensor elements; threshold detection to provide a switched output when a preset field is reached; amplifiers; logic functions including divide by 2 circuits; and various options for outputs. Using such elements, a two-wire sensor can be designed which has two current levels – a low current level in the absence of a gear tooth and a high current level in the presence of a gear tooth.

ADVANTAGES - There are many advantages of manufacturing sensors with on-board electronics. They usually center on size, cost, and performance. In the automotive industry cost is often the most powerful driver. In semiconductor devices, size and cost are always related. The more chips which fit on a wafer, the lower the final cost. In sensor assemblies a single chip solution is almost always less expensive and more reliable that a multi-chip assembly or a hybrid assembly. A two wire smart sensor is preferable to a multi-wire sensor with remote electronics.

On-sensor electronics can increase signal levels to significant voltages with the least pickup of interference. It is always best to amplify low-level signals close to where they are generated. Converting analog signals to digital (switched) outputs within the sensor is another method of minimizing electronic noise. The use of comparators and digital outputs makes the nonlinearity in the output of sandwich GMR materials of less concern. Even the hysteresis in such materials can be useful, since some hysteresis is usually built into comparators to avoid multiple triggering of the output due to noise.

INTEGRATION PROCESS - GMR materials have been successfully integrated with both BiCMOS semiconductor underlayers and Bipolar semiconductor underlayers. The wafers are processed with all but the final layer of connections made. GMR material is deposited on the surface and patterned followed by a

passivation layer. Windows are cut through the passivation layer to allow contact to both the upper metal layer in the semiconductor wafer and to the GMR resistors. The final layer of metal is deposited and patterned to interconnect the GMR sensor elements and to connect them to the semiconductor underlayer. The final layer of metal also forms the pads to which wires will be bonded during packaging. A final passivation layer is deposited and windows are etched through to the pads.

CUSTOM AND SEMI-CUSTOM CHIPS- The process of designing, laying out, and fabricating a semiconductor chip can take several months and requires a significant investment which may be uneconomical for low volume applications. A custom chip provides the minimum size and cost since it is tailored to the specific requirements of the application. An alternative to custom designs for prototypes and for small volume applications is the semi custom design for which a standard semiconductor underlayer has been designed with a variety of semiconductor structures in place. These wafers can be stocked with all but the final metal layer completed. A prototype sensor can be designed to use combinations of these electronic components to deliver its desired performance. GMR material under a magnetic shield can even be used for any resistors required by the circuit. Only the masks for the last few layers need be designed, and the lead time for manufacturing new semiconductor wafers is reduced. On the down side, the chip may be larger than a custom designed chip, but for low volume applications and prototypes the slightly higher per chip cost is more than made up for by lower up front design costs and lead time. Figure 12 shows the layout of a semi-custom chip. Both the transistor underlayers and the GMR and plated magnetic layers are shown.

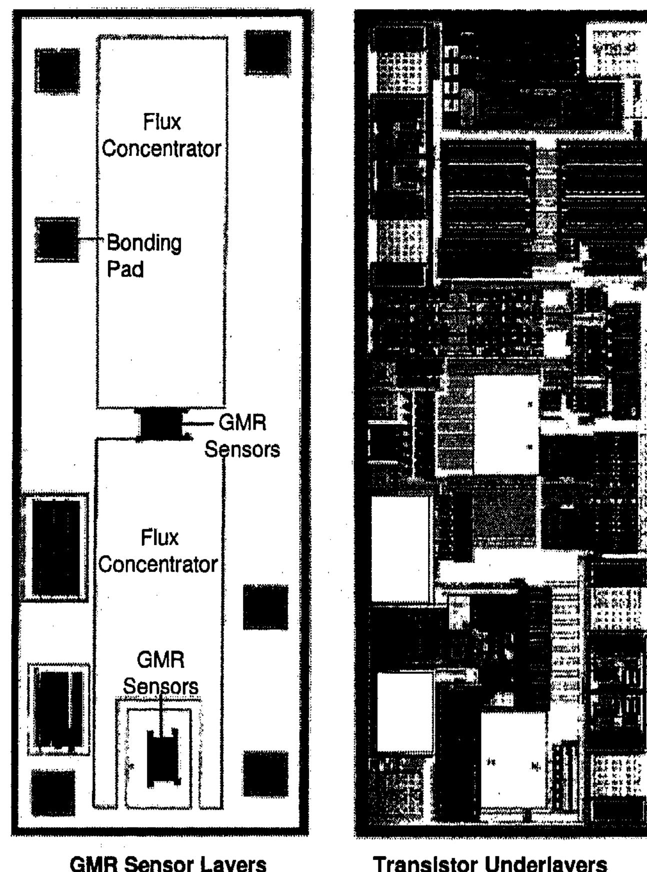


Fig. 12. Layout plots of a semi-custom integrated chip incorporating GMR sensor material. Views of the transistor underlayers without the GMR material and the final chip with GMR and magnetic layers are shown.

CONCLUSION

The discovery of the giant magnetoresistance effect has provided a new and exciting magnetic field sensing material with many applications. The methods of fabricating these multilayer thin films are directly compatible with both BiCMOS and Bipolar semiconductor technologies. Integrated sensors using GMR materials deposited on semiconductor underlayers show great promise for a new generation of sensors with improved cost and reliability for the automotive and other industries.

ACKNOWLEDGMENTS

This review paper has borrowed freely from previous papers and other works written by many of our colleagues at Nonvolatile Electronics, Inc. and especially from the *NVE Engineering and Applications Notes* written by Douglas Guydan We would like to acknowledge our appreciation for their assistance and contributions.

REFERENCES

[1] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Eitenne, G. Creuzet, A. Friederich, and J. Chazelas, "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," *Phys. Rev. Lett.*, vol. 61, 2472-2475 (1988).

[2] J. Brown, "GMR Materials: Theory and Applications," *Sensors Magazine,* vol. 11, 42-48, September 1994.

[3] G. A. Prinz, "Spin Polarized Transport'" *Physics Today,* vol. 48, 58-63, April 1995.

[4] J. Daughton, and Y. Chen, "GMR Materials for Low Field Applications," *IEEE Trans. Magn.,* vol. 29, 2705-2710, (1993).

[5] J. Daughton, J. Brown, R. Beech, A. Pohm, and W. Kude, "Magnetic Field Sensors Using GMR Multilayer," *IEEE Trans. Magn.,* vol. 30, 4608-4610, (1994).

[6] D. Strassberg, "Sensors and coded particles foil counterfeiters," *EDN,* August 15, 1996,

[7] J. L. Brown, "High Sensitivity Magnetic Field Sensor Using GMR Materials with Integrated Electronics," *Proceedings of the Symposium on Circuits and Systems*, Seatle, WA, January 1995.

970605

Development of a Magnetoelastic Torque Transducer for Automotive Transmission Applications

Ivan J. Garshelis
Magnetoelastic Devices, Inc.

Jonas A. Aleksonis and Christopher A. Jones
Methode Electronics, Inc.

Robert M. Rotay
Chrysler Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The development of a transducer for sensing the torque on the output shaft of a four speed rear wheel drive automatic transmission is described. Magnetoelastic polarized ring technology was selected based on its independence from shaft properties and its non-contact mode of sensing. The ring and several intermediate sleeves were attached by press fits onto an experimental shaft. The magnetic field arising from the ring with the application of torque was sensed by flux gate sensing elements. Ability to accurately measure the output torque of an engine driven transmission over its full range of torque and speed was demonstrated by dynamometer tests.

INTRODUCTION

Automotive transmissions serve to match the available torque and speed of the vehicle's engine (or other on-board prime mover) with the tractive effort required at the wheels to attain or maintain the speed and direction of travel desired by the driver. With a *manual* transmission, the driver, on the basis of experience and mastery of the required skills, willfully selects the gear ratio deemed most suitable for the specific driving circumstance (e.g., drive-away, acceleration, hill climbing). With an *automatic* transmission, the driver's desires, relative to the speed of the vehicle, are primarily signaled by manipulation of the throttle. The gear selection lever is repositioned only for directional control and to accommodate special driving conditions, e.g., descending long or steep hills. Based on the engine's response to motion and position of the throttle, as well as to incline and other conditions of the road, the transmission must be automatically guided to quickly select the *most appropriate* gear ratio. The determination of when best to shift from one gear ratio to another and the control of the complex series of events that actually effect the shift require a variety of sensory inputs and a stored knowledge (map) of the engine's speed-torque-throttle position characteristics. The driver's non-involvement with the actual mechanics of shifting results in a heightened awareness of accompanying sounds, vibrations and irregular accelerations. Thus it is the burden of the designer of an automatic transmission to not only ensure that the ratio best able to provide the efficient transfer of power from the engine to the wheels is selected, but that the actual shifting events will take place in a manner that is also smooth and maximally transparent to the driver.

The problem is complicated by variations in the engine map with engine temperature, barometric pressure (altitude), engine wear and between individual engines. It would appear that these problems could be simplified if the actual tractive effort was factored into the control of the transmission since this is the quantity most actively involved in executing the driver's speed demands. In the absence of braking, and except for the rotational inertia of intervening parts, the tractive effort at the wheels has a fixed (for any one vehicle) numerical relationship with the torque on the output shaft of the transmission. Thus, knowledge of the actual value of the instantaneous torque on this shaft should enable the development of improved shift control systems.

The potential for improving shift quality and other aspects of vehicle performance from signals representative of the torque on transmission shafts is well recognized [1-5]. Appreciation of the value of such information has stimulated many investigations into *estimating* this torque from indirect sources, e.g., computations involving engine and torque converter maps [4] or changes in shaft rotational speed [5]. Nevertheless no reports on the development of a *direct* torque sensing means that is sufficiently accurate, reliable and economical for use on production vehicles have yet appeared. This paper reports on the details of such a development.

DEVELOPMENT PROGRAM

A program aimed at developing a torque sensor to a stage sufficient to allow determination of its suitability for the intended use was outlined as follows:

1. Establish a *tentative* specification for the torque sensor. Consider the requirements in sufficient detail to identify preferred sensor locations, expected measurement range, operating environment and performance targets.
2. Critically examine known torque sensing technologies and select the one considered most appropriate for this application. (During this phase it was deemed important to maintain a constant awareness of the goal: a suitably accurate, extremely reliable, economically justifiable and mass producible torque sensor.)
3. Design experimental hardware.
4. Fabricate and assemble necessary parts.
5. Test experimental assemblies under static (non-rotating) loads and under unloaded rotational conditions.
6. Load test an assembled transmission on a dynamometer.

In-vehicle testing was not considered to be part of this development phase since such would clearly require the kind of preparation that would not be undertaken without encouragement from step 6 results. Because of the need to schedule dynamometer testing well in advance, static and unloaded spin testing of the shaft and skeletal sensor constructions were to be relied on to avoid seriously unpleasant surprises during dynamometer testing of the assembled transmission.

SPECIFICATION

The output shaft of a four speed automatic transmission used in front engine, rear wheel drive, light trucks and utility vehicles was selected for installation of a first generation torque sensor. A photograph of the targeted shaft as it normally exists without a sensor is shown in Fig. 1. The sensor was to be located in the region, forward of the speedometer take-off gear, indicated as A A on the photo.

Tentative target specifications were as follows:

- Location: Overdrive unit output shaft of Chrysler Model 44RE/42RE transmission.
- Shaft Diameter: 35 mm at sensor location.
- Shaft Material: Forged AISI/SAE 8620, Carburized, OQ & T (0.040/0.070" case, HRC 58/62) [6].
- Torque Range: 0-2700 N·m.
- Overload: 4700 N·m forward, 2700 N·m reverse.
- Speed Range: 0-5000 rpm (highest accuracy expected to be required in middle ranges).
- Resolution: <7 N·m (1 part in 400).

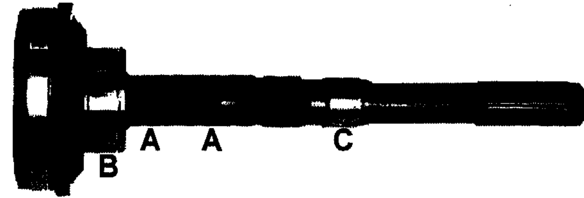


Fig. 1. Transmission output shaft. A A indicates torque sensing region. The shaft is supported in the overdrive housing by ball bearings at B and C.

- Linearity: Desired, monotonicity required.
- Hysteresis: TBD (rezeroing in neutral acceptable).
- Output Signal: 0.2 - 4.8 Vdc corresponding to 0-2700 N·m. CW or CCW directionality to be independently determined.
- Environment: Cast aluminum enclosure. Exposure to transmission fluids. Temperature range: -40 to +150 °C.
- Response: 30 Hz bandwidth (opening estimate).
- Power Supply: 12 Vdc system, as per SAE - J1211.

SELECTION OF TORQUE SENSING TECHNOLOGY

BROAD CLASSIFICATION - A wide variety of sensing technologies are embodied in the many available instruments for measuring the torque being transmitted by a rotating shaft. By one means or another these *torque transducers* all serve to provide an electrical signal that is a linear analog of the torque. All operate by sensing some manifestation of the shear stress which effectively transmits the torque along the shaft. While these devices are more commonly categorized according to their underlying operating principles, for the purposes here we find it more suitable to divide all torque transducers into only two, constructionally distinct categories, namely: those that utilize shaft mounted electrical components and those that do not. In consideration of the rigors of the present application, including circumstances that may arise during manufacture, assembly and repair of the transmission as well as the conditions of actual use, technologies utilizing shaft mounted electrical components were felt to be insufficiently robust. It was also felt that designs utilizing shaft mounted electrical components would be too labor intensive to be cost competitive. In fact, constructions relying on adhesively bonding any components to the shaft were deemed inappropriate. Thus, although strain gauges have been suggested for other automotive applications [7] and capacitive elements have been investigated for use on automotive axles [8], these technologies were not considered further. For similar reasons, as well as the absence of reports on their use in automotive applications, SAW devices [9] were also excluded from more detailed examination.

TORSION BAR - Technologies based on measurement of shaft twist angle were briefly considered. The angle of twist, θ, for a 35 mm diameter (*d*), SAE 8620 steel shaft (*G* = shear modulus = 80 GPa) under the full range torque, *T* = 2700 N·m, was found from

$$\theta = \frac{32LT}{\pi d^4 G} \tag{1}$$

to be only 0.0131° per mm of shaft length. Even if a clear gauging region 100 mm in length (*L*) could be made available, the angle measurement system would need to discriminate displacements as fine as 0.00328° in order to provide the desired 1 part in 400 resolution. Such resolution challenges even instrument grades of "torsion bar" torque transducers, e.g., those based on the light passing through axially spaced optical shutters [10]. It was judged as unlikely that variable reluctance systems, such as those developed for automotive power steering applications (wherein typical twist angles are between 4° and 7° [11]), could be sufficiently refined for use with shafts having the required torsional stiffness. While this difficult angular resolution requirement could be relaxed by reducing the shaft diameter in the gauging region, this would risk plastic yielding of the shaft under peak torque conditions. Also, increasing shaft compliance would alter the spectrum of torsional oscillations in the drive train, with possibly serious adverse consequences. For these and other reasons associated with the manufacturability of needed high precision components, this approach was not considered further.

MAGNETOELASTIC (PERMEABILITY) - Detailed attention was directed to magnetoelastic (often called, "magnetostrictive") torque sensing technologies. It is well known that stress acts to alter the orientation of the *spontaneous magnetization,* the presence of which characterizes a material as ferromagnetic [12]. (Magnetomechanical effects due to stress manifest an intrinsic magnetoelastic interaction generally present in this class of materials.) Among the consequences of this stress induced magnetization reorientation are changes in the permeability of a stressed sample from measured values when in the unstressed condition. The utility of this phenomenon for the measurement of torque on a rotating shaft can be readily appreciated since magnetic coupling enables the determination of properties such as permeability without any need to physically contact the stressed body. The further attractiveness of utilizing changes in the magnetic properties *of the shaft itself* in order to determine the torque being transmitted is obvious. Fleming [13] has studied and compared three common configurations of torque transducers that develop their output signals from stress induced permeability changes.

Although well understood and seeming to be ideally appropriate, reported efforts to use these methods on powertrain shafts nevertheless fail to demonstrate suitable accuracy over the required torque and temperature ranges. The underlying problem stems from the fact that powertrain shafts are designed to fulfill demanding mechanical functions.

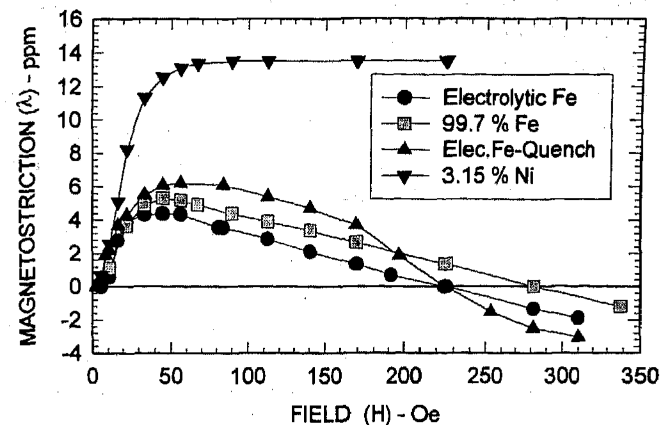


Fig. 2. Effects of composition and thermal history on the magnetostriction of iron and dilute alloys.

Except for those instances when torque sensing is the overriding concern, e.g., in commercial torque transducers, the choices of shaft size, material, fabrication methods, heat treatments and other processing steps are all aimed at meeting mechanical performance and cost criteria. The influences of composition and processing conditions on magnetic and magnetoelastic properties of the shaft are only rarely considered [2, 14]. One generally ignored property, *magnetostriction* (λ), plays an essential role in determining both the direction and amplitude of changes in permeability with stress [15]. While *saturation* magnetostriction (λ_s) [16] is an intrinsic material property, the variation of permeability with stress follows from the variation of magnetostrictive strain ($\delta l/l = \lambda$) with applied field (H), and this is a sensitive function of microstructure as well as composition. Data plotted in Fig. 2 (taken from Schulze [17]) for some dilute alloys of iron clearly shows these dependencies. The reversal in sign of λ seen in all but the 3.15% Ni-Fe alloy is reflected in a variable *direction* of change in permeability with stress in these materials (the *Villari* reversal [18]). Shaft materials exhibiting a Villari reversal are clearly not good choices for this means of torque sensing. SAE 8620 [6] has too little Ni, Si, Al or Co to squelch the non-monotonic λ vs H characteristic found in iron, its principal constituent.

The effective permeability of a shaft surface, even in the absence of stress, varies significantly with temperature as well as with composition and microstructure. Moreover, the term, "permeability" does not define a single, material dependent property, since its value, and the manner and range of its variability with stress, depends on the amplitude and frequency of the measuring field (as well as on other, more incidental influences on shaft magnetization, e.g., the earth's field). Thus, such performance characterizing features as linearity, hysteresis, dynamic range and temperature effects are all dependent on the excitation field parameters in this class of torque sensors.

Additional complications pertinent to this specific application include the potential for drifting calibration caused by *irreversible* changes in permeability associated with stress excursions beyond the endurance limit of the shaft material during peak overloads. Similar drifts would accompany even

subtle microstructural changes associated with prolonged time at temperatures in the tempering range. These, in addition to concerns with normal shaft to shaft variations in characteristics affecting calibration, were considered serious enough to eliminate "permeability" forms of magnetoelastic torque transducers from further consideration. (The existence of the problems associated with the inappropriate and inconsistent magnetoelastic properties of typical shaft materials are well recognized [1, 2, 3, 7, 14, 19, 20, 21]. Persistent attempts at solving these problems, by e.g., characterizing individual shafts [19], ad hoc material development [20], attachment of sleeves of more suitable materials [21] or even remote variations on this basic sensing theme such as inducing eddy currents in Cu conductors electroplated on the shaft [3], have yet to demonstrate adequate performance for this demanding application.)

MAGNETOELASTIC (POLARIZED RING) - By avoiding dependence on permeability, inconsistent or unstable magnetic properties of the shaft do not affect the operation of magnetoelastic torque transducers employing magnetically polarized ring constructions [22]. Figure 3 shows the arrangement of functional elements comprising this type of transducer. The transducer consists only of a ring of magnetoelastically active material attached to the shaft in such manner as to assure that they act as a single mechanical unit and a proximate magnetic field sensor. (Mechanical attachment of the ring to the shaft is most often accomplished by a simple press fit.) As indicated in the figure, the ring is magnetically divided into two regions having oppositely directed circumferential polarizations. When torque is applied to the shaft these polarizations become more or less steeply helical; the ring effectively develops magnetic "poles" uniformly distributed around its central band (indicated in gray) and, of opposite polarity, at its ends. The strength and polarity of the magnetic field associated with these poles is linearly analogous to the magnitude and CW or CCW

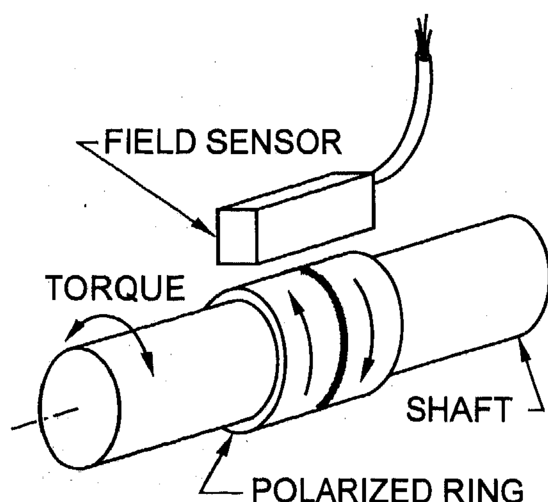


Fig. 3. Functional elements of a polarized ring torque transducer. The arcuate arrows indicate the directions of remanent magnetization (i.e., polarization) in each ring region. When torque is applied, magnetic poles develop along the gray band and at each of the ring ends.

directionality of the applied torque. The output signals from one or more magnetic field sensors thereby provide the measure of the torque. (Further explanation of the operating principles and related details can be found in the referenced technical and patent literature [22-25].) The sensitivity, i.e., the field intensity per unit of torque in this type of transducer is based primarily on the size and *intrinsic* material properties of the polarized ring. This, in combination with other features, including the complete absence of any form of excitation field, greatly simplifies (or may even eliminate) individual transducer calibration. A properly constructed transducer of this type will, by its nature, generate no field (i.e., zero) with zero applied torque, independent of its temperature. Hence, temperature compensation, if needed at all, would be limited to transducer sensitivity (i.e., the "span") and would expectedly be generic for any one design. Calibration drift due to thermally induced microstructural changes in the ring is unlikely since typical rings undergo final heat treatments at 483 °C (for 3 hours), far above the specified maximum operating temperature. The simplicity of the construction portends a reliable and economical device. All of these clearly desirable features, together with advances in the technology development and experience in its use since first suggested for electric power steering applications [25], directed its selection for this application

TRANSDUCER DESIGN AND CONSTRUCTION

GENERAL - Although the functional elements shown in Figure 3 comprise a complete torque transducer, two factors associated with the existing transmission shaft (Fig. 1) necessitated the use of additional components, at least at this experimental stage of development. Firstly, as with all applications involving shafts made of ferromagnetic steels, it is found desirable to interpose a non-magnetic spacing ring between the magnetically polarized ring and the shaft. Spatial separation of the field generating ring from the shaft allows more of the magnetic flux to follow paths through the space on the outside of the ring where the field sensor(s) are located. If the ring is mounted directly on a typical ferromagnetic shaft, nearly 90% of the generated flux will take lower reluctance paths through the shaft and thus not be available for sensing. Besides reducing this loss, spatial separation minimizes any effects that incidental shaft magnetization may have on the operation of the transducer. It has been found in some applications that an austenitic stainless steel "insulating" ring having a radial thickness as small as 1 mm reduces the flux loss to less than 25%. Greater thickness will naturally be even more efficient.

The size (diameter in the sensing region) of the presently available transmission shaft made it necessary to add still another component (ring) to the basic construction. Since the polarized ring generates a magnetic field in proportion to the torsional (shear) stresses over its cross section, it is important to ensure that these stresses accurately reflect the *applied* torque and *only* the applied torque. Should there be any plastic yielding within the shaft or ring, or any slip at their interface under any conditions of applied torque, the stresses will not uniformly relax to zero when the torque is removed.

Under such inelastic conditions there will be a *residual* stress distribution over the cross section with the ring stressed in the opposite direction to that which existed when the high torque was applied. The magnetic field arising as a result of this residual stress will be sensed as a torque even though the applied torque is zero. The field sensor output signal at zero applied torque will thus be different (i. e., there will be *hysteresis*) depending on whether inelastic events occurred during CW or CCW applied torques. Hysteresis will also be manifested as differing values of output signal for the same value of torque, depending on whether the torque is increasing to, or decreasing from, a torque high enough to cause inelastic events. While zero shifts at zero torque can be electronically eliminated (by resetting the signal when in neutral) and minor hysteresis is of little practical concern, it is clear that yielding of the shaft and interfacial slip would best be avoided.

MECHANICAL DETAILS - The extent of this problem and its solution are readily understood from the following analysis. The shear stress associated with the transmission of a torque, *T*, through a shaft of diameter, *d*, varies from zero at the center to a maximum value (τ_m) at the surface, found from

$$\tau_m = \frac{16T}{\pi d^3}. \tag{2}$$

For $d = 35$ mm, under the peak overload torque of 4700 N·m, τ_m will reach 558 MPa. While no torsional yield strength data for carburized 8620 steel has been found, considering that the tensile yield strength (0.2% offset) for the core of a bar of this size is almost certainly ≤ 800 MPa [6], and that the torsional yield strength of steels is typically about 57% of the tensile value (i.e., $\tau_y \leq 456$ MPa), yielding of some portion of the cross-section under overload conditions seems likely. An obvious solution is to enlarge *d* in the region where the ring is to be installed and such would expectedly be the eventual course of action. Presently limited by the existing shaft dimensions, it would appear that *d* could be effectively increased by simply increasing the wall thickness of the spacing ring. After a moment's reflection, it will be realized that the shear stress in each ring is transferred to it via frictional forces arising from the contact pressure at all underlying interfaces. Without unlimited, independent control of friction coefficients, high interfacial contact pressures are required. Except for some exotic (read: expensive) Ni based alloys, non-magnetic materials have insufficient strengths to act elastically under the required contact pressures. Hence it was considered necessary to build up the effective diameter of the shaft in the sensing region by pressing on a heat treated steel sleeve. The problem was further complicated by the fact that the shaft in region AA (Fig. 1) was not sufficiently larger than the bearing journal at C to obtain the desired interference. The shaft was thus built up to the desired diameter by means of *two* hardened steel sleeves, sized such that mounting,, by pressing on the outer one, increased the contact pressure between the inner one and the shaft. To avoid distortion under the heavy press-on forces required, and to provide better control of the alignment of the parts during their assembly, each interface was machined to a shallow taper (typical

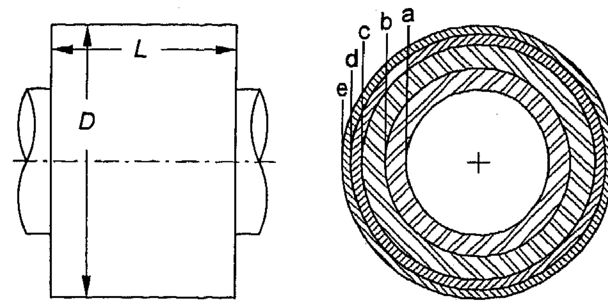


Fig. 4. Details of experimental ring assembly. Inner two rings - OQ & T AISI 4340; "Insulator" ring - Carpenter 22 Cr 13 Ni 5 Mn; Outer ring - Teledyne Vasco C-250, aged 3 h @ 483 °C. Axial length L = 45.7 mm. Outside diameter = 66.0 mm. Nominal radii at each indicated interface: a = 17.6; b = 22.9; c = 28.7; d = 31.2; e = 33.02.

included angle ≈ 1°). Figure 4 shows the details of the final assembly design.

It is of course, only the outermost ring of the four ring set that has an active role in the torque *sensing* function. The use of an interference fit for this ring served the dual purposes of assuring the mechanical integrity of its attachment to the underlying "shaft" and instilling a circumferential tensile (hoop) stress. This is needed to establish the circumferential direction as the stable orientation for the remanent magnetization. An 18% Ni maraging steel was chosen for this ring on the basis of its high strength and desirable magnetic properties [26]. Although it was not considered necessary for the magnetically polarized ring to be any longer than 20-25 mm in order to fulfill its transducing function, it was thought best at this stage to keep all rings the same length in order to reduce the possibility of slippage at any of the interfaces.

MAGNETIZATION - The two, oppositely polarized regions in the outer ring were instilled by simply holding a North-South/South-North pair of bonded ferrite magnets close to the ring surface while rotating the shaft (for about 100 revolutions) [22]. The polarizations were circumferentially homogenized by passing an approximately 300 peak ampere, 60 Hz current, axially through the shaft.

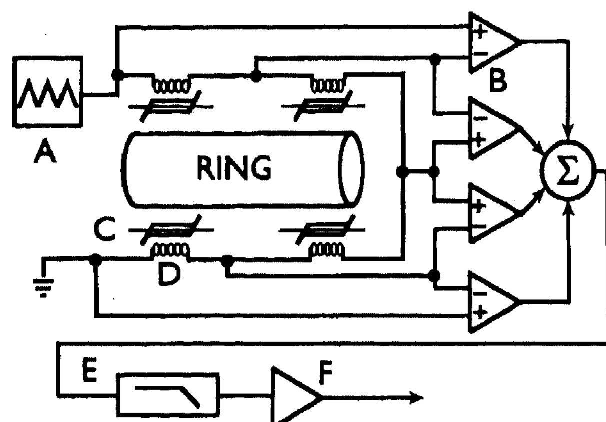
FIELD SENSOR - In general, the magnetic field that arises when torque is applied to the shaft can be sensed by any of several suitable means. Both Hall effect and "flux gate" sensors [27] were used for static load and unloaded spin testing of the experimental transducer shaft. Since the output signals from presently available Hall effect devices often drift with temperature, a flux gate design, having an inherently stable and temperature independent zero field output signal, was chosen for installation into the transmission housing. At this stage in the development it was thought prudent to avoid possible problems from shaft runout, bending stresses or other sources of circumferential inhomogeneity by combining the signals from two, diametrically opposed, sensor pairs. The actual sensors consisted of small diameter, solenoidal windings of high temperature insulated magnet wire, each encircling a single strand of near zero λ_s amorphous wire

(Unitika DC2T). These core/coil "assemblies" were cemented into grooves machined into an aluminum ring that had been turned to fit an axially concentric bore already existing in the overdrive housing. The radial air gap between the ring and sensor surfaces was 0.5 mm.

ELECTRONICS - Figure 5 shows a block diagram of the circuit used with the core/coil assemblies to obtain a field (hence torque) dependent signal. Details of its operation are described in the (lengthy) caption.

TESTING

STATIC TESTS - An apparatus for applying controlled and accurately measurable torques to the stationary shaft was constructed. The shaft was firmly clamped at the forward (large) end, while, through a system of links and levers and a yoke having a matching internal spline, torque in the form of a force couple was applied to the free end. The push and pull forces comprising this couple were applied through tension/compression strain gauge load cells. The forces were



A - Triangle Wave Oscillator
B - Comparators
C - Square loop wire cores
D - Coils
E - Low Pass Filter
F - Output Buffer
Σ - Summing Amplifier

Fig. 5. Block diagram of electronic circuit used with the flux gate field sensors. The triangle wave current (≈ 13 kHz) through the coils sweeps the core wires through major hysteresis loops. In the absence of torque, the magnetization within each of the cores reverses at the same amplitude of coil current. When torque is applied, the associated magnetic field adds to the excitation field on the cores adjacent to one region of the ring while subtracting on those adjacent to the other. Since all core magnetizations reverse at one characteristic value of *effective* field, the cores adjacent to each region now reverse at different current amplitudes, i.e., at different times in the triangle wave. The comparators, which in the absence of torque spend equal (half) periods with high and low outputs, now have asymmetrical high and low periods. Their average outputs, after summing, filtering and buffering, thereby vary up or down depending on the field amplitude and polarity, hence on the torque amplitude and direction.

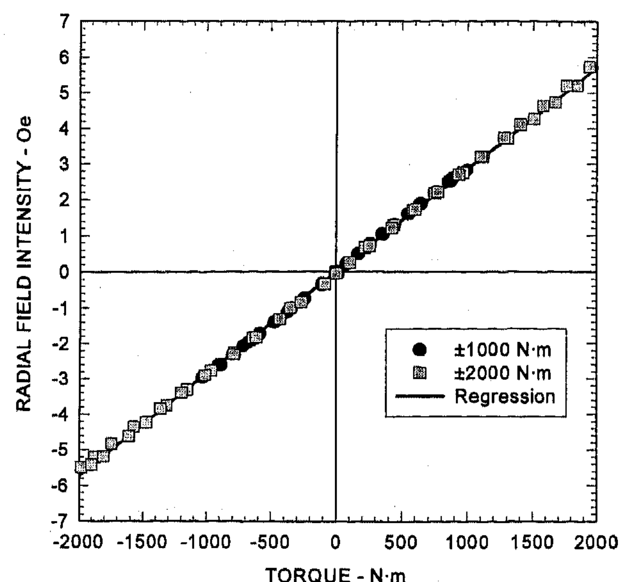
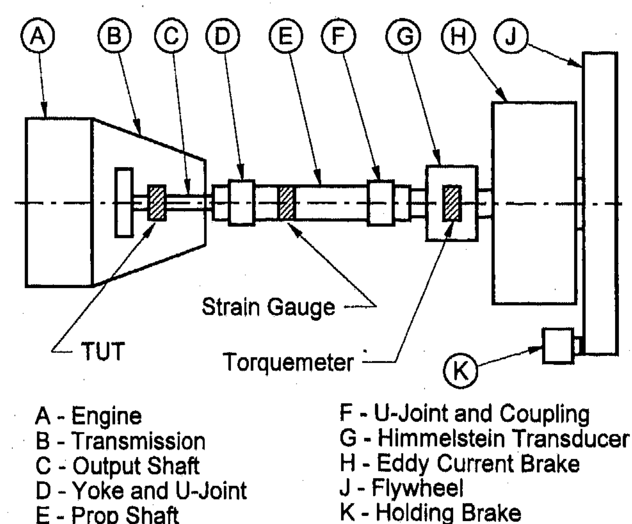


Fig. 6. Variation in the radial field intensity appearing near the axial center of the ring with applied torque. (Distortion of the yoke and load cell limitations prevented accurate exploration at much higher torques.)

produced by a hydraulic "jack" which, by its placement at one or the other end of a lever, provided for the application of CW or CCW torques. A commercial gaussmeter with Hall effect probes was used to measure the actual radial and longitudinal field intensities in the space around the ring. Typical data obtained for complete cycles of increasing and decreasing torques in each direction are plotted in Fig. 6. The excellent linearity and minimal hysteresis are clearly indicated.

DYNAMIC TESTS - After assembly of the overdrive components to the shaft and further assembly into the housing, the overdrive unit was attached to a Chrysler 44RE transmission. The transmission, coupled to a 4.0 liter 6 cylinder in-line engine, was installed in a fully instrumented



A - Engine
B - Transmission
C - Output Shaft
D - Yoke and U-Joint
E - Prop Shaft
F - U-Joint and Coupling
G - Himmelstein Transducer
H - Eddy Current Brake
J - Flywheel
K - Holding Brake

Fig. 7. Arrangement of apparatus used for dynamic tests. The shaded areas indicate the three torque measuring devices.

dynamometer test cell. The driveline arrangement is schematically illustrated in Fig. 7. In addition to the measurement of the output signal from the transducer under test (TUT), torque was measured by a custom built and calibrated strain gauge bridge installed on the prop shaft and by a Himmelstein Model MCRT 8-02TA Torquemeter (rated for 50,000 lb.·in. up to 5000 RPM). Other measured parameters included: engine speed, throttle position, output shaft speed and transmission oil temperature.

All measurement data was sampled at rates appropriate to the specific test and acquired by computer. All details of the test regimens were under computer control except for the selection of the transmission gear ratio. This was either allowed to change automatically by the normal operation of the transmission or else continuous operation in first gear was manually selected. Testing followed several different regimens. For example, while the transmission was kept in first gear, targeted values of speed and output torque were established by interactive adjustments of the throttle position and dynamometer brake current. While some data was collected in this way at various output shaft speeds and torques, one limitation or another prevented the attainment of steady state operation at high enough torques to fully test the transducer. Automatic test cycles provided more interesting data. Prior to the beginning of each such cycle, the output shaft was held stationary by the holding brake on the flywheel while the engine idled. The test was launched by simultaneously releasing the brake and snapping the throttle to a preestablished opening. The engine would accelerate more or less rapidly depending on throttle position and the transmission stepped through its gear ratio changes as the flywheel reached appropriate speeds. After reaching the limiting output speed, the throttle was cut and the flywheel was brought to a halt by the eddy current brake. Tests of this nature were performed for approximately 10% increments in throttle position. Data (at 100 samples/s) from one test run at full throttle (nominally 100%) are shown in Figures 8-12.

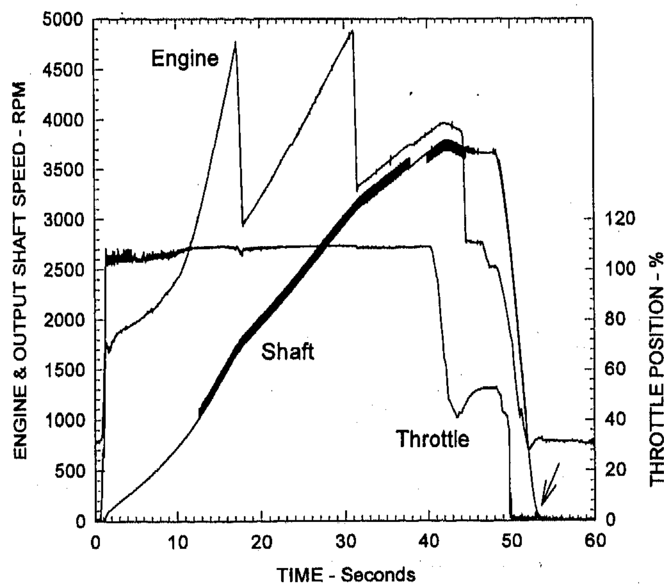


Fig. 8. Variations in engine and output shaft speeds (and throttle position) during 100% throttle automatic test cycle.

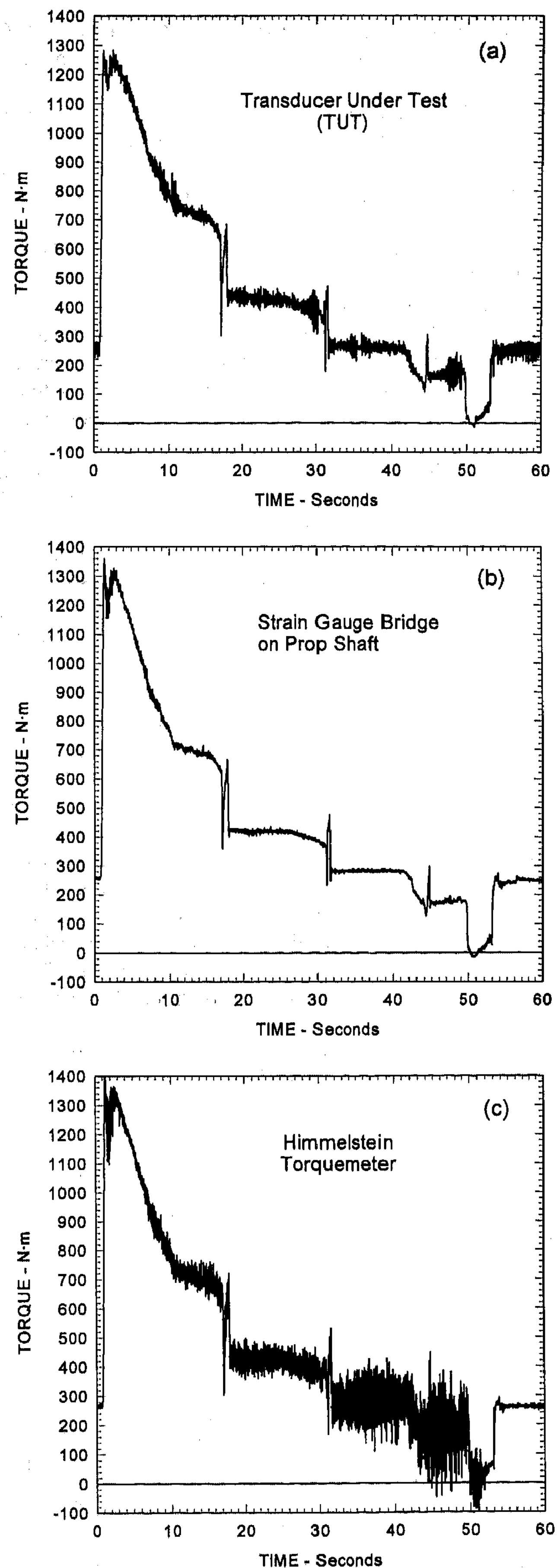


Fig. 9. Torque as measured by: (a) the TUT; (b) the strai gauges on the prop shaft; (c) the torquemeter.

The progression of active events throughout the 60 second duration of the automatic test cycle is shown in Fig. 8. Following the opening of the throttle, which effectively starts the test, the output shaft is seen to undergo a continuous acceleration as the transmission steps between first and second and second and third gears, while decelerating slightly prior to the shift into overdrive. The throttle was cut to about half open just before the shift into overdrive, near 45 seconds into the test, after which the eddy current brake was applied and the throttle shut (near 49 seconds) bringing the shaft to a halt slightly after 54 seconds.

Figure 9 shows the variation in the output torque as indicated by each of the three measurement devices shown in Fig. 7. This figure shows that the measurement of torque under the dynamic conditions found in engine/transmission drivelines is not a simple matter. The precise answer to the question, "What is the magnitude of the torque?" is seen to depend on the type and location of the sensor. While this complicates the performance assessment of any torque measuring device under study, it also shows that there is a wealth of information that torque measurement can provide.

The "noise" seen to be present in varying degrees in these three figures is not noise in the classical sense but represents actual torque variations experienced by each sensor; different in each due to torsional oscillations. It is quite normal for the dynamic torques on a distributed system containing elements of varying compliances and inertial masses to vary from location to location. This is readily seen by examining the expanded data shown in Figs. 10-12. While the TUT is seen to provide an average signal that for the most part accurately follows the prop shaft strain gauge data, a superposed oscillatory signal of $\approx$ 9 Hz is seen to be pervasive. This signal component is believed to have arisen from oscillations of the engine on its mounts, a conclusion supported by its presence even when the output shaft was stationary, e.g., as seen in Figs. 9 and 10 during the first one

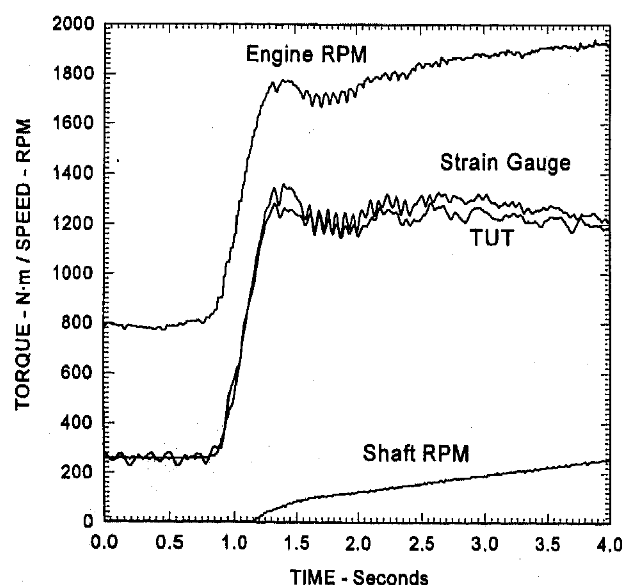


Fig. 10. Variations in engine RPM, shaft RPM and torque, as indicated by both the strain gauges on the prop shaft and the TUT on the transmission output shaft, during the first 4.0 seconds of the automatic test at full throttle.

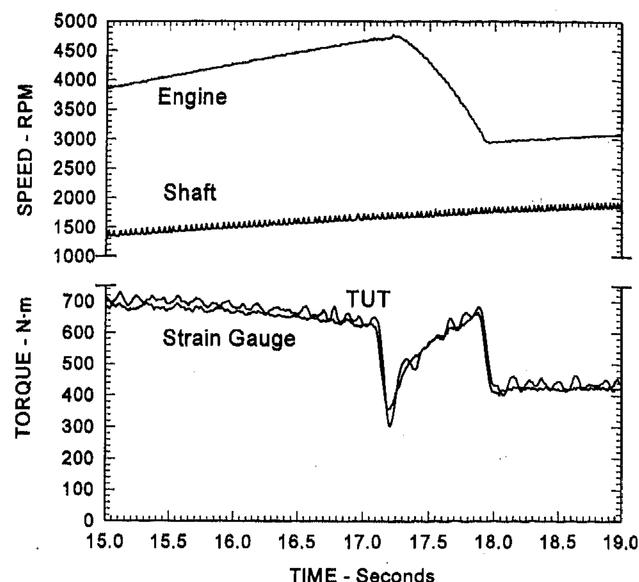


Fig. 11. Variations in engine RPM, shaft RPM and torque, as indicated by both the strain gauges on prop shaft and the TUT, during shift from first to second under full throttle conditions.

second of the test and in Fig. 9 (and the right end of Fig. 12) during the last 6 seconds of the test cycle. These oscillations are seen to be stimulated to substantial amplitudes, both in particular engine speed ranges and during rapid changes in engine acceleration accompanying shifting events and throttle changes. Chaotic and severe bouncing of the engine and transmission was observed during the tests but no records were made to correlate the timing and extent of such events.

Torsional oscillations in the torquemeter records are seen in Fig. 9(c) to be even more severe, but these were of a higher frequency ($\approx$ 13.5 Hz), reflecting a resonant association of the torquemeter compliance with the flywheel and dynamometer

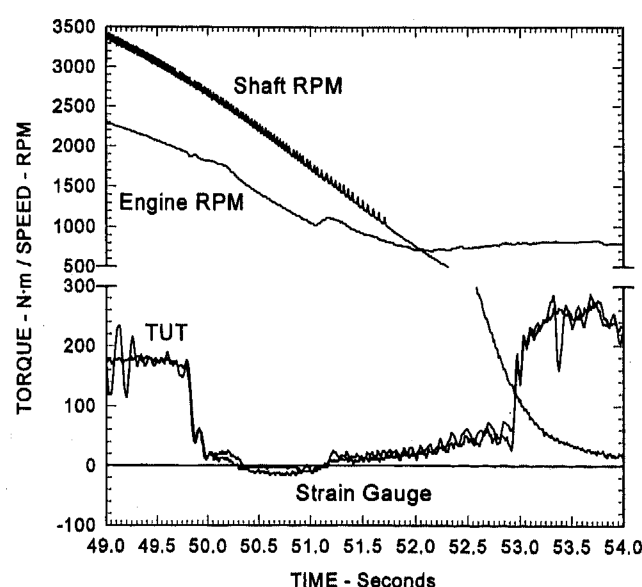


Fig. 12. Variations in engine RPM, shaft RPM and torque, as indicated by both the strain gauges on the prop shaft and the TUT, during final throttle shut down. As seen by the arrow in Fig. 8, the throttle "bounced" to almost 5 % near 53 seconds.

rotor inertias. These oscillations are seen to have been absent (as would be expected) when the output shaft was stationary at the beginning and end periods of the test cycle. The prop shaft is seen to be largely free of the large oscillatory torques found at both ends of the driveline. This too is expected since this shaft is insulated by the compliance of the U-joints and the non-linearity of the torque and velocity transfers associated with the angularity of the coupled shafts.

Other than for the additional oscillatory torque, the TUT signal is seen to track the prop shaft data quite accurately at all torque levels except those over about 1200 N·m. The torque indicated by the TUT is clearly seen, by the expanded data in Fig. 10, to be consistently low from 1.2 - 4 seconds into the test. Since no such fall-off in sensitivity is seen in the static data plotted in Fig. 6, it would appear that a saturation problem may exist in the transfer function of the flux gate sensor. While this certainly needs to be investigated, it is also interesting to note that, as indicated in Fig. 9, the peak torque indicated by the torquemeter *exceeds* that shown by the prop shaft by a similar amount in the same region.

The expanded data in Figs. 10-12 provide an opportunity to examine both the ability of the TUT to follow transient torques and to observe the large amplitude that the oscillatory torques can reach when stimulated, as well as the rapidity with which these are damped.

During the many tests that were performed, over a period of several hours, the transmission oil temperature varied only from RT (≈ 20 °C) to 76 °C, only a small fraction of the specified operating range. Nevertheless, within this limited range no significant changes in the performance of the TUT were observed.

CONCLUSION

It is recognized that the developments reported here are but the first steps towards the ultimate application of torque transducers in automotive transmissions. Even at this early stage of development the described transducer clearly shows the ability to accurately measure the actual transmitted torque using simple and reliable hardware. As the technology advances, it can be reasonably expected that performance will be even further improved. A possible additional benefit from the use of this polarized ring technology is the reduction in overall sensor count, since adaptations of the basic technology have already been shown to provide rotational speed information in the transducer signal [28].

Next steps in the development will include further dynamometer tests utilizing a larger engine and more sophisticated loading means. This will allow dynamic testing up to the full specified CW and CCW torques, including excursions into the overload range. Following this, there will undoubtedly be extensive in-vehicle testing to provide the data necessary for the development of suitable control algorithms. Only then can it be determined whether actual torque measurement on automotive transmission shafts can provide the anticipated benefits.

ACKNOWLEDGMENTS

We would like to express our thanks to Hong Liu for her invaluable assistance with the FEA design and stress analysis of the ring system and to David Cripe for his design and construction of the field sensors and associated electronics. We would also like to thank the managements of our respective organizations for their support and encouragement.

REFERENCES

[1] William J. Fleming, "Magnetostrictive Torque Sensors-Derivation of Transducer Model", SAE Paper No. 890482, (1989).

[2] Hiroyuki Aoki, Junichi Maruyama, Munekatsu Shimada, Katsuji Tanizaki, Shinichiro Yahagi, and Takanobu Saito, "Torque Sensor with Shape Anisotropy", Tech. Digest of the 7th Sensor Symposium, 83-86, (1988).

[3] Jarl R. Sobel, Jan Jeremiasson, and Christer Wallin, "Instantaneous Crankshaft Torque Measurement in Cars", SAE Paper No. 960040, (1996).

[4] Masahiko Ibamoto, Hiroshi Kuroiwa, Toshimichi Minowa, Kazuhiko Sato and Takeshi Tsuchiya, "Development of Smooth Shift Control System with Output Torque Estimation", SAE Paper No. 950900, (1995).

[5] Toshimichi Minowa, Kenichirou Kurata, Hiroshi Kuroiwa, Masahiko Ibamoto, and Masami Shida, "Smooth Torque Control System Using Differential Value of Shaft Speed", SAE Paper No. 960431, (1996).

[6] AISI/SAE 8620, Composition: C- 0.18/0.23; Mn- 0.70/0.90; P- 0.035 max; S- 0.040 max; Si- 0.15/0.30; Ni- 0.40/0.70; Cr- 0.40/0.60; Mo- 0.15/0.25; Fe- balance. For recommended heat treatments and mechanical properties, see, for example, *Engineering Properties of Steel*, Philip D. Harvey, Editor, American Society for Metals, pp 189-193, (1982).

[7] Erich Zabler, Frieder Heintz, Anton Dukart, and Peter Krott, "A Non-Contact Strain Gage Torque Sensor for Automotive Servo Driven Steering Systems", SAE Paper No. 940629, (1994).

[8] J. D. Turner, "The development of a thick-film non-contact shaft torque sensor for automotive applications", J. Phys. E: Sci. Instrum., 82-88, (1989).

[9] anon., Machine Design, 48, Description of a torque sensor based on the stress dependence of surface acoustic waves (SAW) developed by Sensor Technology Ltd., Banbury, UK (Jan. 11, 1996).

[10] Series I and Series II Torque Transducers, manufactured by Vibrac Corporation of Amherst, NH, and described in company literature, are claimed to obtain 1% accuracy with 0.5° full scale deflection using optical shutter techniques.

[11] See for example: U.S. Patent No. 5,394,760 (1995), E. K. Persson, P. K. Webber, and D. L. Perry, "Torque Sensor for a Power Assist Steering System", Assigned to TRW Inc.

[12] Richard M. Bozorth, *Ferromagnetism*, D. Van Nostrand, Princeton, NJ (1951), p 12.

[13] William F. Fleming, "Magnetostrictive Torque Sensors - Comparison of Branch, Cross, and Solenoidal Designs", SAE Paper No. 900264, (1990).

[14] Munekatsu Shimada, "Magnetostrictive torque sensor and its output characteristics", J. Appl. Phys., 73, (10) 6872-6874, (1993).

[15] R. M. Bozorth, op. cit. (Ref. 12), pp 595-627.

[16] R. M. Bozorth, op. cit. (Ref. 12), p 632.

[17] Alfred Schulze, "Magnetostriction I.", Z. Physik, 50, 448-505, (1928).

[18] R. M. Bozorth, op. cit. (Ref. 12), 602.

[19] U.S. Pat. No. 5,495,774 (1996), R. D. Klauber, E. B. Vigmostad and F. P. Sprague, "Magnetostrictive Torque Sensor Air Gap Compensator", Assigned to Sensortech L. P.

[20] U.S. Pat. No. 5,107,711 (1992), H. Aoki, S. Yahagi and T. Saito, "Torque Sensor", Assigned to Nissan Motor Co. Ltd.

[21] U. S. Pat. No. 4,506,554 (1985), "Magnetoelastic Torque Transducer", K. Blomkvist and J. O. Nordvall, Assigned to ASEA Aktiebolag.

[22] I. J. Garshelis and C. R. Conto, "A torque transducer utilizing a ring divided into two oppositely polarized regions", J. Appl. Phys. 79 (8), 4756-4758, (1996).

[23] I. J. Garshelis, "A Torque Transducer Utilizing a Circularly Polarized Ring", IEEE Trans. Magn., 28, 2202-2204, (1992); I. J. Garshelis, "Investigations of Parameters Affecting the Performance of Poarized Ring Torque Transducers", IEEE Trans. Magn., 29, 3201-3203, (1993).

[24] U. S. Pat. Nos. 5,535,555 (1994); 5,465,627 (1995) and 5,520,059 (1996), I. J. Garshelis, "Circularly Magnetized Non-Contact Torque Sensor and Method for Measuring Torque Using Same", Assigned to Magnetoelastic Devices, Inc.

[25] Ivan J. Garshelis, Kristen Whitney and Lutz May, "Development of a Non-Contact Torque Transducer for Electric Power Steering Systems", SAE Paper No. 920707, (1992).

[26] Ivan J. Garshelis, "Magnetic and Magnetoelastic Properties of 18% Ni Maraging Steels", IEEE Trans. Magn, Vol. 26, 1981-1983, (1990).

[27] Daniel I. Gordon and Robert E. Brown, "Recent Advances in Fluxgate Magnetometry" IEEE Trans. Magn., MAG-8, 76-82, (1972).

[28] Ivan J. Garshelis, Christopher R. Conto and Wade S. Fiegel, "A Single Transducer for Non-Contact Measurement of the Power, Torque and Speed of a Rotating Shaft", SAE Paper No. 950536, (1995).

970644

Adaptive Light Pattern - A New Way to Improve Light Quality

Henning Hogrefe and Rainer Neumann
Robert Bosch Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

In most cases of night time driving the low beam light function is used for the road illumination in front of the car. This conventional low beam function has a constant light pattern: no matter whether the driver is actually driving on a straight or curvy road, whether the road is wet or dry, always the same beam pattern is applied. Test experiences of the last years prove that a headlamp with an „adaptive light pattern“ having different optimum light pattern adapted to different driving situations should give a better illumination than the conventional constant low beam pattern which can only be a compromise.

In this paper an outline of the basic idea of the „adaptive light pattern“ is given. Different driving situations and their corresponding optimum light pattern, basic technical concepts and the legal situation regarding homologation of such an advanced lighting system will be discussed.

INTRODUCTION

Automotive headlamps provide for each of the basic light functions (low beam, high beam and fog beam) a fixed static light pattern. Some thirty years ago 200-300 lumen were a good total luminous flux for a low beam headlamp. This amount of light was at best enough to illuminate the most important parts of the driver's night time viewing zone.

All lighting regulations and techniques were set up to ensure a minimum light quality using available light sources.

The introduction of halogen bulbs around 1965 greatly improved the light quality. In the eighties two-reflector-headlamps for separate low and high beam and advanced reflector technologies further increased the amount of light available for the street illumination.

Now, a low beam light volume of 500 lumen and more can be achieved using halogen bulbs in combination with advanced reflector techniques or even 1000 lumen using gas discharge lamps (HID). Having reached this state of lighting performance the question arises, in which direction further development should be continued. Shall we add, as in the past, more and more light - if technically available - into the static light pattern giving more and more intense illumination in the hot spot or broader side illumination? Recent HID-light pattern applications have shown very good results on this way (Ref. 1, 2).

Another very promising alternative is to direct the available light dynamically into the direction where it is actually needed. Why illuminating the hot spot with full power while driving with low speed on a windy road? Why wasting many lumen on large lateral side illumination while driving with high speed on a straight highway? Hence, the idea is to have a dynamic or „Adaptive Light Pattern“ (ALP) which always supplies an adapted, optimal light performance for the actual driving situation.

The legal regulations are not yet completely ready for this innovative concept but the technique is now under full development. The ALP topic can cover a wide range of lighting aspects. We want to concentrate here mostly on the main headlamp functions, especially the low beam, because there seems to be the highest potential for improvement of night time traffic safety.

ALP - AN INNOVATIVE LIGHTING CONCEPT

During night time driving the automotive lighting equipment is responsible for illuminating the driver's viewing field under many different driving situations, street and traffic conditions. The following parameters characterize most of these situations:

- Speed:	low - medium - high
- Road condition:	dry - wet - slippery
- Course of the road:	straight - curvy - intersection ...
- Road topography:	flat - hilly
- Road type:	highway - country - city...
- Viewing conditions:	clear - rain - fog - snow
- Ambient illumination:	dark - street lights ...
- Traffic density:	low - medium - high
- Driver:	age - driving skills ...
- Other parameters:	country, tunnel ...

At the same time the car must be visible to other traffic members without causing glare or other irritations. Regarding this complex task the conventional static headlamp light pattern can only be a compromise. Usually most of the flux available for static illumination is directed to the central parts of the viewing zone including the range zone of the low or high beam (approx. ± 5° hor./vert. of driving direction). In this zone we certainly have the highest demand for good illumination and therefore also the highest requirements by current ECE and SAE standards.
The rest of the available static beam flux is more or less scattered to the left and right sides of the viewing zone. This light pattern is a compromise in the sense that it fairly covers the most important viewing zones but only to some extent the special zones needed for example for curve illumination, fog conditions, etc. This compromise is only for those headlamps a very good overall solution that have a total flux almost as high as in the case of HID headlamps. But the general idea - even for high flux systems - is, that light (e.g. foreground illumination), which is beneficial in the static beam pattern under normal driving conditions is most likely undesirable in some other situations (e.g. wet roads).
These basic considerations about the static and dynamic light pattern seem to be obvious but they have to be verified by scientific investigations.The goal is to evaluate the potential benefits of a dynamic or adaptive light pattern compared to the conventional static light pattern. Independent from basic research, a few technical realizations using rotatable reflectors have been presented years ago (e.g. Ref. 3, 4). First internal field tests and other investigations (Ref. 5, 6, 7) now prove that for most of the above mentioned driving situations different light pattern must be required to obtain an optimal illumination.
The finding is, that no static light pattern, even the sum of all individual light pattern, can deliver the overall optimal headlamp beam pattern. The philosophy of „Adaptive Light Pattern" can achieve this goal by means of detection of the actual driving condition and subsequent switching to the well-adapted light pattern.

The European ECE regulations do not yet allow the realization of the ALP concept. Both, the application of multiple light sources in one headlamp and moving beams are considered as important technical realization possibilities for ALP as will be pointed out in the next section. But it is neither permitted to use more than 1 light source for the low beam function within one headlamp (and more than 2 light sources each for high beam and fog beam) nor is it generally allowed to apply a movable or rotatable reflector.
There is an exception of this practice: an additional secondary high beam may be horizontally rotatable. Reflector parts may also be movable as long as their optical axis is not moved. Also, it is not explicitly forbidden to have adaptive low, high or fog beam if the adapted light pattern always fulfills the current regulation.
A European EUREKA project called AFS (Advanced Frontlighting Systems) with all major European bulb, headlamp and vehicle manufacturers participating (Ref. 8) has been established to make studies and proposals for new regulations to overcome the current restrictive ECE homologation practice. In the second phase the AFS working group is now testing the basics and limits of adaptive lighting in order to propose new regulations in the third phase.
The US Federal Register, Part 571 (FMVSS 108), for automotive lighting is not as restrictive as the European ECE regulation but the subject also is not yet explicitly introduced.

Once we have realized the potential of adaptive lighting in general and supposed there will be some regulatory framework in a few years, what

what kind of variations can be introduced by ALP? Various adaptation possibilities can be thought of:

- Adaptation of low, high or fog beam pattern
- Integration or combination of low, high and fog beam into one beam
- Adaptation of a headlamp to different standards: ECE left hand drive, ECE right hand drive, USA, Japan
- Variation of shape of the cut-off of a low or fog beam
- Variation of beam inclination
- Variation of the signal image of a headlamp (the appearance to opposing traffic)
- Variation of brightness or colour
- On/off - automatic switching of beams

...

The first tests and investigations already mentioned above allow some preliminary conclusions which help to select the most important variations and parameters for „Adaptive Light Pattern" headlamps:

Experimental Experiences:

- Speed is a suitable control parameter
- Distinction between city traffic, 25 Mph-zones, rural roads and highway is meaningful
- Oncoming traffic on rural road: Enhanced SPOT is advantageous
- Turns at intersections: Increased side illumination is very much welcomed
- Driving in convoy: Spot illumination should be reduced in favour of large side illumination
- Wet roads: Reduced foreground illumination helps to reduce glaring reflections to oncoming traffic
- Switching between different light pattern components should be harmonic and not too abrupt

Some of these conclusions are nicely illustrated in Fig. 1 showing schematically different driving situations from the bird eye's view and their adapted light pattern. In the left part of Fig. 1 the corresponding active components of a multi-segment headlamp - one realization technique for ALP - are indicated. The most important control parameter, which is quite easily accessible in a modern car, is the driving speed. Driving with high speed requires the attention of the driver to be directed to the far field in the direction where the car is heading instead of being distracted by the foreground scenery or the outer left and right sides of the total viewing zone. It is well known that the driver's view is attracted by those parts of the scenery that have the highest luminance (Ref. 9). Hence, the adaptive light pattern should support to direct the concentration to the most important zone for the current driving situation. Therefore, at high speed, an additional far field spot illumination should be added, accompanied by reduced or normal side and foreground light.

When driving at low or medium speed usually the spot is not needed any more. In this case, the adapted optimum light pattern depends on other parameters: When there is an intersection ahead, a curve or a broad street in an urban area then an enhanced lateral side illumination to the left and/or right is needed. In

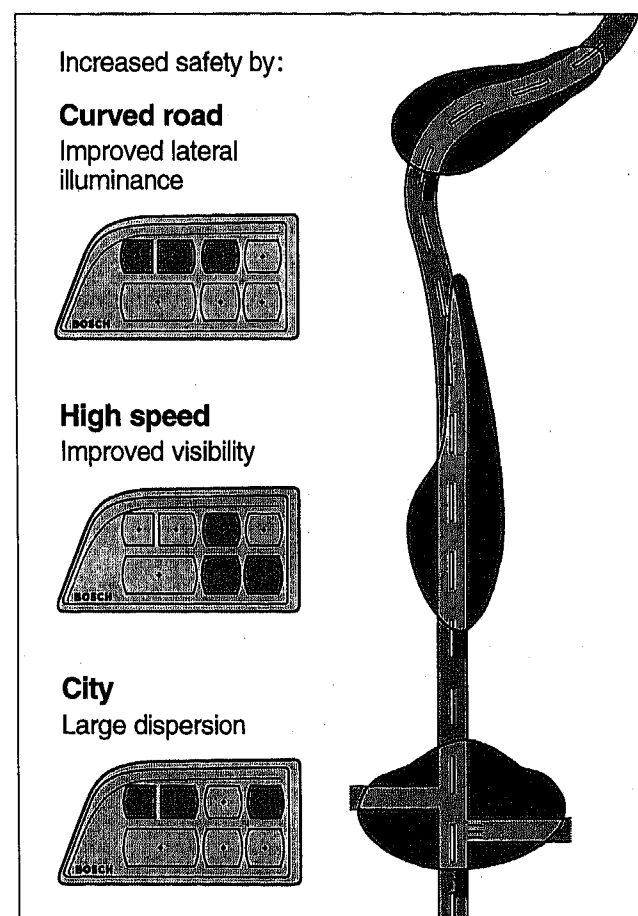


Fig.1: Three standard driving situations with adapted light pattern seen from the bird eye's view.

some cases, e.g. in a city, in a convoy or on wet roads without oncoming traffic (supposed there are means in the own car to detect oncoming traffic), an enhanced foreground illumination may be beneficial. Vice versa, on wet roads with oncoming traffic a reduced foreground diminishes specular reflection by the wet road ahead and therefore also reduces glare to other traffic participants. The transition between different adapted light pattern should always be harmonic and not too abrupt in a manner that has to be found out by detailed testing.

One important measure to assure safety and harmonic transitions is the idea to keep always a „basic light pattern" component static and switched on while the other beam pattern parts, e.g. spot, side illumination etc. will be adaptive.

ALP technique will be more complex than conventional headlamp technique is and it will be undoubtedly a challenge to incorporate it into modern car bodies. On the other hand it can also offer new styling opportunities as it is shown for example in Fig. 2. Recently, some show cars with such stylish model headlamps could already be seen on automobile shows around the world.

TECHNICAL REALIZATION ASPECTS

The design of a headlamp with „Adaptive Light Pattern" has to fulfill the same general requirements as the conventional lighting equipment: cost effectiveness, limited space, stability, life time etc. and it has to deal with the trend to more exotic vehicle front shapes, inclined lenses and low profile reflectors. Additionally, it has to incorporate the ALP-functions.

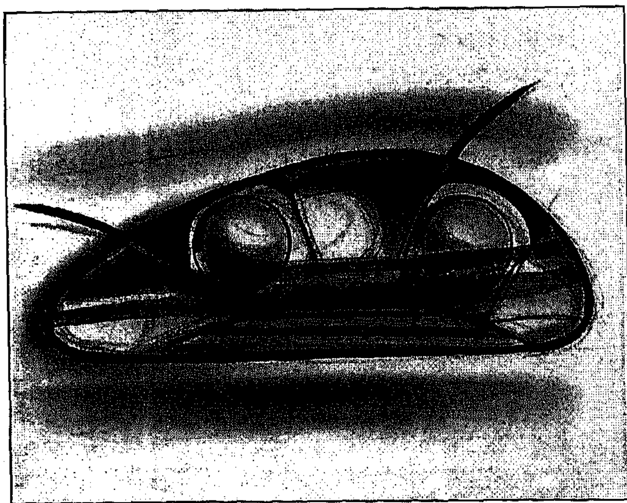


Fig. 2: Example for styling opportunities of ALP headlamps

Once the available space of an ALP headlamp and the cost frame is known it can be decided which technique will be used to make the light pattern adaptive. There is no feasible concept yet known which allows to realize arbitrary light pattern in a single headlamp. In reality, a few major light pattern states, e.g. curve illumination, together with the corresponding adaptation technique have to be selected to realize a powerful ALP headlamp.

There are several techniques to realize an adaptive beam:

- Changes of the light source(s) only
  e.g. moving the bulb inside the reflector, dimming
- Adaptive reflectors
  e.g. moving the reflector (and bulb), moving or deforming parts of the reflector, inserting new reflective surfaces, blocking parts of the beam...
- Adaptive Lenses
  e.g. moving lenses, inserting new lens parts or other optical elements, blocking parts of the lens
- Changes on the whole bulb-reflector-lens system e.g. moving the system, superposition of single beams
- Combinations of the described methods

...

A lot of patent applications and patents about all of these different methods can be found and it would go beyond the scope of this paper to comment on all of them.
Generally, motorized moving parts should be avoided in a headlamp or should be as simple as possible because complex motorized motions are expensive, need much space and are potentially unreliable. Often a simple movement of the bulb or a rotation of a reflector is not sufficient to achieve a well-adapted light pattern.
On the other hand continuous movements have the advantage that a beam pattern can be continuously changed to another beam pattern whereas other solutions often do not allow the realization of transition states between light pattern.
It has to be found out whether motorized motions of headlamp parts can be done in cost-effective technically elegant way in order to be attractive for ALP. Also, we find that blocking of parts of the beam pattern may be a reasonable technique to realize adaptive light pattern but we prefer solutions where no light must be blocked.
The technique, we want to describe here and which has been realized in samples, is the superposition of single beams.

Fig. 3 illustrates this method: An ALP headlamp with this technique consists of multiple sub-units each having a special light pattern (basic illumination, hot spot, curve illumination, etc.). Depending on the driving situation one to three or four of these sub-units will be activated for superposition of their individual light pattern to build the adapted total light pattern. One advantage of this method is that it is rather inexpensive and the assembly is not too complex if it is possible to merge the sub-reflectors to one or two plastic parts. Also, the sub-units can be optimized very well for special illumination tasks (e.g. for the hot spot) and this technique is very reliable as well.

The number of adaptation states must be limited with this technique because each of the multiple sub-units needs considerable space. It is very important to determine in the concept phase of an ALP project the most important driving situations that shall be adaptively illuminated by the headlamp. This yields the total number of sub-units. Then the available headlamp area must be divided in a way that all individual light pattern are well balanced to each other. It is advantageous to use free form reflector layout for some of the beams. This technique, which has been described in an earlier paper (Ref. 10), allows the application of clear lenses and hence avoids some problems with profiled lenses when dealing with multiple closely spaced beams. Fig. 4 shows our early prototype sample which has been realized using multiple free form reflectors incorporated into an existing frame.

This sample received very much interest from our customers as well as from the public and it was useful for some of our first ALP tests. Interesting experiences were collected and contributed to the first conclusions mentioned in the previous chapter. It also turned out that the lighting functions, the balancing of the single components and the operation modes still had to be optimized. Therefore, a further prototype sample is under development. Fig. 5 shows schematically the layout of this new system which again will be based on the superposition method using free form reflectors.

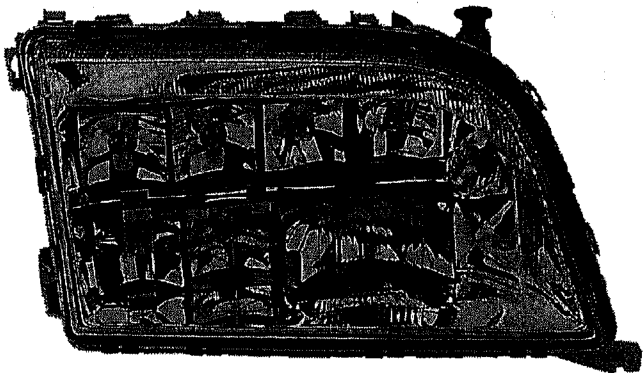
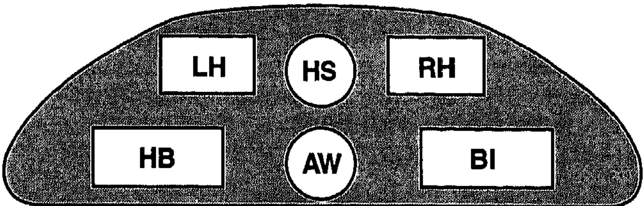


Fig 4: ALP headlamp sample



Single components:

LH Left hand bend
RH Right hand bend
HB High beam
BI Basic illumination
AW Adverse weather illumination
HS Hot spot

Fig. 5: Advanced ALP sample layout

This sample will include the basic low beam illumination which is most often active, it will also include a high beam function, a hot spot illumination, a left and right curve light pattern and an adverse weather beam. Both samples, the early one shown in Fig. 4 and the new one, were designed to fit into existing series production headlamp frames. Nonetheless, it would be also feasible to arrange all reflector components in one horizontal line in order to get a low profile headlamp with not more than 70-80 mm in height.

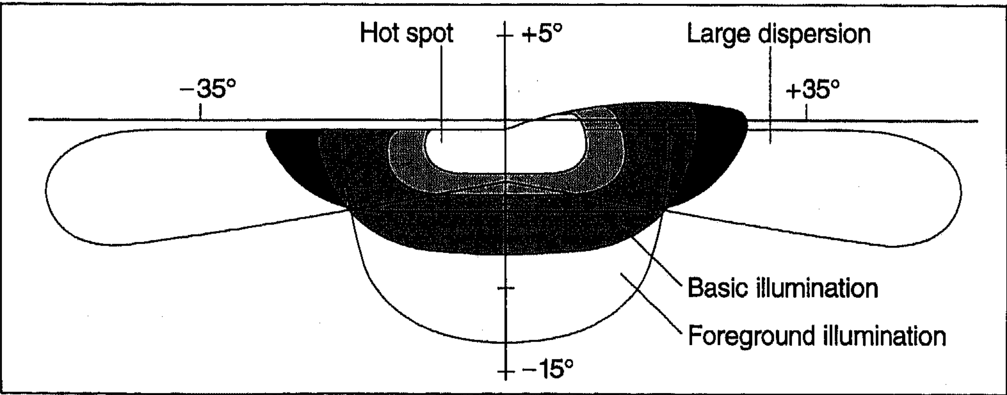


Fig. 3: Light pattern of multiple beam headlamp (Isolux curves in wall projection): The individual light pattern of multiple sub-units sum up to the total adaptive light pattern

The lighting performance of three of the sub-units in terms of isolux plots is demonstrated in Figures 6 and 7. Fig. 6 shows a comparison between the basic illumination and the additional hot spot illumination. The effect of enhanced range on a straight road by additional hot spot lighting is demonstrated.

Similarly, Fig. 7 shows the effect of curve illumination in comparison to the standard or basic low beam. The two figures clearly show the benefit of a well-adapted light pattern and the excellent performance of this prototype headlamp sample. An additional hot spot improves considerably the driver's view when driving on a straight road in an upper speed range.
Likewise, in a narrow bend a special curve illumination makes pedestrians or obstacles visible for the driver which normally will be hidden in the dark.

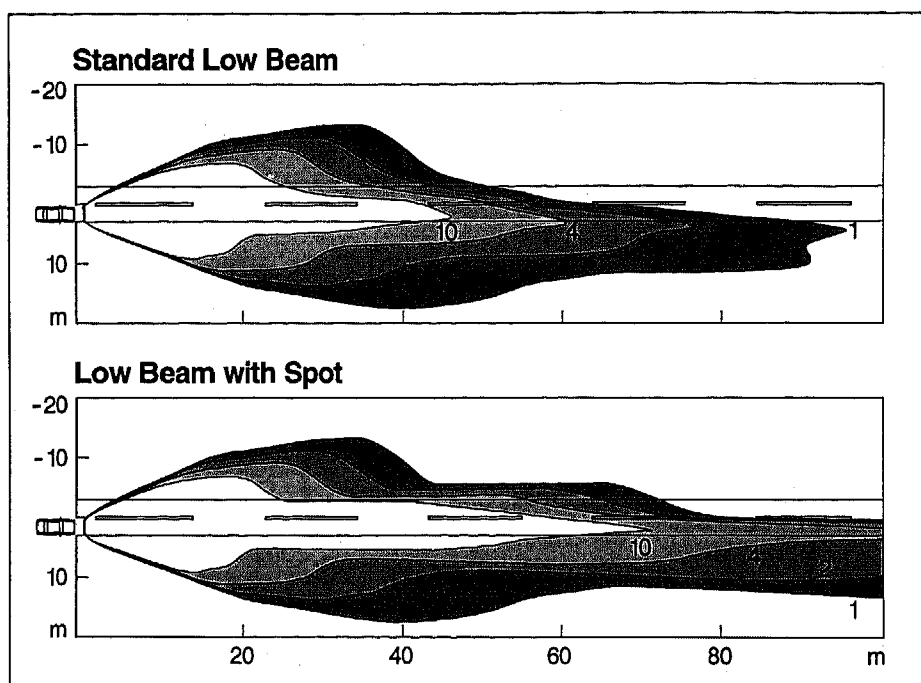


Fig. 6: Comparison between basic low beam and basic low beam with hot spot. Isolux curves from bird eye's view

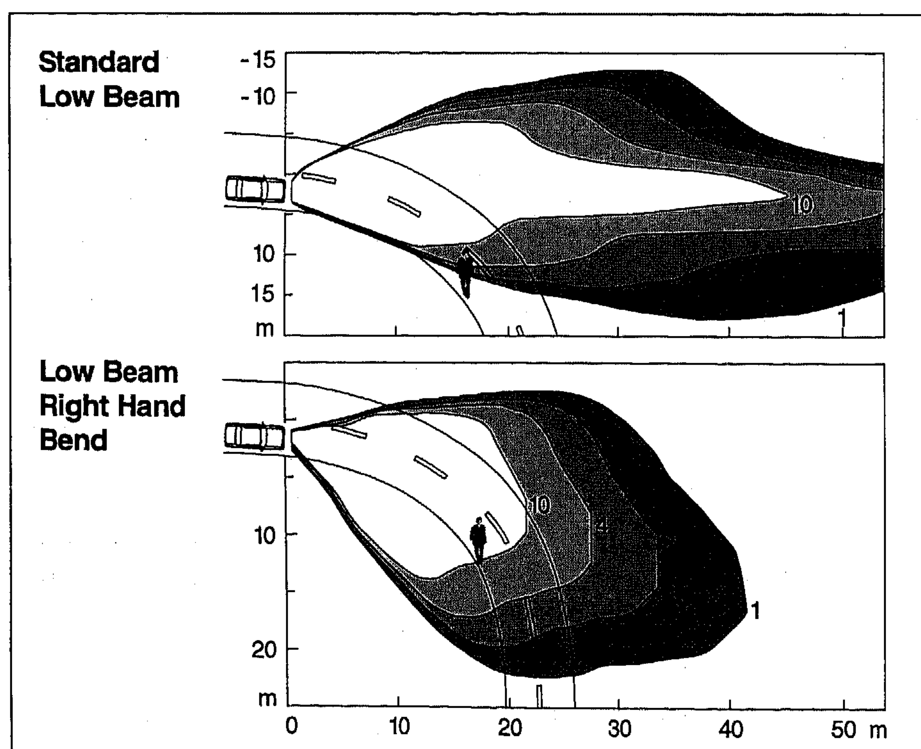


Fig. 7: Comparison between basic low beam and right curve illumination. Isolux curves from bird eye's view

CONTROL OF ADAPTIVE LIGHT PATTERN

The different adaptation states of an ALP headlamp must be switched or controlled in an appropriate manner. Depending on the number of states and their complexity ALP could be operated manually by the driver. But an additional safety system should not impose more tasks on the driver, it should let him concentrate on his main task to drive his car. Therefore an automatically controlled ALP is necessary. This also offers the opportunity to introduce, together with ALP, even more automatic functions, e.g. automatic light on/off or automatic headlamp leveling etc.
For this electronic control the following components are necessary (Fig. 8): Sensors, which capture the relevant driving parameters, an electronic control unit (ECU) including the switching logic, a power unit and a link between these components, preferably an advanced bus system like CAN (Controller Area Network).
The ECU and the power unit should have a rather general layout with a sufficient number of sensor signal input channels and output channels in order to cover a broad variety of ALP application cases. The sensors and sensor signals will be selected according to the specific application.
The sensor signals of speed and the turn signal state may be considered as a basic set of control parameters which almost every ALP headlamp will use. Adding more and more control parameters by supplying the corresponding sensor signals will increase the number of driving situations that can be detected and adapted to (if the headlamp supplies the corresponding light pattern)

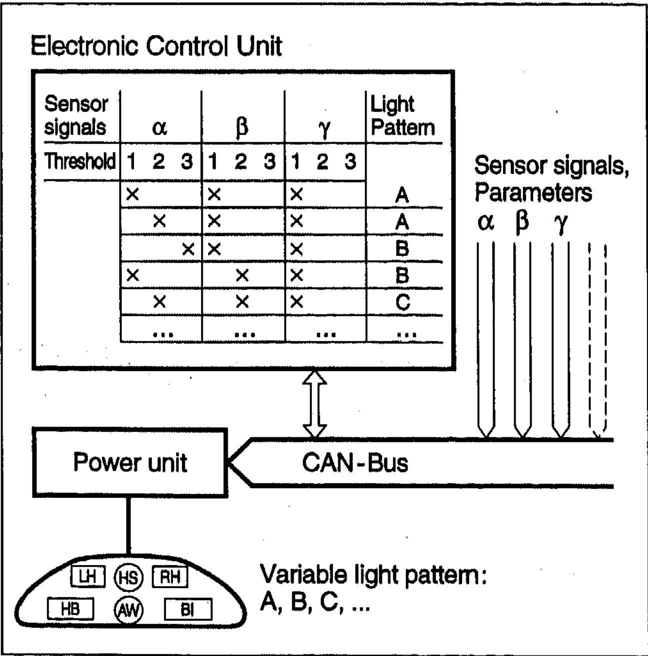


Fig. 8: Basic concept of ALP electronic control.

and it will of course increase the system costs. Hence, a modular control concept can handle ALP applications with a few basic adaptive states and also more complex applications:

	Control Parameter	Light Function
1	Speed, turn signal	Hot spot, improved lateral dispersion intersection illumination
2	+ Steering angle	+ Curve illumination
3	+ Rain sensor	+ Adverse weather condition illumination
4	+ Radar and video sensors	+ Drive in a columne enhanced hot spot in meeting situation
5	+ Navigation	Predictive control to all kind of driving situations

This list , which can be further extended, shows the enormous potential of a well-designed automobile lighting equipment together with an intelligent electronic control concept. To keep the system costs within an acceptable range it is worth while thinking about an integration of the ALP control with other more and more frequently used advanced electronic systems in the car. This may be for example the ABS (Antilock Braking System), VDC (Vehicle Dynamics Control) or ACC (Automatic Cruise Control) which can deliver speed, steering wheel angle, distance and other signals. Also the on board navigation, e.g. the Bosch Travelpilot, which is increasingly often integrated in upper class vehicles, can be combined with the ALP control. Using navigation as a signal source has the advantage that certain signals which are related to the road map (e.g. curve radius) can be delivered in advance. Thus, a curve or an intersection can be adaptively illuminated before reaching actually the point when the steering wheel is turned or the turn indicator is operated. Night time driving tests with the available prototype samples will help to find out the specific properties of predictive and non-predictive control. The ECU

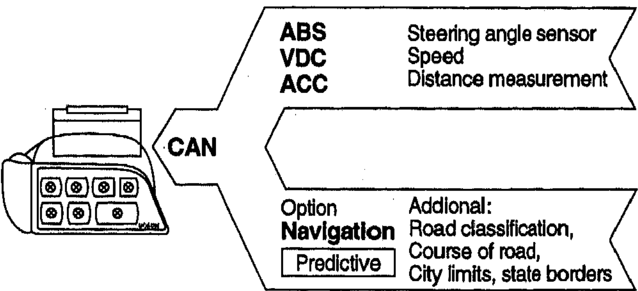


Fig. 9: Possible integration of ALP control with other advanced electronic vehicle control systems

logic will then be programmed correspondingly to apply the adapted light pattern with the right timing.

CONCLUSION

Headlamps with „Adaptive Light Pattern" (ALP) can improve current headlamp illumination through continuous automatic adaptation of the headlamp's street illumination to the actual situation.
Prototype headlamps are now ready for test driving. Advanced electronic control concepts for the complete car are available that can realize either simple direct automatic control or even high class predictive control (Fig. 10).
Additionally, the legislation process is started to formally allow this technique to be applied to series headlamp products in the future.

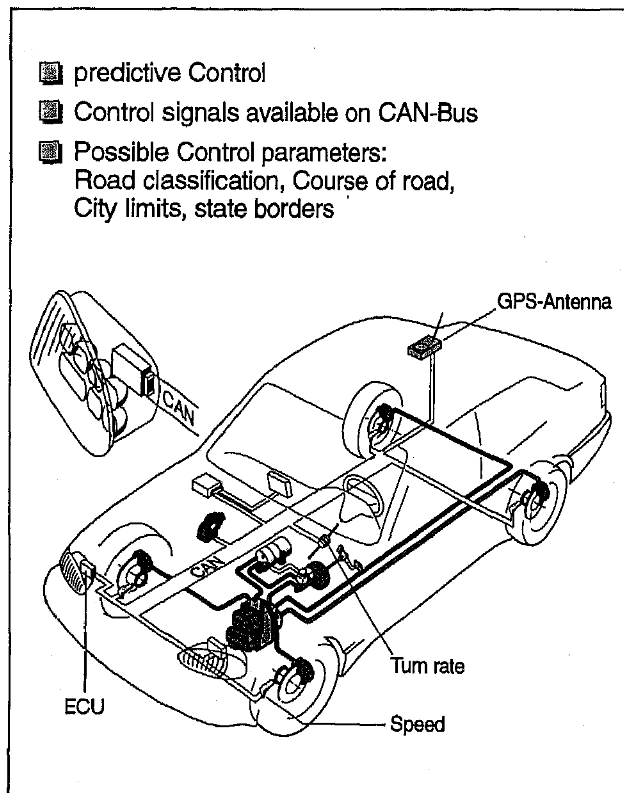


Fig. 10: Car equipped with ALP headlamps including predictive electronic control

Adaptive light pattern headlamps will contribute to more safety in night time traffic. They will help the driver in orientation and will serve as active guidance devices by means of the variations of the light distribution on the road.

REFERENCES

(1) W. Huhn, G. Hege:
High Intensity Discharge Headlamps (HID) - Experience for More Than 3-1/2 Years of Commercial Application of Litronic Headlamps, SAE Technical Paper Series 950591, 1995

(2) R. Neumann:
Improved Projector Headlamps Using HID (Litronic) and Incandescent Bulbs,
SAE Technical Paper Series 940636, 1994

(3) K. Wada, K. Miyazawa, S. Yagi, K. Takahashi, H. Shibata:
Steerable Forward Lighting System,
SAE Technical Paper Series 890682, 1989

(4) S. Kobayashi, M. Hayakawa:
Beam Controllable Headlighting System,
SAE Technical Paper Series 910829, 1991

(5) J. Damasky:
Geometry of the Road Area and Effects on Motor Vehicle Lighting,
Progress in Automobile Lighting (PAL), Symposium, Darmstadt, 26.-27. Sept. 1995

(6) R. Neumann:
Future Lighting Technology for Automotive Headlamp Systems
Progress in Automobile Lighting (PAL) Symposium, Darmstadt, 26.-27. Sept. 1995

(7) J. Damasky:
Lichttechnische Entwicklung von Anforderungen an Kfz-Scheinwerfer,
Dissertation, Darmstadt, 1995

(8) AFS - Advanced Frontlighting Systems,
EUREKA project No. 1403,
Phase 1 - Feasibility Study, Final Report, 1996

(9) M. J. Flannagan, M. Sivak, E. C. Traube, M. Aoki:
Effects of Foreground Illumination,
Wet Pavement, and Driver Age on Pedestrian Detection Distance,
University of Michigan,
Transport Research Institute,
UMTRI-95-23, 1995

(10) H. Hogrefe, R. Neumann:
Application of Free Form Reflectors in Modern Headlamp Systems,
SAE Technical Paper Series 950593, 1995

970645

Intelligent Lighting

Emmanuel Groh
Valeo - Lighting Systems

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Lighting and signaling functions are key factors for road safety and driving comfort at night. Night-time traffic represents 20 % of total mileage, but generates more than 50 % of fatalities. This percentage is even higher in adverse weather conditions.

New developments are being proposed to improve safety when driving at night. Systems which take account of environmental factors modify the beam pattern in accordance with driving conditions. Automatic beam modification is one of several projects grouped together in the wider approach known as *Intelligent lighting*. This approach differs from current vehicle lighting techniques which, being based on legal and regulatory requirements, entail design and testing in static conditions.

Intelligent lighting involves electronic data acquisition and processing, with the target of altering the shape, the relative intensity, the aim and the direction of the headlamps in accordance with the prevailing conditions. These external conditions include rain, fog, speed, road geometry and traffic density.

This highly innovative lighting system requires homologation by the existing laws and regulations, which is currently under review.

INTRODUCTION

THE PRESENT SITUATION - The current situation is defined by laws and regulations which consider lighting as a collection of static functions.

The conditions under which these different lighting functions are to be utilized are clearly defined, for example: high beam function when there is no traffic in front of the vehicle, low beam in traffic, parking lights in urban areas if ambient light levels are adequate etc.

These fixed functions correspond only partially, however, to actual traffic conditions. In the event of bad weather, or on a winding road, the driver often finds his lighting inadequate. The utilization of fog lamps in the event of dense fog is an appreciable safety factor in comparison with the simple utilization of the low beam. But in all cases, switching between the functions is the responsibility of the driver. The notion of automatic adaptation of lighting to the external conditions is rarely implemented, and is sometimes prohibited.

An initial analysis should enable us to determine which are the new functions liable to be of interest to drivers by amending deficiencies in present-day lighting systems. The possibilities are wide ranging, although there does seem to be a general desire to see certain specific functions. Of course, not all these new functions will necessarily be accepted by the different regulatory or legal systems.

The main new functions concern bad weather lighting, lighting for winding roads, special lighting for rural or urban zones and lighting for highways or more generally for lighting adapted to speed.

Several investigations into these different themes have been carried out by various equipment manufacturers and automobile constructors with the aim of proposing technical solutions which, on the one hand, would provide increased comfort and safety for drivers, and which on the other, could be accepted by different national regulatory systems.

The first task is to define the type of lighting which can best meet the needs of any particular given situation.

The second task is to define the technical solutions which will allow us to modify the power and the geometry of conventional beams in order to obtain a new function corresponding to the lighting required for this particular situation.

The third task, the development of automatic lighting management, constitutes a particular problem. How far can we allow a part of the lighting to be entrusted to automatic systems? What functions need such automatic responses? What functions can be managed automatically while maintaining correct safety levels? What normal functions can be automated to procure improved safety and responsibility?

PRINCIPLES OF INTELLIGENT LIGHTING

Data is acquired by sensors. This data reflects the parameters which are used to define whether or not the lighting function needs to be activated, and to configure the function appropriately.

The data is processed by systems based on micro-controllers or microprocessors. These systems analyze the situation of the vehicle relative to the lighting function being utilized and, if necessary, command the reaction of the lighting.

This reaction can take different forms:

Automatic activation or deactivation of one or more lighting functions.

Modification of a lighting function.

Warning to the driver that his lighting is unsuitable for the situation.

Warning to other drivers that their own lighting is unsuitable.

The first general principle is shown in Figure 1.

DATA ACQUISITION - An intelligent lighting system can acquire data relative to four types of parameters: (figure 1.b)

- The road environment: the type of road, existence of bends, gradients, passage through urban areas, tunnel, mountainous zone, etc.

- The ambient conditions: weather, luminosity

- Road traffic: density of traffic, oncoming or preceding vehicles

- The state of the vehicle: speed, acceleration, attitude, functional state of the lighting

A part of this data is already utilized by other systems present on the vehicle. For example, the speed of the vehicle is used by the dashboard display, the attitude and the acceleration by intelligent suspension systems, the ambient luminosity by certain air conditioning systems, or the presence of rain by certain wiper systems.

Potentially useful data may be acquired by sensors of all types. However certain information sometimes proves very difficult to acquire in a simple physical form. Defining and analyzing the surrounding traffic is extremely complex with reasonable resources, i.e. with resources compatible with the automobile industry.

DATA PROCESSING - The data acquired must be routed to one or more processors and control blocks. Future automobiles will certainly incorporate can type (controller area network) data transportation buses. This type of bus enables data to be distributed to the entire vehicle over a paired wire physical medium, which will enable the bulk of automobile system wiring harnesses to be significantly reduced. Each processing unit takes the data it needs, while each sensor unit transmits the data it acquires as quickly as possible by means of a communications controller.

The complexity of the processing necessary to such systems is increasing. Present processing systems are sufficiently powerful to respond to these new demands, at costs which remain low. The automobile lighting sector is no exception to this trend.

MODIFICATION OF LIGHTING - Automobile lighting is subject to laws and regulations which are relatively strict. The idea of developing lighting which is responsive to the surrounding conditions is already present in the existence of the four standard functions, parking, low beam, high beam and fog lights. The way to obtain increased safety and comfort is to improve the capability of lighting to respond optimally to surrounding traffic conditions.

This improvement implies both the modification of the four existing functions and the creation of new functions. To achieve this, it is necessary to provide the means to be able to modify the geometry, direction and distribution of the lighting, and finally to vary the power of the beams.

From this viewpoint, a number of different paths are being explored. The first is to increase the number of lamps, i.e. create additional lamps to perform new functions. This solution has the disadvantage of being costly and bulky, but it is by far the most simple; the second path is to investigate systems which can be used to modify an existing function in order to provide another. With this solution, for example, more than one light source may be used, the geometry of the reflector may be modified or the power supply may be increased or decreased.

BENEFICIAL FUNCTIONS

Future lighting systems must be able to adapt to a wider range of situations than at present. Drivers and manufacturers want lighting systems that respond to specific situations. Each country, each individual, has different requirements, but there are a certain number of situations which are generally perceived as dangerous or uncomfortable with present lighting. In other words there is a general desire for innovative new functions corresponding to a certain number of particular points.

Of the situations where current lighting is perceived as inadequate, bad weather is by far the one which poses the most problems. The presence of fog is a particularly critical situation. Present day fog lights are often judged inadequate, or too dazzling. Progress could be made by controlling the light flux in ratio to the density of the fog, for example, or by improving light distribution.

Loss of visibility on winding roads also represents a danger that needs to be eliminated. Glare, particularly as regards oncoming drivers in rain is another a problem of the same type. In both cases, modification of the geometry of the beam would provide a response adapted to the situation.

Automatic control of lighting is a particular challenge. While it is obvious that bending headlamps must be able to modify the beam automatically, the automatic activation or deactivation of the main functions poses a problem of safety. Automatic deactivation can result in a fatal accident very rapidly, so fail-safe controls for such functions are absolutely vital. Automatic control must include a notion of absolute reliability which must be absolute for certain functions. Advanced technical research needs to be carried out to guarantee that this type of automatic control system are fully certified.

SOME EXAMPLES OF DEVELOPMENT

THE BENDING LOW BEAM - Operating principle:

The purpose of this system is to provide the light necessary in the direction of the movement, principally around the bend when the vehicle is turning through a curve.

- Description of the system: (Figure 2)

The data required comes from several sources. The first is the steering wheel angle which enables the vehicle's angle of turn to be calculated. The speed can also be used to determine this target angle more precisely. The acceleration of the vehicle can also be usable data, together with the activation of the turn signals.

The most interesting data would be the direction of driver's eyes, but such a system of detection has not yet been adapted for automobiles.

Light can be provided in the target direction (Figure 2.b) in different ways: by pivoting the normal lamps, by pivoting additional lamps, or by activating other light sources, for example.

Valeo has developed experimental prototypes which sense the steering angle and which use additional lamps (Figure 2.c) whose movement is controlled by a system based on a microprocessor. Figure 2 shows how the system works:

The system provides the driver with illumination of around 20 lux in the target direction. At night, current illumination in a direction 30° to the right or left relative to the vehicle is normally less than 1 lux, which is inadequate for a correct view of the road.

THE MODIFIABLE HIGH BEAM - Operating principle:

The principle of this function is to be able to distribute the light of the high beam differently according to the speed of the vehicle. The beam will tapered to a longer range at high speed and wider at slower speeds.

- Description of the system: (Figure 3)

Speed data is already utilized for the instrument panel. This data can be recovered and transmitted to an electronic processing unit. A variation of the system utilizes acceleration/braking data when the vehicle is already at high speed; significant acceleration gives a more pointed beam, significant braking gives a wider beam. In this case, the acceleration data is acquired through the pedals or by a speed differential measurement.

Valeo modifies the beam by the interposing a lens between the high beam light sources and the reflector. (Figures 3.b and 3.c)

The pointed beam gives a maximum of 185 lux, as against 80 for the wide beam, but this latter provides 5 lux at 10° as against less than 2 lux for the pointed beam. (Figure 3.d)

AUTOMATIC ATTITUDE CORRECTION - Operating principle:

The purpose of attitude correction is to maintain the low beam at the correct height in order to avoid dazzling oncoming or preceding vehicles while at the same time ensuring the driver has optimum lighting.

- Description of the system : (Figure 4)

The most interesting data in this case is the distribution of the light beam. One good way of acquiring this data is by determining the light distribution on the ground. Another way is to determine the attitude of the vehicle relative to the ground immediately in front of the vehicle. This attitude can be calculated by measuring the attitude of the vehicle relative to the ground beneath the vehicle, that is by measuring the clearance of the vehicle's axles.

Valeo proposes attitude correction systems using 2 sensors to measure the height of the axles, a processing unit and lamps fitted with actuators used to change the attitude of the lamps.

Variations can take account of the ground distribution of the lighting, of the speed, and of the vehicle's acceleration. A variation might take account of the vehicle's speed to provide a function known as the highway or high speed function. In this case, the processor raises the low beam. (of less than 1 %rad, depending on speed). Another variation might take account of weather conditions, so that the processor lower the low beam (of about 1 %rad) in bad conditions, particularly in fog conditions.

The results obtained enable glare vis-a-vis other vehicles to be significantly reduced, while at the same time ensuring better lighting. An additional advantage is that the lamps are always correctly adjusted.

CONCLUSION

The current situation of vehicle lighting systems, with regulations principally based on static conditions, is necessarily undergoing change. This change is of interest for drivers for two essential reasons: SAFETY and COMFORT.

Intelligent lighting systems in fact procure improved comfort because they provide the driver with lighting which corresponds more closely with his visual needs in terms of amount, shape and position. By adapting to surrounding traffic conditions, intelligent lighting systems also provide drivers with improved safety, particularly in difficult conditions. The need for better lighting is urgent in certain traffic situations, particularly for elderly or visually impaired persons .

The automatic control of certain lighting functions will also provide further improvements in safety and comfort. For example, automatic deactivation of rear fog lamps when a vehicle approaches from the rear will avoid glare for the following driver, thus providing increased comfort and safety.

The types of function described here indicate a clear way forward. The utilization and combination of the functions described will undoubtedly improve automobile lighting.

The future may even see automobile lighting systems designed for data acquisition by camera, or possibly by multi-wavelength reception. Image processing could be used to acquire the total vehicle situation, and to define the best possible lighting. Other safety systems could benefit from devices to determine such functions as braking when closing on an obstacle or a vehicle. This type of approach reinforces the overall vision suggested by intelligent lighting, not only for the lighting system, but more generally for the complete transportation system that is the automobile.

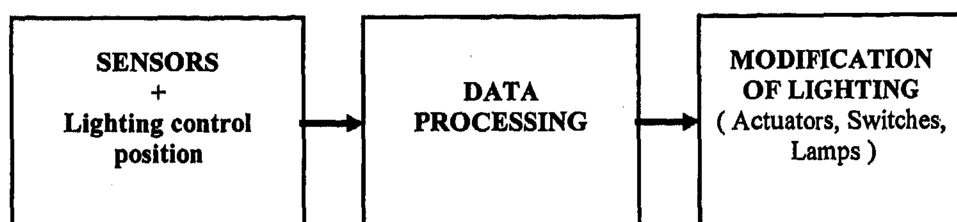


Figure 1

PARAMETERS	STATE OF THE ART	FUTURE SOLUTIONS
Weather conditions	Existence of adequate sensors, special difficulties with regard to fog detection	Possibility of determining these parameters at reasonable cost
State of the vehicle	Existence of adequate sensors	These parameters are easily acquired, existing or easily adaptable solutions
Road environment	Research, development of special sensors (tunnel, street lighting etc.)	Limited to particular points, or complex system with image, infrared, UV or multi-wavelength processing
Road traffic	Complexity of effective systems, development of systems detecting following or oncoming vehicles	Complex system with image, infrared, UV or multi-wavelength processing, or limited to detecting non-complex circumstances

Figure 1.b

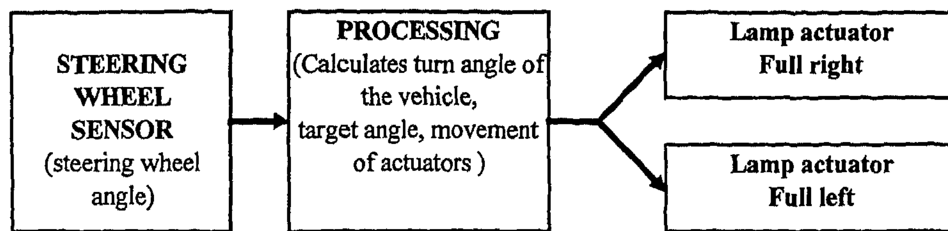
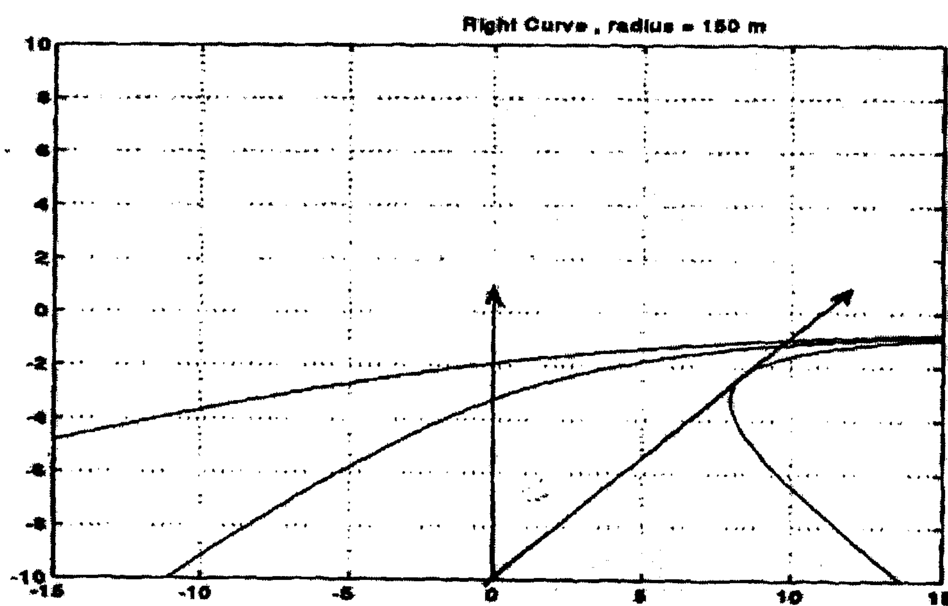
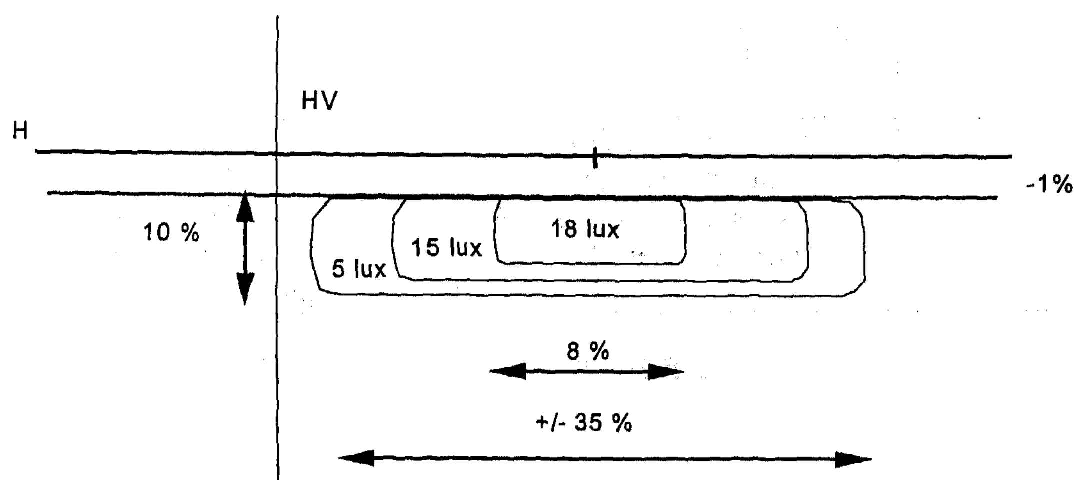


Figure 2



Direction of the vehicle, target direction
Figure 2.b



Additional Lamps for bending low beam system

Figure 2.c

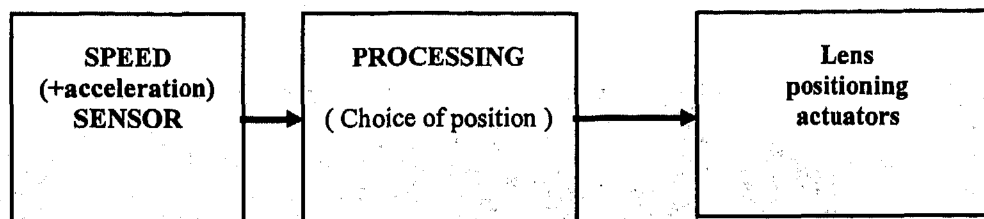


Figure 3



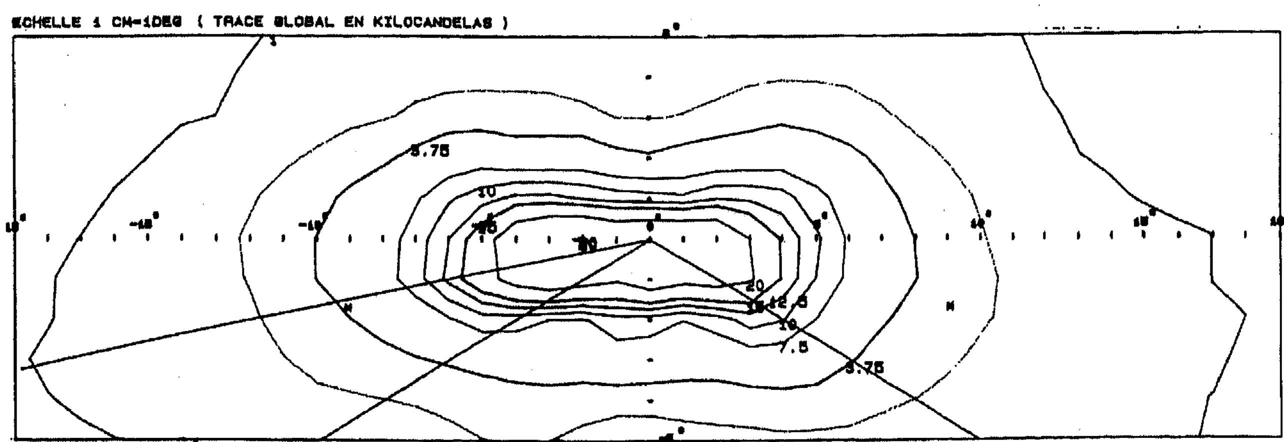
High Speed Position

Figure 3.b

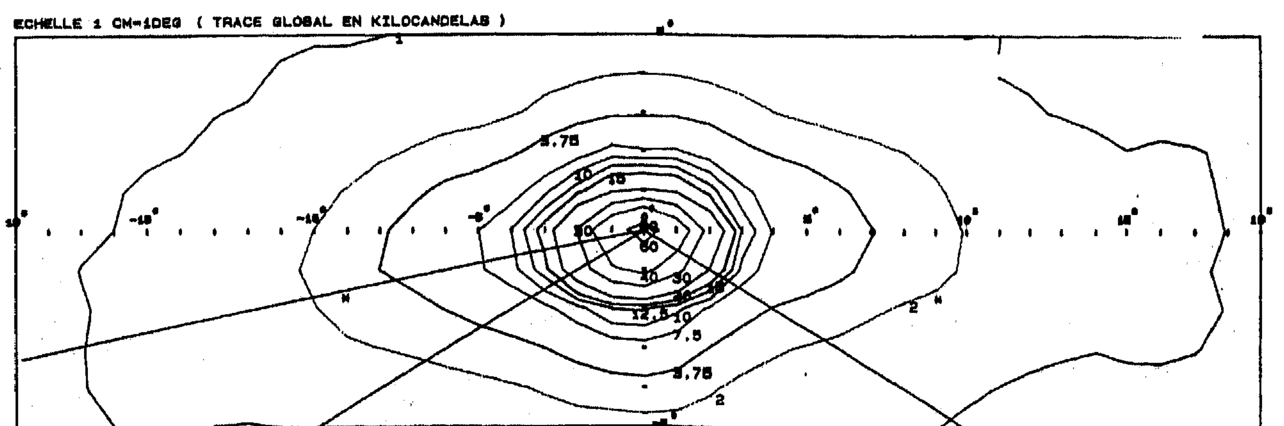


Low Speed Position

Figure 3.c



Wide high beam (with lens) in kilocandelas



Pointed high beam (without lens) in kilocandelas

Figure 3 d

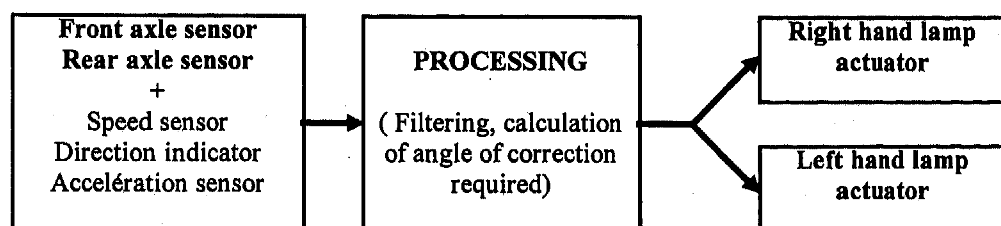


Figure 4

970647

Variable Headlamp Beam Pattern - Lighting Requirements for Different Driving Situations

Joachim Damasky
HELLA KG Hueck & Co.

Wolfgang Huhn
BMW AG

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Light distributions of headlamps are defined as static values. Neither FMVSS nor ECE allow variable beam pattern for low beam headlamps. New electronic systems and sensors available for car applications facilitate a situative lighting, that prepares an optimum lighting for each driving Situatuion.

This paper shows the necessity of different beam patterns for certain driving situations. Based on an road geometry analysis and an investigation of objects in the street the optimum light distributions are carried out for the single situations.

1. Introduction

New technologies and light sources open new possibilities in headlamp design. On the other hand the infrastructure of the available informations in cars has changed in the last years dramatically. New cars with digital bus systems have all informations available everywhere in the car. Also the efficiency of the microprocessors increased in the past. As a result of all these changes during the last time but also of the changes in passenger car traffic an adaptive variable headlamp beam pattern seem to be possible and necessary.

Paper No. 960920 shows the fundamentals of the investigation. A road geometry analysis, based on an video tape recording of more than 5500 km european roads, seen from a driving car, delivered the necessary data for the different driving situations.



Fig. 1 Evaluation of objects in the road based on video

The video tapes were analysed with the help of a special computer-video system. The analysis took account of the course of the roads, the brows and dips, and the positions of all objects relevant to lighting technology.
The small black lines are adapted to the road markings and describe therewith the course of the road. The radius of the curves, of crests and sags but also the with of the road is therewith exactly determined.
The white lines, adjusted to the base point of an object, give the exact distance between object and video camera.

The white squares mark the position of objects relevant for lighting equipment, such as traffic signs, delineation reflectors or the mirrors of the car ahead.
A result of the video tape analysis the road situations were differenced into the following types:

1. Urban Roads
2. Smaller country roads (straight or bended)
3. Larger Country Roads (straight or bended)
4. Motorways

Parameters were the width of the roads, Radii of curves, public lighting e.g..
In the following the details of the certain roads an the resulting beam pattern are explained.

1. Urban roads

Task of a urban beam pattern is the illumination of the foreground in front of the car, but also the objects in the roads like pedestrians and traffic signs. It has also to ensure the recocnition of cars with a minimum of dazzle to other road users.

Appr. 1100 km roads in european urban situations were analyzed. The position of traffic signs but also of all objects like pedestrians were determined. The following figures show examples of the evaluated data.
Figure 2 shows the distribution of overhead signs in urban roads, Figure 3 the locations of pedestrians.

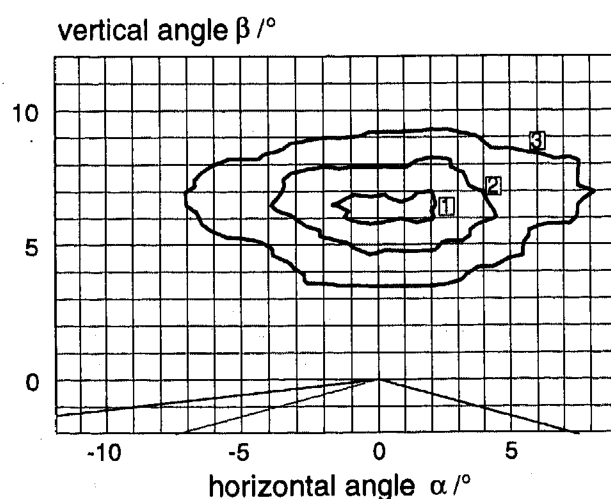


Fig.2 Locations of overhead signs on urban roads
dist. 50 m,
1 = 10%-, 2 = 50%-, 3 = 90%-propability-area

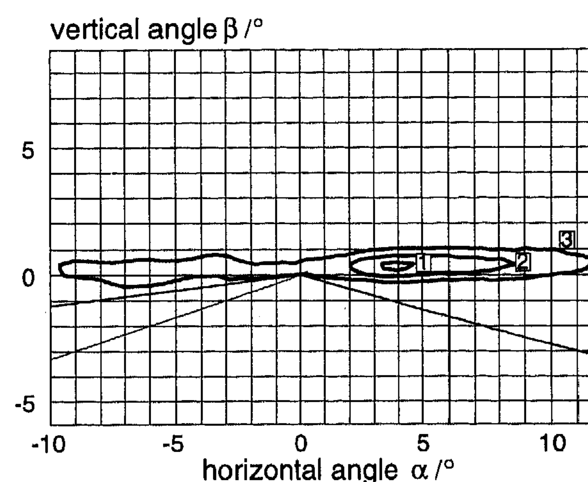


Fig. 3 Distribution of pedestrians on urban roads
dist. 50m
1 = 10%-, 2 = 50%-, 3 = 90%-propability-area

Based on the results of the investgation the following parameters have to be considered:
The foreground illumination has to provide the adaptation level for areas with lower public lighting. For less glare caused by dynamic movement of the body of the car the sharpness of the cut-off line has to be softer than for a country or motorway light.
In the areas of the pedestrians a minimum illumination is recommended. The positions of fig.3 mark the center of the body, the feet positions are appr. 1.14° lower.
In the areas of the traffic signs, overhead, left hand and right hand mounted, a minimum illumination must be ensured.
As the result of the investigation of the urban roads the following beam pattern for an urban light was developed.

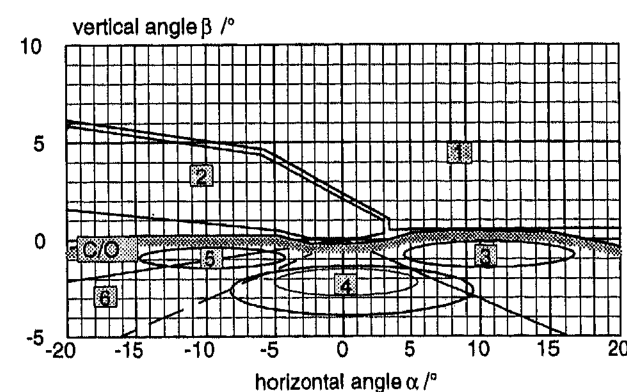


Fig. 4 Beam pattern for urban roads
zone 1 overhead traffic sign E < 2 lx

2 glare E < 1,0 lx
3,5 pedestrian feet area E > 25 lx
4 foreground right hand E >15 lx
6 foreground 5 lx < E < 25 lx, homogeneous luminance on road

It shows the consideration of the oncoming drivers in zone 2, the location of traffic signs in zone 1, but also the position of pedestrians and bicycles. Zone 4 and 6 is the area for the adaptation level and foreground illumination.
A first headlamp realized for urban traffic shows a wide illumination without a sharp cut-off line. Figure 5 shows a scematic birdseye-view of the road isolux diagram.

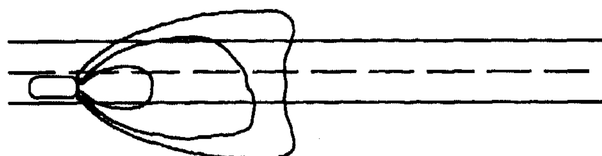


Fig. 5 scematic birdseye-view of the urban traffic light

2. Country Roads

The evaluation of the new country road beam pattern was described in the SAE-Paper Nr. 960920 from 1996. Only a short summary should be given here. For the country road light distribution appr. 2500 km country road have been evaluated. Considered objects were: traffic signs, delineation lines and reflectors, eyes of oncoming traffic, rearview mirrors of ahead traffic, but also parameters like curves and crests/sags.
On country road glare becomes a more important problem. Compared to urban surroundings, where the adaptation level is caused by public illumination higher and the glare sensitivity decreased a lower dazzle illumination is necessary.
The side illumination becomes more important for the driver too , from this area he gets the most important visual information for visual guidance and delineation of the road. The distribution of the objects was shown in Paper Nr. 960920, so figure 6 shows the proposal for the country light distribution.

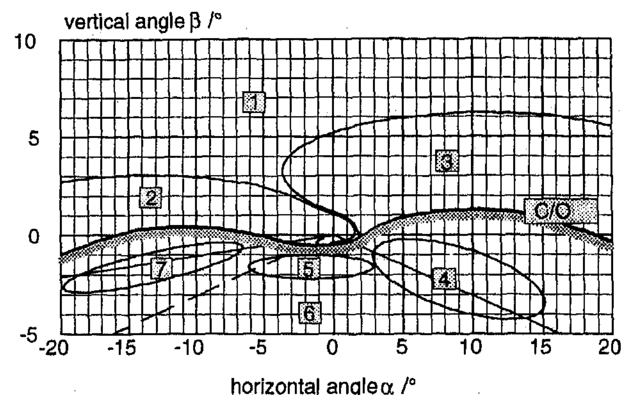


Fig. 6 Country light (fixed mounted part)
zone
1 overhead traffic signs 0,2 < E < 1,5 lx
2 glare area 0,5lx < E <1,0 lx
3 right shoulder traffic signs E < 2,5 lx
4 foreground right E > 15 lx
5 fixation area at Cut Off 50lx > E > 25lx
6 foreground 5lx <E < 15 lx, homogeneous resulting luminance
7 foreground left E >10 lx
on the 25m measuring screen

The realised light distribution. is shown as a scematic birdeye view.

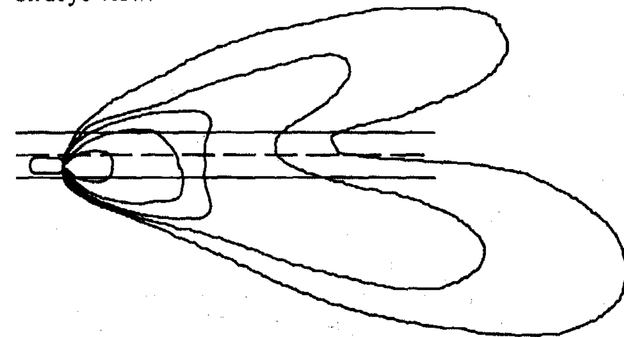


Fig. 7 scematic birdseye view on country road beam pattern

Out of all parameters of the geometry, but also considering lighting parameters like adaptation and glare the beam pattern shown in figure 6 and 7 was developed for country roads. In this light distribution no moving part of the light is included for the smaller, curved roads.

Test drives with this headlamp showed the advantages of this system: much better visual guidance and increased comfort without increased dazzle for other traffic.

But these test drives showed too, that the static light distribution is not an optimum solution for narrow, curved road. In these situation a moving part of the light, that lits into the curve, is necessary and very comfortable.

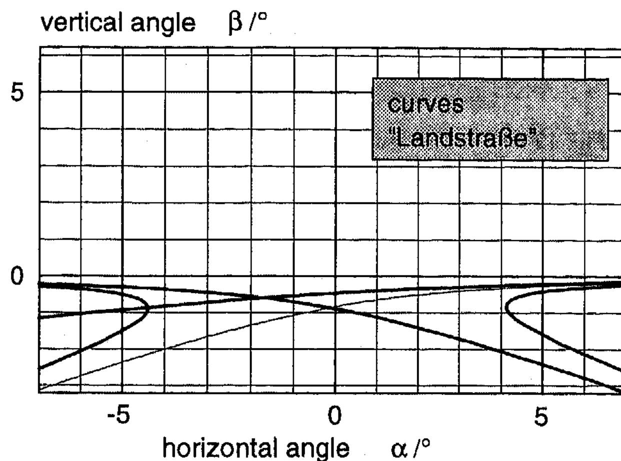


Fig.8 90%-area of curves on smaller country roads ("Landstraße") up to a distance d = 100m

As a result of the statistical evaluation but also from test drives a curve radii of appr. 2500 m is a conveniant threshold for a curve or bend lighting. Above this curve radius no bending lighting seems necessary, but below it provides a save and comfortable steering in curves.
The moving part of the light distribution needs no vertical component in its Cut-Off, the visual range can be shorter than the range of the low beam. The speed in curves with radii under 2500m is usally lower than the maximum speed on country roads. The maximum intensity should be in the range of the illumination of the low beam.

The lines in Fig. 8 show the edges of the road. They have to be illuminated dynamically. This could be realised by an electronic steering unit, that uses a gyro-sensor, a steering angle sensor or a navigation system as the necessary signal input.

3. Motorway lighting

A third driving situation is the Motorway. On motorways the lighting parameters differ from country roads by radii of curves, geometry of object locations, speed of vehicles, dynamic movement of car bodies and oncoming traffic.
On one hand the glare illuminations can be higher than on country roads, the distances between the oncoming cars are much larger. On the other hand the reaview mirror glare, resulting from the illumination above the horizon should be reduced.

The necessary parameters for a motorway lighting are:

- large viewing distance
- optimal optical guidance
- traffic sign illumination
- less dazzle for other road users

For alls these parameters data have been evaluated.

90% of the radii of curves on Motorways are larger than 5000 m .This is one of the results of the statisical evaluation. The number of lanes per driving direction is in Europe mostly 2, less than 5% have more lanes.
The problem of dazzle of overtaken traffic and traffic ahead increases with the possible light intentsities in the direction of the mirrors. Therfore the position of the mirrors was determined.
Fig. 9 shows the location distribution of left rearview mirrors of cars ahead on motorways in a distance of 100 m, Fig.10 the inside rearview mirror.

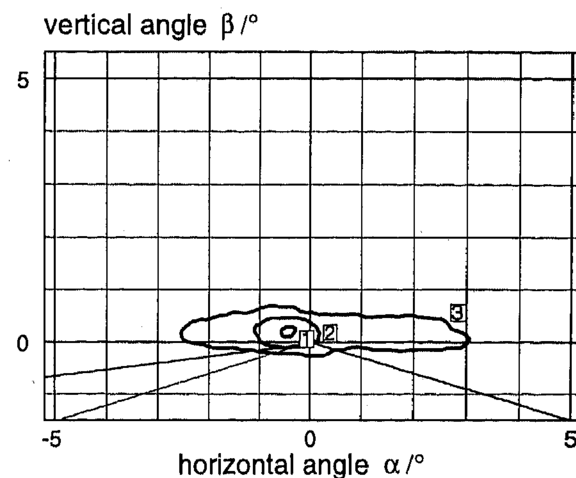


Fig. 9 Position of left rearview mirrors of ahead cars on motorways , distance 100m
1 = 10%-, 2 = 50%-, 3 = 90%-propability-area
related to headlamp, mounting height 0,65m

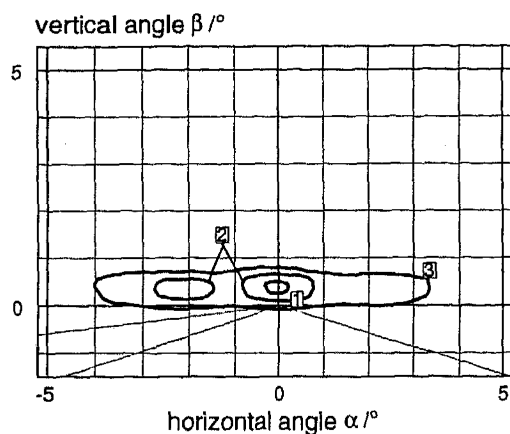


Fig. 10 Position of inside mirrors of ahead cars on motorways , distance 100m
1 = 10%-, 2 = 50%-, 3 = 90%-propability-area related to headlamp, mounting height 0,65m

The different 10 % and 50 % areas are a result of the lanes. It shows, that the observation car mostly drove on the right lane.

The areas shown in picture 9 and 10 show the most important glare areas on motorways. The glare illumination should not exeed a value of 1lx.

For save driving on night time motorways the legibility of traffic signs, right shoulder mounted and overhead, is important. Therfore the position of traffic signs was also evaluated.

The position of overhead signs on motorway is shown in Figure 11.

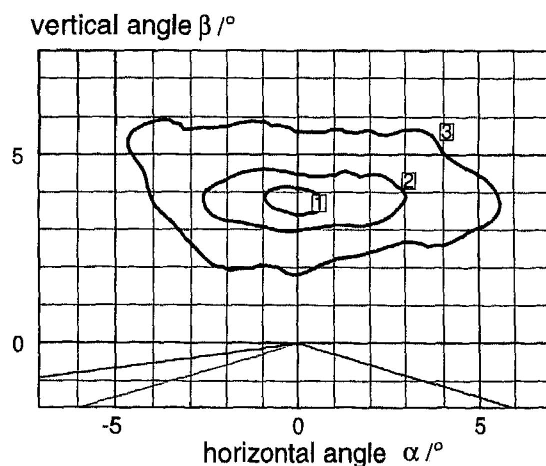


Fig. 11 Location of overhead signs on motorways
1 = 10%-, 2 = 50%-, 3 = 90%-propability-area related to headlamp, mounting height 0,65m

The result of the evaluation of these data, but also of consideration of the dynamic car body movement lines certain parameter out.

1. Sharp Cut-Off-Line with an almost horizontal shape
2. Defined rearview mirror glare of appr. 1lx
3. Illumination of traffic signs in certain zones.

Fig. 12 shows the resulting motorway beam pattern.

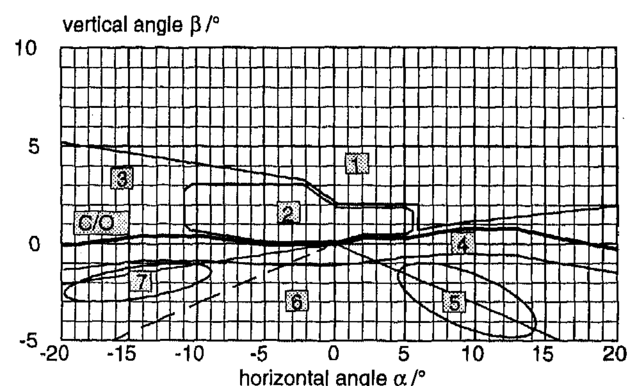


Fig. 12 Motorway light distribution

zone 1 overhead signs E < 2 lx
2 glare area E <1,0 lx
3 shoulder mounted signs E < 1,5 lx
4 fixation area 100lx > E > 25lx
5 foreground right E >15 lx
6 foreground 5lx <E < 25 lx, homogeneous resulting luminance
7 foreground left E >15 lx
on the 25 m measuring screen

Based on this proposal a motorway headlamp was realized. The light distribution of this headlamps is shown in Figure 13.

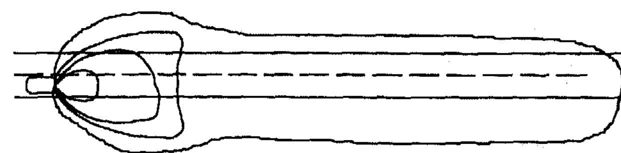


Fig. 13 birdseye view on scematic motorway beam pattern
It shows a long distance lighting for the distances up to 150 m in front of the car and a wide foreground illumination for the optical guidance.

A very comfortable driving on motorways is the result, confirmed by test drives. Test drives showed, that

maximum illumination of more than 100 lx are also possible and very comfortable to drive.
Motorway light in this shapes with high light intensities have to be combined with an dynamic levelling system to avoid glare for ahead and oncoming traffic.

4. Adverse weather lighting

One of the most important light distributions is the adverse weather lighting. Increased glare values and reduced object luminances decrease the avialable visual information for car drivers. Especially older driver feel very unsure and uncomfortable under these conditions.
The main problem is the glare caused by reflection on wet pavements.
As a basic investigation the reflection behaviour of wet pavements was determined. Especially the foreward reflection coefficient is of interest for the glare of oncoming drivers.
For this reason the measurement point 4 D exists in the US-Regulations. Modern free shape form reflectors enable a big amount of light direct in front of the vehicle. Therefore the reflection coefficient for foreward reflection was measured for different distances in front of the test car. The result is shown in Figure 14.

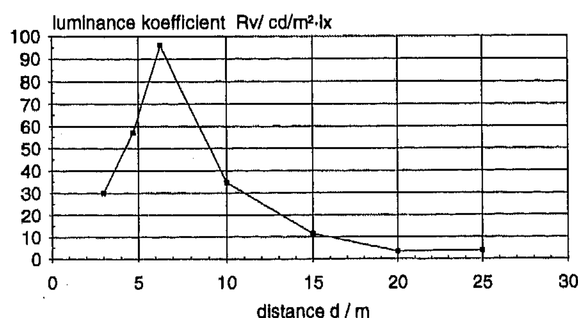


Fig. 14 Foreward reflection on wet pavement, distance between illumination and observer 30m

The Figure shows, that the maximum is not at the distance of 10 m, where it should be for illumination angle = reflection angle. The headlamps were mounted on a height of 65 cm, the observation height was 1,2 m. The maximum value was shifted to the illumination source. A limit for the foreground illumination is necessary.
Also a fog luminance computing software was worked out. With this system the calculation of the self-dazzling luminance in fog situations is possible.

The result of the investigation of adverse weather conditions is the development of the following light distribution, shown in Figure 15.

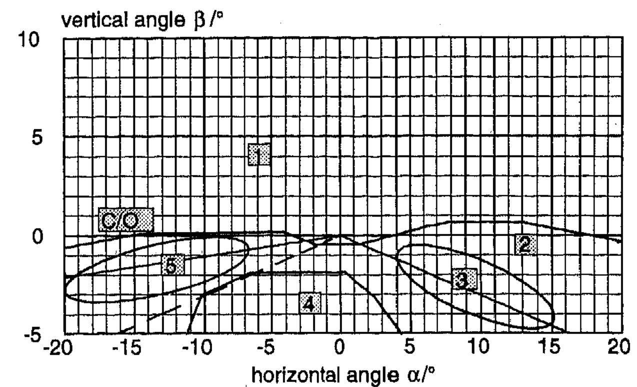


Fig. 15 Adverse weather beam pattern

zone
1 dazzle and self dazzle area E <0,2 lx, homogeneous
2 foreground 5lx <E < 25 lx, homogeneous luminance
3 delineation right E > 30 lx
4 reflex-glare zone E <5 lx
5 delineation left E > 10 lx

The necessary headlamps are in development. Zone 1 is the glare zone, where a low and uniform light level is recommended. Zone 3 + 5 are necessary fo the optical guidance, zone 4 is the reflection glare zone.

5. Eureka Project No. 1403 AFS

European lamp manufacturers, setmaker and carmaker work together for a situative, adaptive lighting in an Eureka-Project AFS. AFS (Advanced frontlighting systems) includes the following companies:
Carmaker: BMW/ Rover, Mercedes Benz, Volkswagen/ Audi, Ford, Volvo, Renault, PSA
Setmaker: Bosch, Hella, Valeo, Magneti-Marelli
Lamp manufacturers: Osram, Philips
The target of this partnership is a new ECE-Regulation, that allows more freedom in Lighting design towards a situative lighting.
The work is carried out in three phases:
Phase 1: feaseability study.
In this phase the target was to carry out if an AFS system is possible. Therfore the different functions:
1. City light
2. Country road light
3. Bending light

4. Motorway light
5. Adverse weather light
6. Overhead light
were tested in static and first dynamic test.
A market study in Germany, France, Italy and Sweden on customer demands sets priorities.
This phase was finished in summer 1996 an the conclusion was the recommondation for proceeding.
In the now running phase 2a test drives to all funktions were carried out in Germany and France under participation of all members.
Variable systems, which allow very different light distributions were build up and used for these test.
The phase with the scientific research work will start in 1997.
Target is a proposal for a new regulation before year 2000.

References

[1] EUREKA project no. 1403

[2] Damasky, J: "Lichtechnische Entwicklung von Anforderungen an Kfz-Scheinwerfer" (Development of Lighting Requirements on Motor Vehicle Headlamps) (Dissertation - Darmstadt 1995)"

970650

Development of Active Headlight System

Toshiaki Aoki, Hideki Kitamura, Kenji Miyagawa, and Makoto Kaneda
Stanley Electric Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The highest fatal traffic accident rate occurs on a curved road at nighttime. The "Active Headlight (AHL)" system developed by Honda R&D and Stanley Electric is the system that the beam pattern of the headlamp changes and more headlamp light is distributed toward a curve direction. The AHL system can direct a light toward a curve before the vehicle enters a curve by using road information from a navigation system. When the AHL system moved in nighttime, visibility on curves and stability of driving were improved and glare to an oncoming vehicle was acceptable.

INTRODUCTION

The number of nighttime traffic accidents has been increased in Japan [1]. Figure 1 shows the number of both daytime and nighttime traffic accidents on a yearly basis, assuming numbers for 1982 are 100 each. For the first 10-year period following 1982, the number of fatal accidents at nighttime rose almost 1.5 times. The fatal rate of nighttime accidents is higher than that of daytime accidents; the ratio is about 3 to 1. Figure 2 shows the fatal rate categorized by the type of road and time (daytime or nighttime). Fatal accidents on a curved road at nighttime are the highest among all traffic conditions.

To decrease these fatal accidents on curved road at nighttime, Honda R&D and Stanley Electric developed the "Active Headlight (AHL)" system. The parts of the development results were presented as the vehicle "Honda ASV-1" at the Japan Ministry of Transport ASV (Advanced Safety Vehicle) project in August 1995, and after that we continued improvement.

ANALYSIS OF ACCIDENTS ON CURVED ROAD

More than 90 percent of driving depends on incoming visual information [2]: insufficient visibility can be a critical factor triggering improper driver response. The visibility and eye-points, taken by an eye-marker camera, were recorded both at daytime and nighttime, and the comparison is shown in figure 3. At nighttime, it is harder, to see what is beyond a curve. Therefore, drivers tend to set their eye-point closer to themselves. Most of the drivers' eye-points at nighttime was distributed in the area illuminated more than 3 lux on the road.

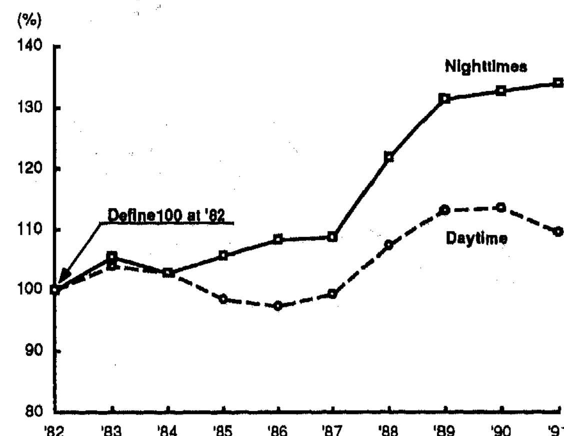


Figure 1. Fatal accidents number (in Japan)

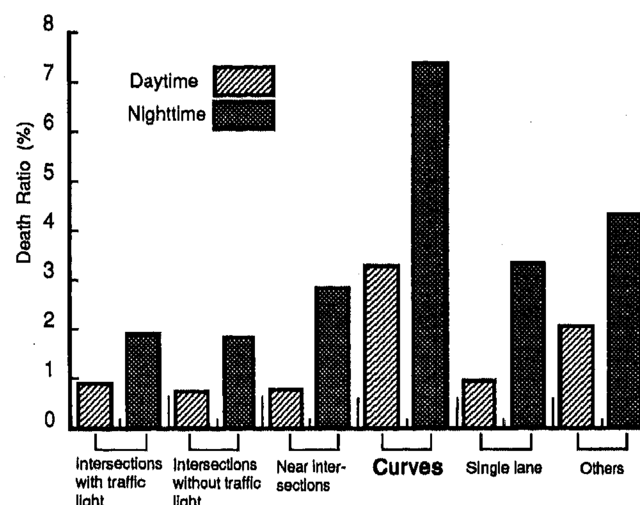


Figure 2. Fatal rate categorized by the road type

The Nagata's report [3] suggested that inexperienced drivers tend to set their eye-points closer to themselves compared with experienced drivers while moving around a corner, and that insufficient driving skill and closer eye-points are interrelated. Hirao defined [4] a margin time "Td" as the time the vehicle reaches to the driver eye-point. He reported driving becomes inconsistent when "Td" value is too small. "Td" tends to decrease particularly at nighttime, that cause inconsistent driving and accident.

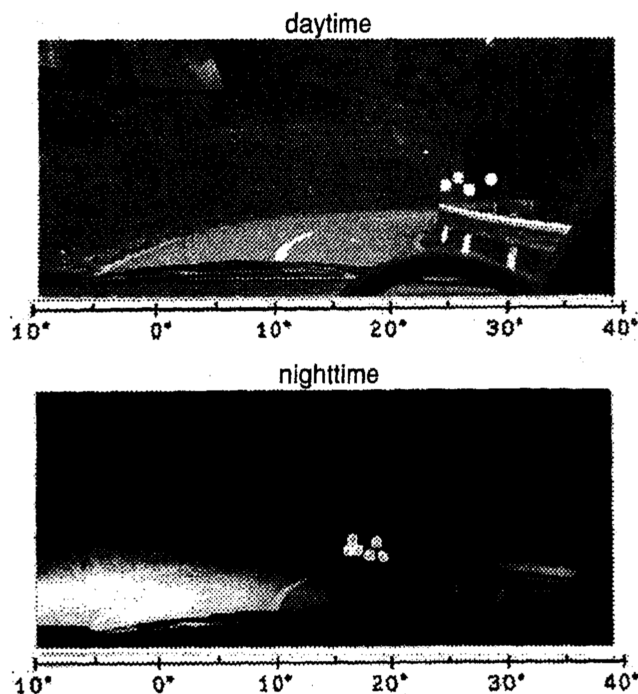


Figure 3. Eye-points distribution on a curve road

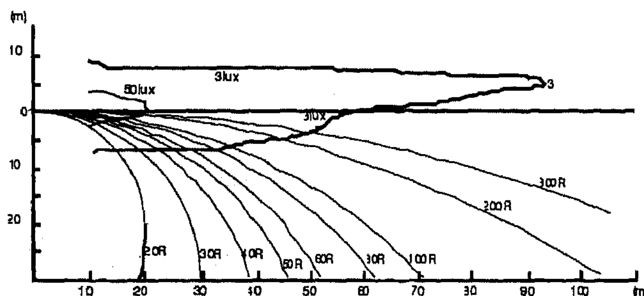


Figure 4. Conventional headlamp beam pattern

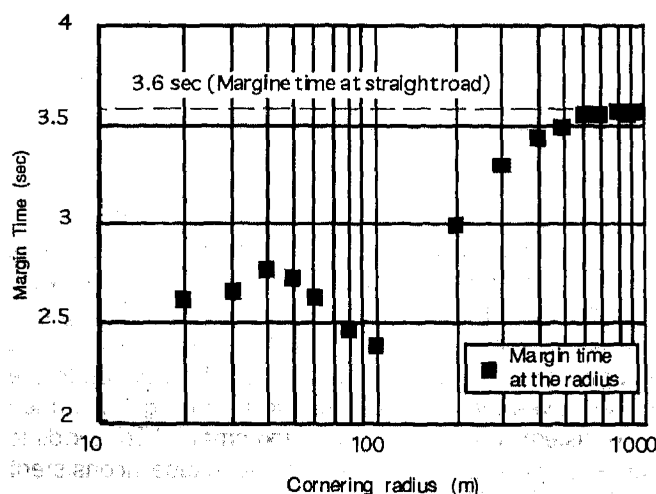


Figure 5. Cornering radius and margin time

Figure 4 shows the beam pattern of a current headlamp with various curved roads. Assuming that the driver's eye-point directs on the 3 lux line and the vehicle runs centrifugal force of 0.3 G, we estimate the margin time of each curves. Figure 5 shows the calculated margin time on curves when the figure 4 headlamp is applied. The margin time on straight road driving at 60 Km/h is about 3.6 sec. The margin time of each curves is evidently small.

The Active Headlight (AHL) system is developed to improve visibility, to increase the margin time and to stabilize the driving.

ACTIVE HEADLIGHT SYSTEM

OUTLINE

The main structure of the AHL system is shown in Figure 6. The system calculates headlamp turning angle using a steering angle, vehicle speed, turn signal and information of curves from a navigation system. In the preliminary experiments, drivers tended to see a turning direction before they operated the steering wheel. The system received a distance to a coming curve and the radius of the curve from a vehicle navigation system so that the system can direct the headlamp beam before the vehicle enters a curve. Curved roads ahead were predicted by the method of corner notice and warning [5] used in the "Honda ASV-1."

HEADLAMP

A HID is used as the light source. The system is used for lower beam only. The reflector of the headlamp is split into two parts, top and bottom. The top half reflector can rotate horizontally. The right-hand lamp can be rotated 50 degrees to the right and 20 degrees to the left. The left-hand lamp can be rotated symmetrically to the right-hand lamp. The top half reflector is designed to reflect the light under the horizon. The bottom half reflector is fixed and this part alone can meet low beam requirement. The total candle of the top reflector and the bottom reflector is meet low beam requirement at any rotational angles.

On a straight road, the movable reflectors are directed beams straight ahead, improving the visibility of an object far ahead. At a curved road or a cross-section, the movable reflector on the turning side is rotated more than the reflector on the other side so that the visibility in the turning direction is improved while overall light distribution is kept smoothly.

CONTROL

There are three ways to control the rotation angle of the reflector.

1) Control by steering angle
   When a vehicle is on a curved road, the AHL system calculates curve radius by the steering angle and the vehicle velocity, and determines the rotation angles of the reflectors.

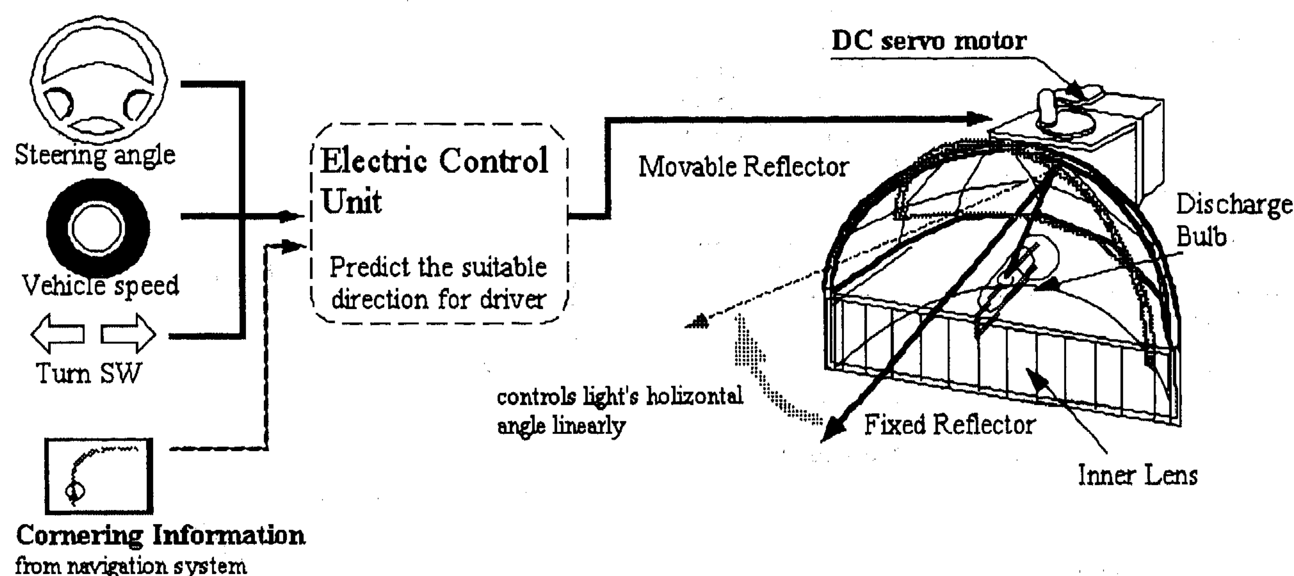


Figure 6. Structure of the AHL System

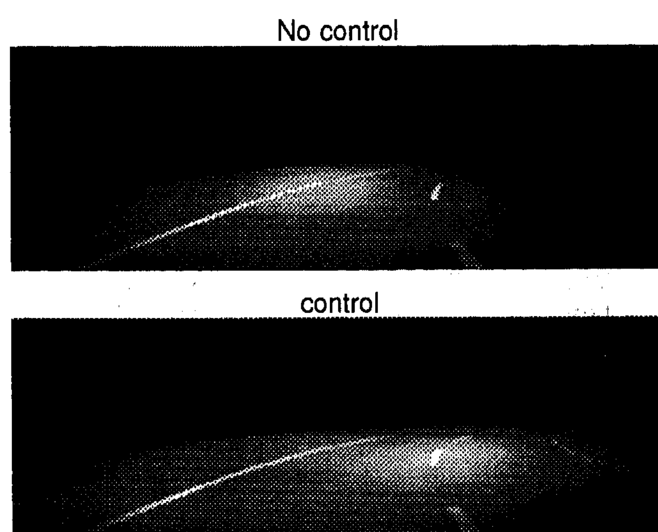


Figure 7. Illuminated area image by computer graphics

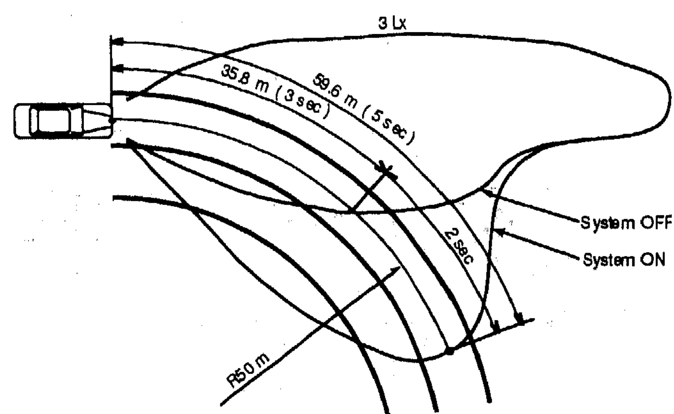


Figure 8. 3 Lux area of the AHL system

2) Control by turn signal
   When the turn signal is on, the system rotates the reflectors with a large angle. In case of the steering angle exceeds a predetermined value, the control will be switched to the "control by steering angle" mode.

3) Control by road information from the navigation system
   When the vehicle approaches a curved road in some distance, the navigation system sends an expected curvature information to the system. The system rotates the beam to adequate direction. In case of the steering angle exceeds a predetermined value, the control will be switched to the "control by steering angle" mode.

VALIDATION

EVALUATION OF VISIBILITY

Visibility on a curved road with a 50m curvature (hereafter called "R50") was evaluated using the AHL system. Three lux ISO-lux lines on a road surface of the AHL system and a conventional headlamp are shown in Figure 8. When the AHL system is on, the illuminated area with 3-lux reaches 60m along the curve. In case of the conventional type, illuminated area with an intensity of 3-lux reaches only 36m. The AHL system improved the visibility of the turning direction. If the vehicle runs at 45 km/h along the "R50" with the AHL system, a driver can detect road conditions two seconds earlier than with conventional one. Illuminated area images by computer graphics with each headlamp system are shown in Figure 7.

EVALUATION OF GLARE

We have also evaluated glare level for an oncoming vehicle. In the test of glare, 8 subjects sat in a stationary vehicle and they graded the glare level of an oncoming vehicle with AHL system and with conventional system. They graded with the de Bore scale. Figure 8 shows the results of the study. The AHL system increases glare, but it was still acceptable level.

	Left curve	Right curve
Conventional	7.0	4.5
AHL	5.8	4.9

Figure 9. The results of discomfort glare

SUBJECTIVE EVALUATION

Subjective evaluation of the AHL system has been done by Honda on the winding course. Six subjects drove the vehicle equipped the AHL system with two modes. The one mode was the AHL system was on, and the other mode was the AHL system was off. Figure 9 shows the results of the evaluation. The most subjects answered that the vehicle with the AHL system was comfortable, confident and better visibility on the curves.

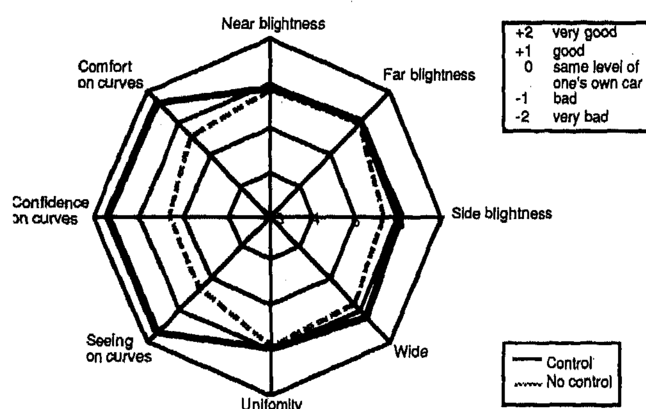


Figure 9: Subjective evaluation

EFFECT FOR EVASIVE OPERATION

The evasive action of a vehicle, driving on a constant radius, was evaluated to assess how the system would affect the driving. A subject drove the test vehicle on a constant 30m radius curvature at 40 km/h speed. Obstacles were placed at random on the course without knowing the driver. The steering angle was recorded by a sensor. The test results are shown in Figure 10. With the AHL system, evasive action became quicker and steering speed was decreased, steering wheel operation became stable. However, with the conventional system, evasive action was less responsive and steering speed tended to increase, resulting in more corrective steering action and unstable driving. Driving with the AHL system was closer to daytime driving than driving with conventional systems.

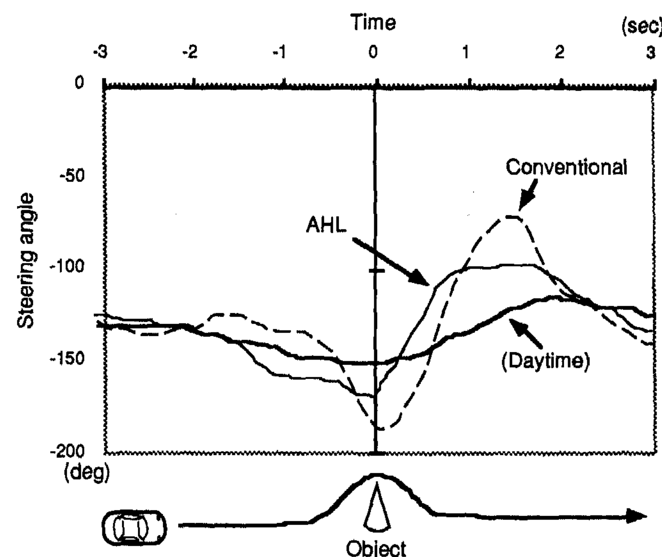


Figure 10: Evasive Operation

CONTROL BY ROAD INFORMATION FROM THE NAVIGATION SYSTEM

The most advanced feature of the AHL system is illuminating curve direction before the vehicle enter a curve by using the road information from the navigation system. The prediction method of a curve is called "Zone Judgment Method" developed by Honda. The detail of the method is described in the reference paper[5]. The AHL system can get the road information 5 seconds before the vehicle enter the curve. Waiting for 4 seconds, the AHL system starts to rotate the reflector to curve direction. The control angle is adjusted at the entrance of the curve.

CONCLUSION

In 1950's, there was a vehicle which third headlamp moved together with the steering wheel. The AHL system achieved accurate headlamp beam direction control and beam control in advance the vehicle enters a curve by using high accurate sensors and an advanced vehicle navigation system. As described validation section, the visibility on curves and the stability of driving task were greatly improved by the AHL system.

Last two years, high performance HID headlamps were introduced in the market. It seems to be still difficult to illuminate adequately in all curve conditions with this HID light source.

The AHL system can't be on the market immediately by the restrictions. Headlamp regulation should be changed along with the technological progress.

REFERENCE

[1] Night time traffic accidents: trend and characteristics, "OWL" congress for safety, 1993.
[2] Hartmann, E.: Driver Vision Requirements, 1970 International Automobile Safety Conference. New York, (1970)
[3] Nagata et al.: Vision and driving action of inexperienced drivers on a stationary circle, Automobile technology theses, vol. 1/2/3, No. 3 1992.
[4] Hirao, Osamu: Nighttime accidents in Japan and research on how to deter them, "OWL" congress for safety, 1993.
[5] Tamura et al.: A study of self-reliant cornering speed control system, the 14 th ESV, 1994 section.

970651

Progress and Perspectives of LED-Technology for Automotive Signal Lighting

Detlef Decker
HELLA KG Hueck & Co.

Uwe Thomas
Hewlett Packard

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

Of all the light sources for automotive applications, only the LED increased its efficiency by a factor of more than 100 within the last quarter of a century. The use of new chip materials and new LED packages raised the emitted flux in the latest "TS-AlInGaP" process to more than two lumens per LED. Offering an extremely high service life, low power consumption, little mounting depth and quick response time the first LED application for signal lights was the Central High Mounted Stop Lamp (CHMSL).

In order to continue this success for other signal functions further technical problems have to be solved: Highly efficient secondary optics have to collect the light and to form it into an adequate beam pattern. - New methods of electrical connection are needed to enable LED's to follow the 3D-shape of the car.

1 Introduction

The latest developments of highly efficient LED technologies in combination with an increasing number of package options enables the LED to compete cost efficiently in the area of automotive exterior signal lamps. Today LED's are already being used in many automotive interior and exterior lighting applications.

With several 10.000 hours life time, an LED exceeds the life time expectancy of a motor vehicle under harsh automotive operating conditions. This means that an LED lamp does not need to be replaced at regular service intervals anymore. Another major safety benefit is the instant turn-on time of an LED lamp. A study by the University of Michigan found out that an LED lamp is seen 170 msec earlier than a bulb. At a speed of 70 mph this translates to a possible reduction of braking distance by up to 15 feet.

LED's come in small package sizes and have little weight. That makes it easy to arrange the LED's in various array formats to serve difficult design requirements. With a new mounting option LED's even can be attached to a three-dimensional metal substrate which follows the curvature of a car body. The low power consumption of the LED technology helps to minimize the vehicle's electrical load, to reduce the diameter of copper-wires and minimize the heat generation of the LED lamp.

The color emitted by an LED depends on the composition of the semi-conductor material being used. The band width of an LED's emitted color spectrum is very narrow. It emitts pure colors and thus tinted filters to generate the required color are not needed. This makes it very easy to design unicolor appearances, e. g. a dark gray look for the entire signal lamp.

2 Chip Technology

2.1 LED Efficiency

In the past the efficiencies of conventional LED technologies such as GaAsP (Gallium Arsenic Phosphate) or GaP (Gallium Phosphate) were not sufficient to generate enough flux for exterior automotive signal lamp applications. These technologies have an efficiency of about 1 to 2 Lumen/Watt.

A breakthrough was made with an LED technology called AlInGaP (Aluminium, Indium, Gallium, Phosphate) about four years ago. This material has an efficiency of about 20 Lumen/Watt which is equal to the efficiency of an unfiltered incandescent bulb (see Fig. 1).

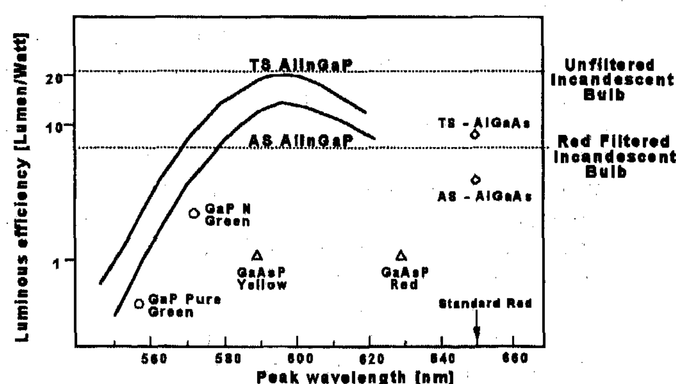


Fig. 1 Performance of LED Technology

With this tenfold increase in efficiency, suddenly the minimum required LED count for a signal lamp decreased by a factor of 10 for the same amount of flux required.

Light emission with AlInGaP is not limited to certain wavelengths. With variation of the doping levels, light green, yellow, orange and red colors are possible. AlInGaP generates its peak efficiency in the orange (~ 600 nm) color range. Improvements in the fabrication of the AlInGaP LED die will increase the efficiency even further.

2.2 AlInGaP Chip Structure

The crystal substrate of Hewlett Packard's AlInGaP LED's is made of a 200 µm thick GaAs (Gallium Arsenic) layer (see Fig. 2).

On top of the substrate there are three differently doped layers, each 1 µm thick, consisting of variations of Al, Ga, In and P. This structure is called a hetero junction structure.

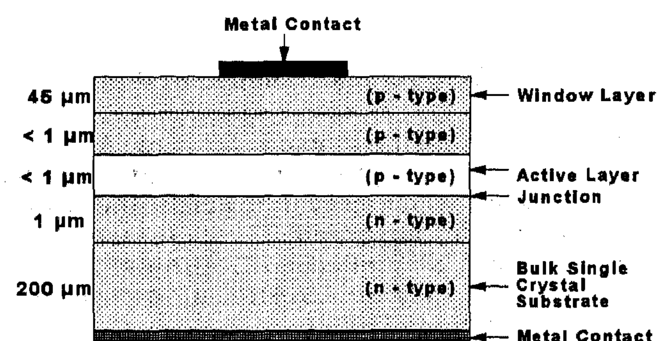


Fig. 2 AlInGaP Heterojunction LED Structure

The middle layer and the bottom layer form the pn-junction. The middle layer is also called the "active layer" because the distinct doping of the three layers forces the electrons mainly to recombine within this active layer. Due to its relative thickness this active layer provides more room for the electrons to recombine. This results in a higher maximum allowed drive current with hetero junction LED structures.

Standard AlInGaP LED's reach their maximum recommended current density within the junction at 50 mA while conventional LED technologies approach this limit at 30 mA. High current densities cause partial overheating which damages the crystalline structure of the LED die. At these areas, so called darkline defects, light ceases to be generated. The LED has degraded. The active layer of a hetero structure can handle higher current densities and thus is less susceptible to degradation.

On top of the three AlInGaP layers there is a 45 µm thick GaP window layer which creates the interface to the metal contact. The majority of the surfaces of an AlInGaP die cube do not expose any materials with metal contents. Only the sides of the very thin 3 µm AlInGaP layers have an aluminium content. Thus there is very little surface which could oxidise in a humid environment. The AlInGaP technology has shown excellent degradation results under the most stringent stress conditions such as high drive current, temperature cycling and humidity.

2.3 Absorbing versus Transparent Substrate

The GaAs substrate used for the AlInGaP LED absorbs light. Photons generated in the active layer which travel towards the GaAs substrate are absorbed. That is why this technology is called AS-AlInGaP. The efficiency of such an AS-AlInGaP structure is about 10 Lumens / Watt.

With a new process called "wafer bonding" the absorbent GaAs substrate can be replaced by the transparent GaP substrate (see Fig. 3).

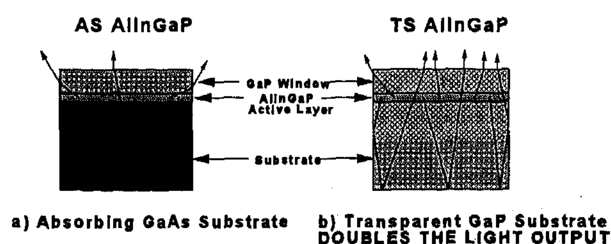


Fig. 3 HP High Performance AlInGaP LED Chip Structures

This technology is called TS-AlInGaP. With the help of a reflective coating on the back of the new GaP substrate all photons are reflected back through the material which increases the efficiency up to 20 Lumen/Watt. This is twice the efficiency of an AS-AlInGaP LED.

Both the AS- or the TS-AlInGaP dies are cubes with an edge length of 0.25 mm and can be built into any existing LED package.

3 LED Design

3.1 Packaging

With the use of silver conductive epoxy, the back of the chip (substrate) is connected to the cathode lead of the leadframe. At the location where the die sits the cathode lead is formed to a little reflector cup to capture the light which emits from the sides of the LED die. The top of the die is metalized and connected to the anode lead of the leadframe via a 25 µm diameter thick golden wire bond. The leadframe is now dipped into a mold cup filled with optical grade epoxy. The shape of the mold, the insertion depth of the leadframe and the reflector cup surrounding the die determine the radiation pattern of the LED. Also the different indices of refraction of the LED die, the epoxy and the ambient air have to be designed into the shape of the cathode reflector cup and the epoxy mold cup.

For a narrow emission angle package more light from the LED die has to be redirected into a narrower solid cone angle. This causes some optical losses. The wider the radiation pattern of the LED package, the more optically efficient it is.

3.2 Design with LEDs

The design objective of an LED signal lamp application is to generate as much flux as possible while taking the maximum operating limits into account. The maximum drive current should not be exceeded because of the current density as described above. The maximum junction temperature of the LED is related to the higher temperature at which the optical grade epoxy starts to expand rapidly and increase the risk of catastrophic failures such as broken wire bonds or lifted LED dies.

The following formula determines the LED junction temperature:

$$T_j = T_a + P_d * R_{th\,(j\text{-}a)}$$

where:

Tj	= junction temperature
Ta	= ambient temperature
Pd	= power dissipation (Vf * If)
Rth(j-a)	= thermal resistance junction to ambient temperature

This formula shows that by reducing the systems thermal resistance the drive current through the LED can be increased while keeping the junction temperature below its maximum. A higher LED drive current does increase the LED light output because the two are proportional to each other.

3.3 High Power Package

With the development of a power LED package the junction-to-LED pin thermal resistance was reduced to 150 °C/W. This power package is flat, which shortens the path for the heat to travel by bringing the LED junction closer to the pcp. The diameter of the leadframe is larger, which gets the heat out of the package more quickly. A new epoxy development has shifted the maximum junction temperature up to 125 °C.

Standard 5 mm LED packages have a junction-to-pin thermal resistance of 250 °C/W and a maximum junction temperature of 110 °C.

Also the power package has been designed for maximum optical performance. The radiation pattern options for this package offer wide viewing angles only to minimize the losses caused by larger angle refractions of the redirected light as described above. Such a power LED can run at 70 mA and with the TS-AlInGaP die it typically generates 2.5 lumens.

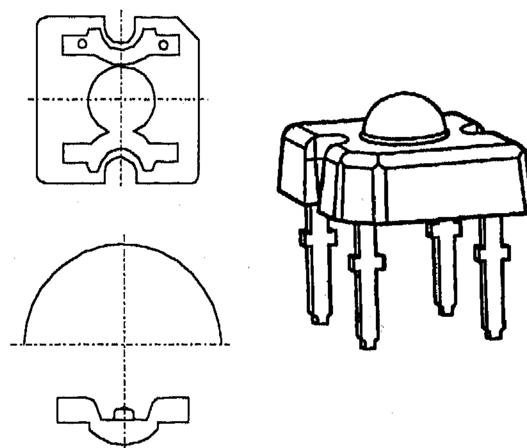


Fig. 4 Outline of Power LED Package

About 12 of these TS-AlInGaP power LEDs generate enough flux to meet the photometric specifications of a Center High Mounted Stop Light. The power consumption of such a unit would be about 5 Watt. under the assumption of typical system thermal properties [Rth(j-a) = 300 °C/W] the temperature rise of the junction above ambient is about 50 °C.

4 Applications

Needing the features "high reliability", "quick response time" and particularly "little mounting depth", the CHMSL was the first application of LED signal lighting.

Naturally this technology is also being investigated for other signal functions. The use of LED's - particularly for complete combination rear lamps - places new requirements on the optical and mechanical layout due to the more complex basic conditions for such lamps:

- The shape of the signal field does not always allow integration of a (rectangular) LED matrix.
- Modern combination-rear-lamp styling frequently requires shapes other than lines and rows; variations in brilliant or high contrast technology are frequently preferred.
- Low-profile versions also have to be curved, because today's combination rear lamps are almost always integrated in freely curved surfaces.

5 Optical Design

Different optical design possibilities result depending on the geometry of the lamps and the desired styling.

5.1 Fresnel Lens Systems

This classic optical system for capturing the light rays emitted in a cone shape from the LEDs and passing them on effectively parallel to one another is always suitable when the luminous surfaces can be composed of a number of individual segments. In addition to square surfaces (see Fig. 5) the overall surface can also be composed of individual hexagonal / trapezoidal or similar surfaces.

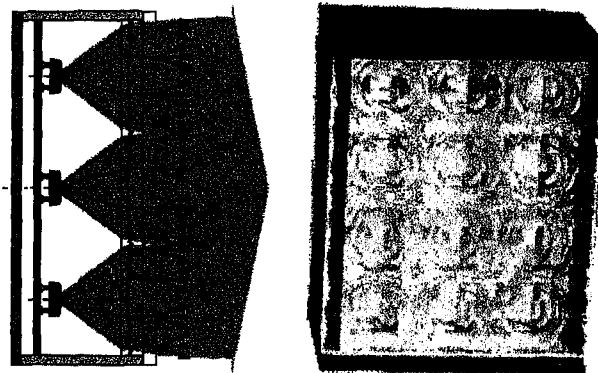


Fig. 5 Fresnel lens optics: Principle and sample

5.2 "Flat Mirror" Optical Systems

Fresnel lens systems require greater overall depths for the illumination of large surfaces. This applies especially to tail lamp functions as the above-mentioned new LED-Technology allows the requested lighting pattern to be generated with only one or two single diodes. The overall depth can be reduced considerably by reflecting the light rays. Turning the combination "LED - Fresnel lens" 90 degrees towards the optical axis of the lamp means that a greater area can be illuminated by the inclined mirror on which the light falls (fitted with prisms-flat mirror) than would be the case if a Fresnel lens surface were used directly. Simultaneously the "spots" of the individual LED's are not as clearly visible. Since the reflecting mirror coating remains visible in the cold state, the system provides a brilliant impression (see Fig. 6).

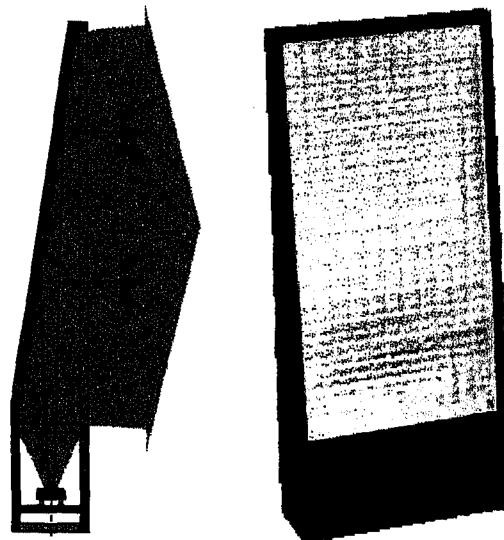


Fig. 6 LED flat mirror system / schematic drawing and example

5.3 Light Guide Technology

Light guide technology can be used for particularly flat versions. The light is carried by total reflection after entering a transparent material. As these light guides can be formed out of thin plastic materials and can be provided with reflecting elements (e. g. prisms) these designs allow the flattest versions of all.

Contrary to incandescent bulbs the "cold" LED's can be installed directly next to the light guide, eliminating increased overall depths and volumes for the light source and input segment. Figure 7 shows an example for a flat taillamp.

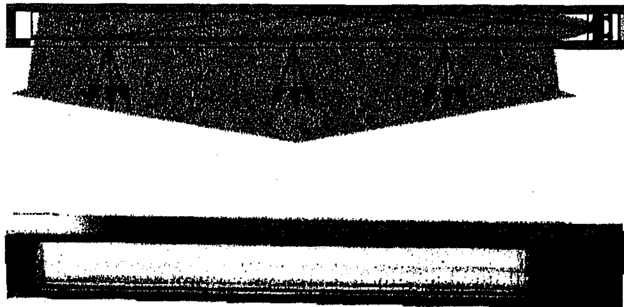


Fig. 7 Light guide taillamp / schematic drawing and example

6 Connection Technologies

Selection of the technology depends considerably on the basic geometry of the lamp. The possibilities are evaluated below:

6.1 Circuit Board Installation

This classic solution for mounting electronic components also provide an excellent solution for the LEDs in the safety brake lights mentioned at the beginning. Since the circuit board is a flat component, curved lamp contours require that the LEDs be arranged in terraces on individual circuit boards connected by wire bridges at equal distances from the lens. These limitations can be eliminated to a certain extent by the use of flexible circuit boards.

6.2 Snap-LED's

This connection technology developed by Hewlett Packard fixes the LEDs with a type of rivet within a stamped grid. Certain sections of the grid can be shaped (see Fig 8) to allow terrace-shaped structures by means of deformation. This connection technology appears to be the least expensive alternative for complex 3D structures.

Another advantage of this principle is that the LEDs are not subjected to any type of thermal load during processing. Conduction of the heat away from the semiconductors can be improved and the danger of cold solder points is eliminated.

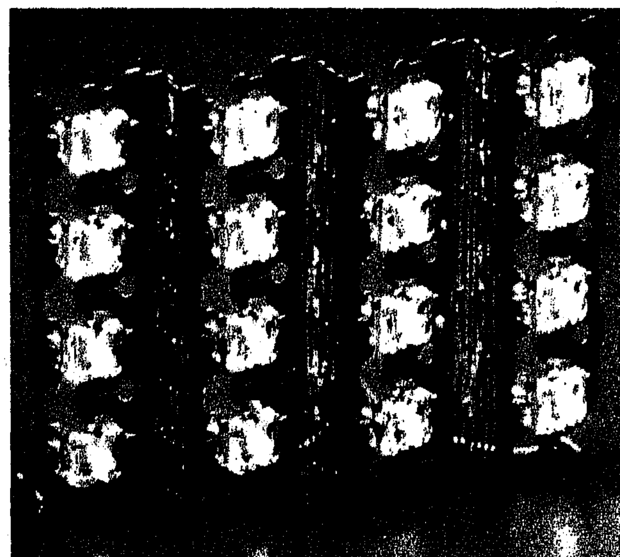


Fig. 8 Application example for SNAP-LEDs on a 3D carrier

6.3 Moulded Interconnected Devices

Moulded Interconnected Device Technology (MID) produces conductive connections by means of galvanized conductors on complex three-dimensional plastic surfaces. The LED's can be connected by soldering. In addition to automation, the many process steps still required for production of the circuit board represent a major obstruction for this technology.

6.4 Alternative Concepts

Dependent on the basic geometry of the lamps and the type of diode used, other alternatives are imaginable. One possibility here is wire connection technologies using industrial robots. The connections could be completed for example by using snap-lock connectors.

7 Prospects

Looking at the LED itself one of the major tasks for the future will remain to increase its light output further. With a reduced number of required elements the competitiveness of LED-lamps will grow even more.

Apartly from this the challenge to achieve the white color for the signal functions "Position", "Reverse" and "Daytime Running Light" will so remain. Attaining better efficiencies for the green and blue color will enable this goal be realized by mixing red, green and blue by color addition to "white".

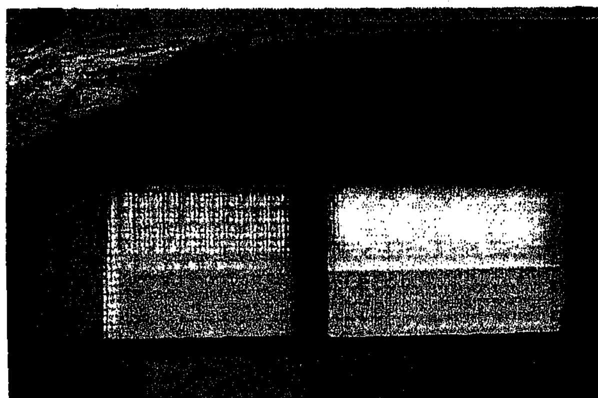


Fig. 9 Prototype of a combination rear lamp

The possibilities for application will increase, too. In addition to the mentioned use for complete combination rear lamps shown as a prototype in figure 9, use in side marker lamps as well as additional side repeater flasher lamps will provide the next applications for LEDs. Since the high reliability of LEDs offers particular advantages in the commercial vehicle sector, the next fields of application for LED signal lamps will be found here.

LED technology can also be used to great advantage for new concepts for the rear signal appearance on motor vehicles, because the size of the signal field as well as the overall brightness can be varied just as easily as can, for example, the light-up frequency using LEDs with one simple electrical control. This ensures that the control or adaptation of future signal appearance patterns on motor vehicles can be implemented simply in terms of optical technology.

LEDs can be used for many signal functions - as shown - even before the registration codes allow implementation of such visions. It will soon be necessary to examine new lamps more carefully to determine whether incandescent bulbs need be used at all.

970652

LED Automotive Rear Combination Lamp (RCL) Dual Level Driver Electronics

Derek S. Mallory
North American Lighting, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

For the current class of available automotive LEDs (Light Emitting Diodes) used in CHMSLs (Center High Mount Stop Lamps), voltage regulation is required for operation at an optimized output light flux. We have built and measured the performance of resistive, voltage reference, and pulse width modulation DC-DC voltage regulators in order to achieve the required dual level control for RCL (rear combination lamps). These three types of driver electronics for dual level operation of a RCL are compared for performance, and efficiency under various operational conditions.

INTRODUCTION

Current automotive circuitry achieves the required voltage level with passive resistors on the boards located near the LEDs. A typical high output LED is operated at 2.5 V compared to 12.8 V supplied from the battery. With an array of four or three LEDs in series, 22% to 33% of the power is wasted as heat in the voltage divider resistance. The additional heat generated on the circuit board from this voltage divider reduces light output for the LED array.[1] This is of major importance since photometric requirements are tested after having the LED lamp on for 30 minutes. A 30% decrease in light output due to heating of the LED/circuit board is not unusual over this time.

The use of LEDs in a RCL will require dual level operation for stop/tail functions which can be easily achieved by an external driver electronics module. Removing the on board regulation circuitry to an external driver control unit will reduce heat in the lamp package, and improve performance of the LED lamps. Additionally, different DC-DC voltage regulation techniques can be utilized to further increase the operational efficiency of the driver electronics.

MAIN SECTION

PHOTOMETRICS - For dual level control to be considered, it is first useful to understand the motivation behind the need for dual level voltage regulation. SAE J585 and SAE J586 address tail and stop lamp photometric requirements respectively.[2] These typically assume a dual filament bulb to meet the requirements. SAE J586 sets the minimum intensity ratio for stop (high) operation to tail (low) operation at 5:1. Table 1 shows a comparison of maximum and minimum allowed values at H-V for stop and tail operation.

Compartments	1	2	3
Stop max (min)	300 (80)	360 (95)	420 (100)
Tail max (min)	18 (2.0)	20 (3.5)	25 (5.0)

Table 1. Stop and tail requirement comparison at H-V test point with maximum and minimum requirements in candela. A 5:1 minimum intensity difference is required.

The 5:1 intensity difference would correspond to minimum stop lamp intensity and maximum tail lamp intensity at H-V. It is of note that a standard dual filament bulb accomplishes this ratio with a 7:1 intensity difference between stop and tail.[3] Additionally since the tail filament is usually defocused ratios of ~9:1 are achieved in practice. Thus, a more reasonable intensity ratio for an array of LEDs would be in the range of 7:1 or more for design purposes. With LEDs this can be achieved by voltage regulation, or current regulation. For the purposes of this paper, we only consider voltage regulation as a possible means to achieve the dual level light output.

VOLTAGE REGULATION SUBCIRCUITS - The following is an introduction to DC-DC voltage regulation circuitry that would be useable to drive LED arrays for RCL applications. For our design purposes, we considered two different arrays of 48 LEDs. The two parallel arrays

were accomplished by **1)** 16 rows of 3, and **2)** 12 rows of 4 LEDs in series. For the purposes of this comparison, high operation for the individual LEDs was at 2.5 V @ ~70 mA, and low level operation was at 1.9 V @ ~30 mA. Table 2 shows the complete operational conditions required by the LEDs.

Mode	16 rows of 3	12 rows of 4
high (8.40 W)	7.5 V @ 1.12 A	10.0 V @ 0.84 A
low (2.74 W)	5.7 V @ 0.48 A	7.6 V @ 0.36 A

Table 2. Typical operating voltages for one stop/turn/tail LED array with total current.

RESISTIVE CONTROL - Figure 1a) is an example of a resistive network that can be utilized for dual level LED control circuitry. Currently, this is the standard arrangement for single level operation of CHMSLs. It is recommended to operate the LEDs in arrays of three or four in series.[4] Since the LEDs typically cannot handle large reverse voltages or currents, a protection diode is added to the array.

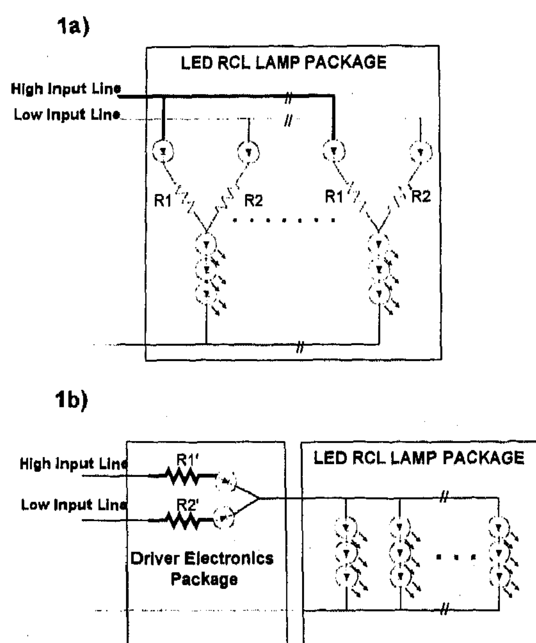


Figure 1. a) An example of a possible on-board resistive network for dual level control of an LED array. b) Shows a schematic of the driver electronics using power resistors R1' and R2' for dual operation.

Another resistive option is to operate the array with a single power resistor voltage divider as in figure 1b). This circuit could be placed in a control module that is attached to the lamp housing. This has the advantage of removing the resistive heating from the local board area near the LEDs. However, the large wattage power resistors, and high current diodes required are not the most power efficient options available as will be seen.

All resistive control methods have potential drawbacks. During operation at higher or lower voltages than design, the LEDs will be run at incorrect voltages and light output levels. Large voltage spikes and EMI will not be filtered from the LEDs. Large wattage power resistors add considerable weight compared to other control methods.

VOLTAGE REFERENCE - Figure 2 is an example of a Darlington voltage feedback circuit for DC-DC operation. This is a standard method widely used in automotive circuitry to provide reference voltages. The output voltage is determined by the voltage divider. This fixes the transistor into saturation and the output V_e is determined by the voltage drop across V_{be} (which is typically 0.6 V). Thus,

1) $$V_e = V_b - 0.6\ \text{V}.$$

The Darlington voltage reference will provide a fixed drive voltage at reasonable efficiency if operated in saturation mode. The efficiency depends upon the current gain of the transistor used. The current gain of a typical medium power output Darlington is low (10-20) when driven in saturation for high output current.[5] Thus to drive an input load of 1.5 A, ~100 mA injection current will be required. The lower the current gain of the transistor, the higher the losses in the voltage divider and Zener diode reference input to the transistor base V_b.

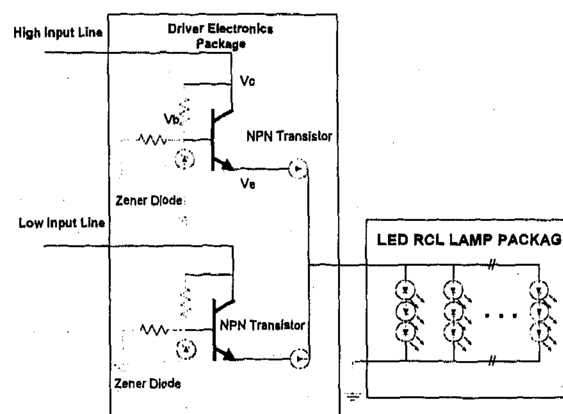


Figure 2. Example of dual level Darlington voltage references for driving a LED RCL lamp.

For operation at higher input voltages than originally designed, the Zener diode will clamp the base voltage to the required output level. However, when operated at lower than design input voltages, the output voltage to the array will fall linearly with the base voltage divider circuit. Efficiency can also be lost at low input voltages due to the transistor dropping out of saturation mode.

This DC-DC voltage regulator will filter some high frequency voltage spikes from the LED array, due to the transistor switching times. But, additional filtering would be required to completely isolate the array from electromagnetic susceptibility on the input line. Tradeoff considerations for output voltage vs. input voltage would also need to be made when selecting the reference Zener diode values.

PULSE WIDTH MODULATION (PWM) - Figure 3 is an example of a pulse width modulation circuit for dual level voltage regulation with voltage feedback. Two options exist for using the PWM circuitry. The first is to use rectification or wave shaping on the output. This will limit the efficiency of the overall circuit depending upon the rectification circuitry. The advantage is it will supply a filtered DC voltage for which the LED light output is fully characterized, which may help meet EMC requirements.

The second method for PWM operation is to drive the LEDs with no rectification. This is the most efficient method power wise to operate a LED array. A major consideration for using unfiltered square wave output is the possibility of EMI problems (Electromagnetic Emissions). Typical drive frequencies are in the range of 60 Hz to 400 Hz.

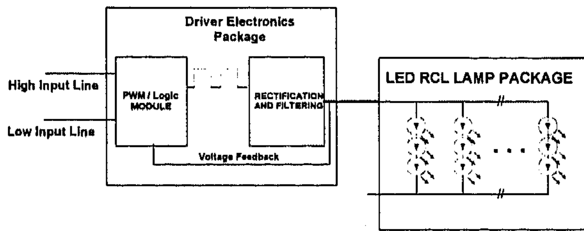


Figure 3. Pulse width modulation with rectification for dual level LED voltage regulation in RCL applications.

An additional advantage of this more complicated system is the operation at over voltage and reduced voltage. Light output (and power supplied) can be feedback controlled by adjusting the duty cycle of the LEDs,[6] effectively allowing constant controlled light output over a large input voltage range from 9 to 18 V.

ELECTRICAL EFFICIENCY - Table 3 shows the relative efficiency comparisons between the three main types of driving circuitry we constructed. The comparisons are for high and low operation of the driver circuitry with the LEDs in rows of three in series. The power dissipated by the LED array in low and high operation was given in table 2.

For the driver circuit modules above, the heat from the losses is separated from the LED array circuit board. It is apparent that the PWM method for voltage regulation is the most efficient. However, PWM is the highest cost system and will have possible EMC problems. The greatest energy savings are achieved during high mode operation, which would be the stop/turn signal functions. This high efficiency translates into weight savings as well, since a smaller heat sink would be required for the control circuit.

Driver Circuit Type	3 LED Array
Resistive Low Mode	3.4 W
Resistive High Mode	5.9 W
Darlington Low Mode	1.5 W
Darlington High Mode	3.5 W
PWM Low Mode	1.2 W
PWM High Mode	2.1 W

Table 3. Comparison of driver circuit wattage required for operation of a 3 LED in series array. Total power dissipated by the RCL lamp is the sum of driver circuitry plus LED array given in table 2.

THERMAL CONSIDERATIONS - When designing LED rear lamps, performance at elevated temperatures must be considered. The junction temperature of an LED needs to be kept below 125° C to insure no failures. These failures are typically caused by the wirebond becoming disconnected from the LED die when the plastic around the die passes through the liquid/solid glass transition at 125° C. Design of the circuit board, location, venting of the lamp, external temperature, and operational voltage all play important roles in determining the final junction to ambient thermal resistance.

For the dual level control methods mentioned in this paper,

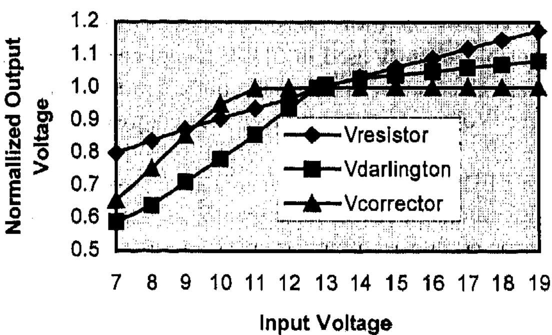


Figure 4. Experimentally measured normalized output voltage for the three types of regulation circuits driving 3 LEDs in series. Operation mode is for "high" output of 7.5 V across 3 LEDs at 12.8 V input.

operational voltage is what can be controlled. Figure 4 shows the normalized measured output voltage for input voltages varying from 7 V to 19 V for the three different dual level control methods. In figure 4, Vresistor is using the power resistors shown in figure 1. Vdarlington is using the Zener diode driven voltage regulator from figure 2. Vcorrector is using a voltage feedback circuit similar to the PWM module in figure 3.

Figure 4 further highlights the problems that can occur if LED lamps are driven above the designed voltage of 12.8 V. If resistors or the Darlington voltage reference circuitry have been used, additional heating will occur in the LEDs from the higher drive current. Only control circuitry with feedback will limit the maximum drive voltage for the LEDs, thus eliminating additional thermal factors. It is not a stretch of the imagination to envision dual level control modules that are temperature compensated to further reduce LED drive voltage at elevated temperatures. Further study of total lamp system thermal and photometric interactions is needed to propose circuitry to effectively handle these concerns.

CONCLUSION

The LED systems that are employed in future automotive applications will need to be designed cost effectively with an understanding for the optical, photometric, thermal, and overall system performance interactions. The demands for better fuel economy and electric cars will continue to motivate that the most efficient electrical voltage regulation and lighting be employed. LEDs are an excellent consideration for this task, and will need dual level drive circuitry to meet these requirements.

A voltage regulation driver circuit can be used to operate a LED array for dual level RCL applications. The lowest power consumption option is accomplished by a PWM DC-DC voltage regulator. However, PWM modules cost more and have possible EMC considerations. Resistive methods of control currently used in CHMSL applications are not as efficient electrically or thermally but are much simpler to design. Voltage limiting by feedback control may be needed for LEDs because of thermal considerations. For future LED RCL lamps, a case by case approach to these considerations needs to be approached when selecting the optimum dual level drive system.

ACKNOWLEDGMENTS

I would like to acknowledge the efforts of Mr. Benjamin Strauss of North American Lighting Inc. for assisting with construction of the LED lamp hardware and circuitry. Also, Ms. Robin Ray from Hella Electronics for discussions on electrical requirements and circuitry.

REFERENCES

1. LED Lamp Thermal Properties, Publication No. 5964-8921E, Hewlett-Packard, 1996.

2. SAE Ground Vehicle Lighting Standards Manual, Published by: Society of Automotive Engineers, Inc., 400 Commonwealth Drive, Warrendale, PA, 1995. pp. 115-117.

3. Miniature and Sealed Beam Lamp Catalog, GE Lighting, Nela Park, Cleveland, OH. 1995

4. Private Communication with Mr. Jim Loeffler, Hewlett-Packard.

5. Motorola Bipolar Power Device Data Book, Motorola Literature Distribution, P.O. Box 20912 Phoenix, AZ, 1995, pp. 3-188.

6. "Pulsed mode vs. DC operation," From: Operational Considerations for LED Lamps and Display Devices, Publication No. 5091-90704E, Hewlett-Packard, 1996, pp. 19-22.

970653

Innovations in Low Pressure Discharge Lighting

George J. English and Harold L. Rothwell, Jr.
Osram Sylvania Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The automotive lighting industry has undergone major changes over the past 25 years. The sealed beam lamp, the industry standard for half a century has been replaced by a number of new aerodynamic headlamps. Over this same time period signal lighting has undergone dramatic changes; from a standard hand signaling for brake and turn, to complex multi-cavity systems for running, turn, and stop lighting. In the interior of the vehicle similar changes have also occurred which have dramatically improved driver visibility of the instrument cluster. In all three application areas major changes are continuing to occur particularly in the light source and optics of the illumination system. This paper discusses some of these changes and compares the various light sources and optical systems used to illuminate the space around the automobile and the communication between vehicles. While the forward illumination has been the dominant element in lighting, signaling to improve accident avoidance is becoming of equal importance. The central theme of this paper is to provide information about the various sources with a particular emphasis on the neon discharge light source and its auxiliary system components that can expand our communication ability.

INTRODUCTION

The automotive lighting industry has undergone major changes over the past 25 years. The sealed beam lamp, the industry standard for half a century, has been replaced by a number of new aerodynamic headlamps. The gas-filled tungsten incandescent evolved to the hard glass halogen. Today, High Intensity Discharge (HID) sources are being used on several vehicles both in the United States and Europe. Signal lighting has undergone a dramatic change from the days when hand signaling was standard for brake and turn with a simple red tail marker lamp to multi-cavity systems and now Light Emitting Diode (LED) and Neon Low Pressure Discharge (LPD) lamps are available. In the interior of the vehicle similar changes have also occurred. Simple gas filled incandescent lamps have been replaced with multi-colored edge-light cluster lighting in addition to LED and Sub-Miniature Fluorescent (SMFL) lighting techniques. In all three application areas major changes are continuing to occur particularly in the light source and optics of the illumination system.

While forward illumination has been the dominant element in lighting, signaling to improve accident avoidance is becoming of equal importance. However, it should be noted that the requirements for signaling as well as for forward lighting have historically been developed based on criteria involving human interaction with the performance properties of incandescent source and optics. In general, the evolution of new light sources and optics produce different performance characteristics. These differences must be continuously evaluated during and after the development process to ensure that they provide improvements to the intended safety requirements derived for the incandescent function.

LIGHT SOURCES FOR ROAD ILLUMINATION AND SIGNALING

FORWARD LIGHTING - The simple glass headlamp that served so well for so long in the United States was an incandescent filament mounted in a parabolic reflector with a hermetically flame-sealed lens to ensure an inert atmosphere. The reflector collimated the light from the tungsten wire source and a glass spread lens was used to distribute the light to meet the required beam pattern. In the mid-seventies, hard glass tungsten halogen capsules replaced the incandescent filament source and the bonded beam replaced the flame-sealed glass lamp. Plastic optics replaced the glass leading ultimately to the aerodynamic headlamp with replaceable bulbs (9004, 9005/6 and 9007). More recently, HID sources have found their way into headlamp applications. [1] The blue-white light of the source interacts favorably with road markers, signs and paint. [2] Under wet driving conditions this source enhances illuminated surfaces, providing a visual aid to drivers. The development of improved low glare optics using these sources will allow older people to extend their driving years by keeping them more mobile during evening hours.

SIGNAL LIGHTING - Even though signal lamps have traditionally been incandescent vast improvements in optics and system designs on the rear of the vehicle have occurred. Automotive lighting engineers are providing illuminated areas to outline vehicle shape, and provide direction and braking information. Rear-end collisions have been reduced by the addition of brake/turn/tail lamps and reflex material for passive identification. As a further improvement, the third brake lamp, the Center High Mounted Stop Lamp (CHMSL)

has been added for transmitting vehicle de-acceleration information.

Incandescent signal light sources have changed from a metal based bayonet incandescent to hard glass tungsten halogen wedge base and S-8 style clear and amber bulbs. The latter sources have greatly improved bulb alignment, reduced contact problems and are becoming standard with some automotive manufacturers. During this evolution, optics for these applications have transformed from simple round three- or four-inch diameter parabolic reflector lamps with a plastic pillow colored lens to large red lens systems with multiple incandescent sources. As insert and multi-color molding technology evolved, single piece lens packages with different lighting functions (color filters) have become dominant. High contrast luminaries have been introduced to reduce the glare effects produced by the sun from these lamps. As with forward lighting new sources are beginning to replace the incandescent signaling lamp. For instance, LED's and LPD lamps were introduced into CHMSL's. [3]

HID LIGHT SOURCES - At least in a general way the HID and LPD sources are similar because they both are plasma discharges and they both need ballasting. The operating pressures, electrodes, arc gaps, and lamp chemistry are, however, significantly different. Briefly, the HID source is a high luminance device ideally suited for focusing-type optics that are used for illuminating subjects such as roads and pedestrians. However during a cold start, the initial output radiation is limited to xenon and some mercury. After approximately one second the mercury pressure builds, providing a mercury spectrum which is visually bluish-white. Within five seconds the discharge becomes visually white as the contribution from sodium and scandium metals produce additional emission. The discharge processes that provide emission of light are driven by initial high power levels (75 watt for the steady state 35 watt HID source) applied to the arc tube during the start-up phase. Within approximately ten seconds the heat provided by the applied power provides a gas mix of metal halides such as scandium iodide, and sodium iodide. Thus, the color chromaticity and luminance are varying until the bulb wall steady state temperature is achieved after several minutes.

SMFL LIGHT SOURCES - Low power narrow diameter fluorescent lamps can be used for interior lighting applications. Typically instrument cluster back lighting and general lighting with a dome fixture are good examples. The efficacy of the SMFL lamp is quite good, ranging above 35LPW. The lighting process involves a low pressure discharge through mercury gas which emits strongly in the ultraviolet (UV). A phosphor material, coated on the inside of the bulb, absorbs the UV and converts the energy into longer visible wavelengths. Unfortunately any mercury-based discharge depends on the volatility of mercury. Particularly for fluorescent lamps the pressure of mercury is very important. At temperatures below 0^0Centigrade(C) the fluorescent lamp does not work well because the vapor pressure of the mercury is low. Typically electrical heaters are attached to the outside of the SMFL tube to heat up the mercury. On the other hand, as the temperature increases above approximately 30^0C the increase in the vapor pressure of mercury causes the UV to be self absorbed and results in a reduction in efficacy.

LPD LIGHT SOURCES - The LPD, on the other hand, achieves its spectral characteristics nearly instantaneously. The bulb chemistry is simple in comparison to the HID source. Its electromagnetic spectrum is determined by the excitation of atomic species in the gas phase. For automotive applications neon gas, which emits a reddish spectrum, is the most useful LPD source. However, even for pure neon the emission spectrum will depend upon the applied power characteristics as will be discussed later in this paper.

The LPD sources provide much faster response time than the incandescent filament because the tungsten wire must be heated to incandescence. Wire heating is achieved by passing an electric current through it. As the power is increased the filament's resistance turns the electrical energy into heat, thus raising its temperature. At about 1000°Kelvin(K) the radiation from the wire begins to shift from the infrared into the visible portion of the electromagnetic spectrum. As the filament temperature increases, a higher percentage of radiation is formed in the visible region of the spectrum. The filament size, shape, wire size, applied voltage, and power of the light source are related to its steady state radiating level. The resultant rise and decay time of the light output is on the order of hundreds of milliseconds. On the other hand, the low pressure gas discharge of neon are atomic radiators and can produce light significantly faster.

The neon gaseous discharge is less dependent upon the temperature of the environment, and more on its own internal continuous operating temperature. The steady state temperature in typical luminaries is reached in about 20 minutes, with a light output reduction of about 10 to 15 percent comparable to effects found with LED's. It should also be noted that the incandescent source is usually designed to operate at or near its operational limits to provide high luminance and efficacy for a desired life requirement. Therefore, the applied voltage variations to the source affect its output and life expectancy. The current state of voltage regulation found within most vehicles permits significant variation in lamp life and output. HID and LPD sources can be regulated within a few percent over an input voltage range from 8 to 18 volts. This helps maintain a regulated output and protect the source.

LED LIGHT SOURCES - The LED has also found applications in automotive. The red LED is used primarily for CHMSL units. Typically twenty or more individual diodes are clustered or arrayed in order to achieve acceptable light levels. The LED also has a fast rise time since the emission is due to solid state "gap" transition. Barrier junction noise is

dependent on the heat level which contributes to a light output reduction at elevated temperatures.

Because of heat transfer limitations the LED power levels are limited to about 0.1 watt. In spite of today's most advanced fabricating techniques heat sinking is required when large clusters of LED's are used for continuous operation. This sets the appearance characteristic of the LED lighting module as an array of point sources. The high luminance LED source is highly directional because of the fabrication technique of a plastic collection lens over the solid state component. This highly directional feature can present two problems that need to be solved in packaging of these small sources. First, the mounting of the LED needs to be controlled precisely or the array will have an element that appears to have burned out at certain viewing angles. Second, to create enough spread for the signal light pattern, the peak intensity pattern of the individual LED's needs to be spread to achieve off-axis coverage for the lighting function. In order to achieve this condition more LED wattage will be needed. This will require more diodes because the light per LED is limited, about fifty diodes would be required to perform the turn or brake function. This number of diodes with the most advanced technology will produce luminaries with a blotchy lit appearance, unless large optical packages or highly diffuse one are utilized. The latter indicated method will require a larger population of diodes to meet photometric requirements.

INCANDESCENT (TUNGSTEN) LIGHT SOURCES - The tungsten source for rear vehicle lighting is of low wattage and compact size and in general is placed in optical packages with a great deal of light spreading capabilities to form the desired lighting pattern and appearance. The reflectors are generally deep and highly diffuse in comparison to the specular forward headlamps. This is done to reduce the impact of bulb size, position, and tolerance limits, of multi-functional bulbs, etc.. The tungsten source is filtered by the use of the colored lens. Presented in Figure 1a is the CIE Chromaticity Diagram indicating the color shift necessary to transform a 2600°K incandescent source, in this case a tail lamp of the wedge or S-8 design, to the recommended red signal region. About two-thirds of the visible light must be absorbed by a red lens filter material in order to achieve the desired effect. Thus, about 3% of the lamp wattage is converted into useful red light. The remainder goes into heat.

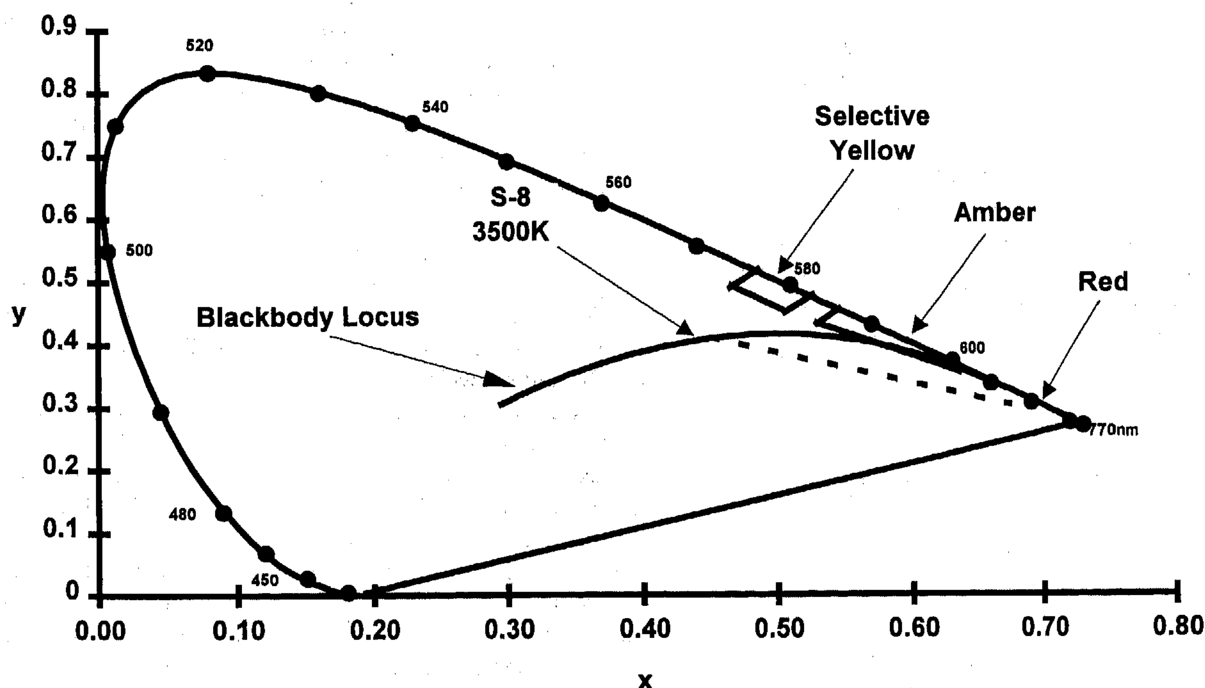
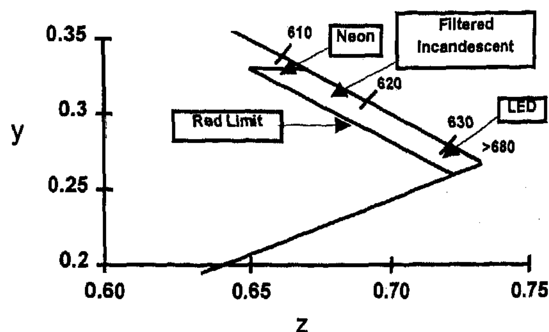


Figure 1a. - CIE Chromaticity Diagram indicating color areas. Dotted line represents color shift of incandescent through a red filter.

An exploded view of the red corner of the CIE is shown in Figure 1b. The positions corrsponding to the various sources are indicated. The red limit box is the specification for red light as defined in the SAE Signal and Lighting Handbook.



Figures 1b - Transformation of white light from S-8 absorption of lenses (approximately 2/3 of visible light energy is lost).

SPECTRAL PROPERTIES OF THE LIGHT SOURCES - The LED generates light in a narrow wavelength band and thus is highly efficient, unlike the continuous broad electromagnetic spectrum of the incandescent source or the multiple line neon spectrum (see Figure 2). A comparison of the typical visual radiating properties of the three sources is shown in Table 1. In the past, it has been difficult to make a hybrid system using LED's and tungsten with good color matching characteristics. The LED source could not be filtered to bring it into the region of the chromaticity diagram where the red filtered incandescent was located. The basic problem is the inherent narrow spectral band of radiation at about 650nanometers(nm). In order to achieve color matching between the two sources the incandescent source, which emits in a continuous broad band, had to be filtered further into the red region. This required higher bulb wattage's to provide sufficient intensities for various lighting functions. More recently new LED sources have been developed that radiate closer to the filtered incandescent chromaticity.

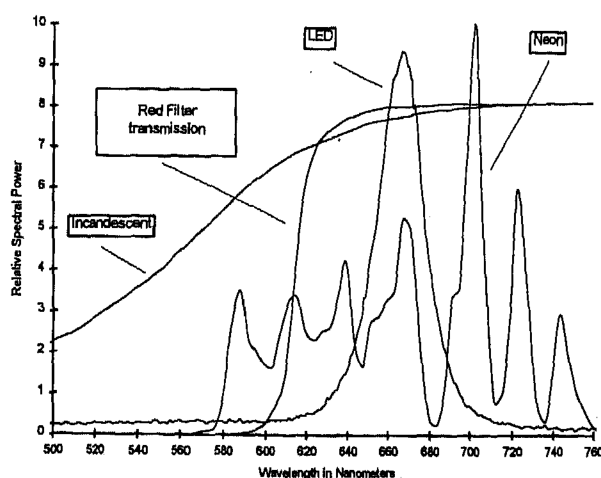


Figure 2. - Spectral power distribution of Incandescent, Filtered Incandescent, Neon and LED sources.

The neon source, while not a continuous spectrum radiator, has a series of atomic emission lines that generate light from the yellow-green to the deep red (between 580 and 750nm). A small amount of filtering between 580 and 620nm can shift the system deeper into the red to match any filtered incandescent source. Any filtering to the red will reduce the lumen value of the neon radiation because the eye responds more to the yellow-green than to the deeper red stimulus. The filtering can reduce the effective light of the neon radiation by about one quarter. Note that the unfiltered red neon spectrum meets the SAE color specification and that a crystal clear colorless lens could be used in this case, minimizing the amount of wasted wattage.

SIGNALING CHARACTERISTICS

Signaling with the gaseous type of sources (also LED's) allows for a number of advantages not available today with the current tungsten or mercury-base sources. Even with the advent of radar, sonar, and other navigation aids, mariners and air craft pilots still depend on visual aids. Communication by means of colored and white lights are readily accepted by humans. Information can be superimposed within the field of view of an observer, and can be easily utilized while monitoring driving conditions and the surroundings. The improved use of signal lamps and reflex materials/or signs and sign illumination has reduced fatalities over the past several years. To date, with the exception of xenon flash lamps, communication between two vehicles has been performed by the use of tungsten sources or horns.

An incandescent source needs a finite amount of time to build up to maximum outputs on the order of two to three hundred milliseconds. The tungsten decay time is less than, but similar to, that of its rise time. This limits the number of on-off lighting cycles per time interval, limiting the frequency of pulses for signaling lights. Thus, tungsten can only provide strobing information at about 1 to 2 hz rate. While the eye is limited by its critical fusion frequency to about 10 hz for a highly distinctive effect, a person can typically detect light modulation to about 50 hz. The voltage applied to the bulb will have a significant effect on this rise time as well as on the maximum steady state output available from the source. After turn-on, the light increases approximately exponentially with time then approaches the steady state value asymptotically. Figure 3 and Table 2 show time to 90% light output as a function of voltage. The light output of an incandescent source and its spectrum are temperature dependent which in turn, depends on the percentage of its rated applied power. A ten percent reduction in voltage reduces maximum output by about 30%, or a 10% increase can cause a 70% percent reduction in life.

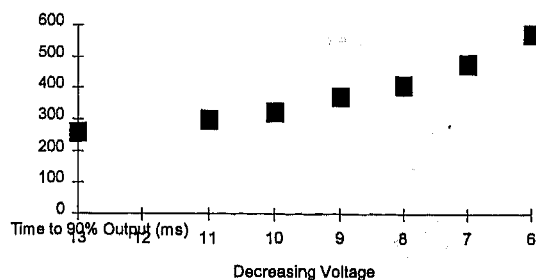


Figure 3 - Delay in Reaching 90% Level of Steady State due to Voltage Reduction in Incandescent Light Sources

The neon source can generate light nearly instantaneously by the ballast once a threshold voltage is applied to the ballast. This instantaneous rise time aids in the recognition of the warning as noted by M. Sivak. [4] The improvement translates into a significant margin of safety in bringing the vehicle to a stop (shown for various speeds in Table 2). The mean reaction time for several sources is shown in Figure 4.

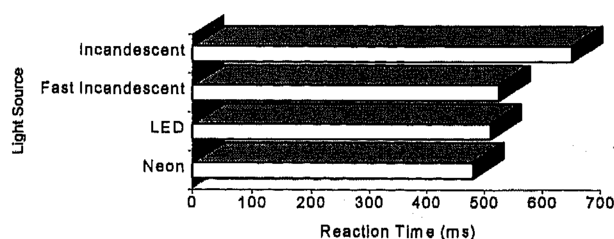


Figure 4. - Mean reaction time by light source

Signal recognition can be enhanced by several methods with the LPD. First, the intensity can be increased. However, limits are quickly reached because of glare issues. Strobing or pulsing will also increase conspicuity but this causes an annoyance after several cycles. Switching between two levels, however, softens the disturbance. Lastly, the rise time of the light source can aid the recognition process. As stated by Sivak:

> "In other words, under the conditions of the present study, the decreases in rise time from that of the standard incandescent lamp to those of either the fast incandescent, LED, or neon lamps had larger effects on reaction time than an eight-fold change in luminous intensity." [5]

Therefore, in order to achieve better recognition of brake signaling the addition of several pulses of light switching from high to low before the constant high level red intensity could be effective. In addition to helping improve braking and turning information, the neon source can conserve energy and improve visual recognition of a disabled vehicle by providing a flashing warning distress signal. The hazard signal lamp function can be achieved with the neon source and ballast very efficiently. A significant amount of the flashing tungsten source energy is used to drive the source through the heating stage. Also, if the vehicle is having battery problems the peak voltage available for the flash mode will be decreasing. By contrast the ballast low operating point could be set at 8-volts or some low voltage before a significant decrease is allowed to occur in lamp wattage. The neon source could provide a significant pulse of light, when the tungsten lamp would be putting out only a small percentage of its rated light.

OPTICS DESIGN CONCEPTS

TROUGH - Both double and single filament incandescent lamps are found in multi-signal functions such as tail, brake and turn applications. The diffuse reflectors and high spread lenses are mainly used to provide a uniform appearance to reduce glare while meeting photometric intensity requirements. Over the past few years both a gem or jeweled and a uniform smooth appearance have been used for vehicle styling identification. At night, these uniformly lit packages are desirable and considered important because they are viewed for long periods of time when vehicles are traveling in the same direction. Because of their light emitting characteristics, the LED and tungsten sources require a significant amount of spread to provide this uniformity at visually accepted levels. Many small point sources in deep optical packages are used to achieve this uniform redistribution of light.

For a given length of tube, the neon source luminance can be increased (or decreased) by adjusting the current density in the discharge column. The current variation and therefore the luminance is limited by the lamp electrodes (also tube bore size, gas pressure) and ballast characteristics. The intensity is uniformly distributed along the tube and the total light output is then directly proportional to its length for a given current density. Optical packages can be fairly simple since the light source can be treated as a line source. As an example, an illuminated reflector seen from various points is depicted in Figure 5. The flashed area is a measure of the intensity at each observation point. On axis at point A, an observer sees the reflector uniformly illuminated or fully flashed. This is the maximum or peak intensity level of the system. When viewing the reflector from other locations in the plane, containing the lamp and the parabolic axes, the projected reflector area is reduced. This foreshortening of the optic system reduces the intensity. The reduction will continue until the combination is viewed end on at location B. The observer at point C sees the specular reflector dark, with light only being provided by the direct view of the linear light source. Light can be redirected to this location by the introduction of a spread lens, spread elements in the reflector surface, or by changing the specularity of the reflector surface. Intensity in this direction also may be increased by increasing the source diameter. Peak intensity values over 100 candelas

are readily achievable with specular reflector surfaces, and the neon LPD lamp.

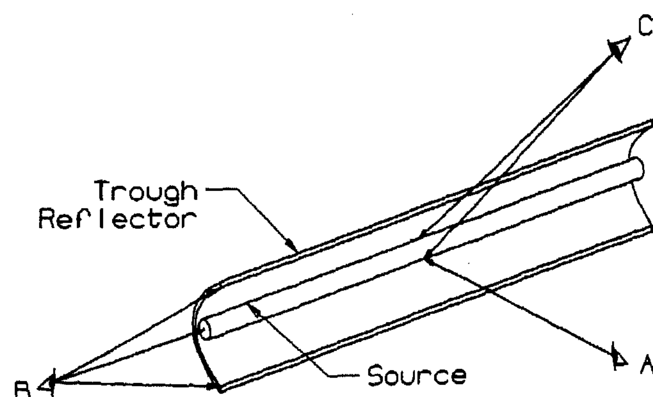


Figure 5. - Primary viewing locations of trough cylindrical reflector with LPD source.

The luminance of a source currently being used in the Ford Explorer CHMSL is about 17,000 cd/m^2 for 28-watts of power and an arc tube length of 0.7m. This signal light system with red lens and EMI protection produces a tail function and a brake function with 150 candela, and a system luminance of over 9,000 cd/m^2. This value is comparable to many lamps utilizing incandescent light sources. The optics to provide the desired photometric performance with the neon LPD also present the lighting designer and stylist with new opportunities. These cylindrical optics with linear sources have lighting distributions that differ from the LED and incandescent systems. A compact source illuminating a parabolic reflector of revolution provides a very symmetrical beam with minor irregularities due to the source shape. On the other hand a cylindrical reflector optic provides good horizontal illumination and conserves the light vertically. With a smooth, highly specular, parabolic reflector the vertical distribution is a function of source diameter and reflector focal length. Further spread in the reflector can be achieved by the use of faceting, spread elements such as grooves, changes in surface finish, or by changing the reflector cross-section shape. The spread function can provide both horizontal and vertical components to the beam pattern. Spread control also can be achieved by lens elements in the cover.

Fresnel reflectors can be used to reduce package depth as shown in Figure 6. The discharge source in this case is located at the top of the lighting fixture. When power is applied to the luminaire, the entire surface becomes flashed in certain directions. The finish of the reflector could be highly specular or partially diffuse. The lens could be red or clear, and may have patterns on it.

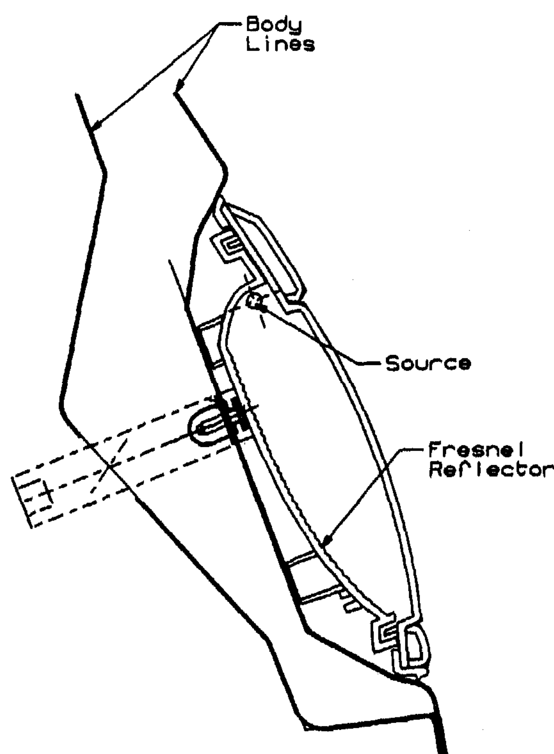


Figure 6. - Compact Fresnel reflector with LPD to reduce packing space and sheet metal deformation.

APERTURE - The luminance of a gas discharge source confined to glass cavities of tubular geometry can be further improved by the addition of a reflective coating (See Figure 7) applied either externally or internally to the arc tube. The coating material can be metallic or high reflectance dielectric. For high gains (five-to-eight times) the reflectance of the coating should be in the mid to upper ninety percent range. The gain in luminance due to the coating will depend directly upon the coating reflectance, inversely upon the angular opening in the coating, and inversely on any self absorption of the light by the gas. An aperture lamp intensity distribution can be approximately represented mathematically by a strip of lambertian material the size of the opening in the lamp coating. Lamps with aluminum coating on the exterior of the tube produce gains in intensity over two, but provide a different luminance profile that is ideal for improving light spreading in optics. The horizontal spread is fairly uniform, but the vertical spread is not the expected Lambertian distribution, which would indicate possible absorption effects within the gas. [6] A typical intensity distribution is shown in Figure 8 for a 2mm aperture opening. The total integrated light output decreases as the angle of view is increased reducing the effective area opening of the aperture. However the peak value actually increases slightly for angles off from the center. While a gain is achieved in the forward direction the increase off axis appears to be associated with reflection and absorption effects in the discharge. This distribution poses no real problem to the optics and in some cases can be helpful to the optics designer to achieve better vertical spread.

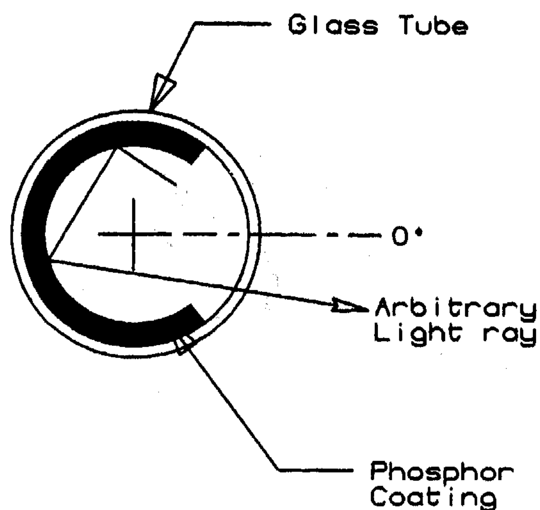


Figure 7. - Crossectional view of aperture lamp with example light ray.

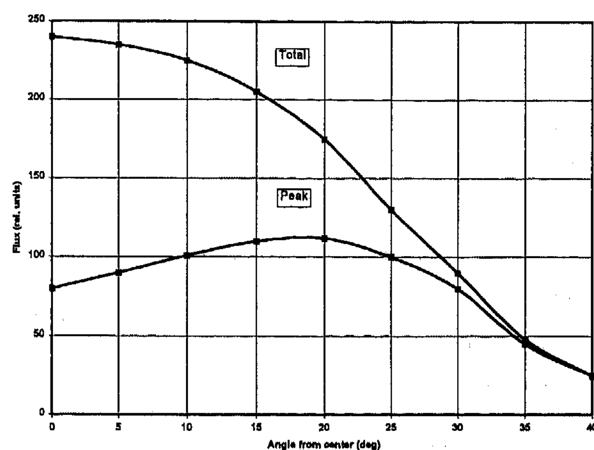


Figure 8. - Luminous distribution fo a 2mm aperture for viewing angles transverse to the lamp axis.

Aperture lamps are very useful for designing small package size luminaires. A thick condenser lens can be placed directly in front of the aperture to redirect the intensity distribution pattern as is shown in Figure 9a. A typical intensity distribution for angles perpendicular to the axis of the optic is shown in Figure 9b. The system was comprised of a three-by-five millimeter diameter aperture arc tube of opening 3mm, and a glass lens of 10mm diameter. The optical system generates a 30° included angle beam of light. An optic of about 10mm in height and length between 300 and 450mm will comply with the 25.8cm^2 area requirement for CHMSL's. The intensity produced by an LPD operating at 26 watts with this design is about 150 cd. Such optics can be located in the automobile at the top of the rear window in the headliner. The small package size and higher luminance can penetrate moderate solar glass windows, and the headliner location provides better viewing for reverse and forward driving.

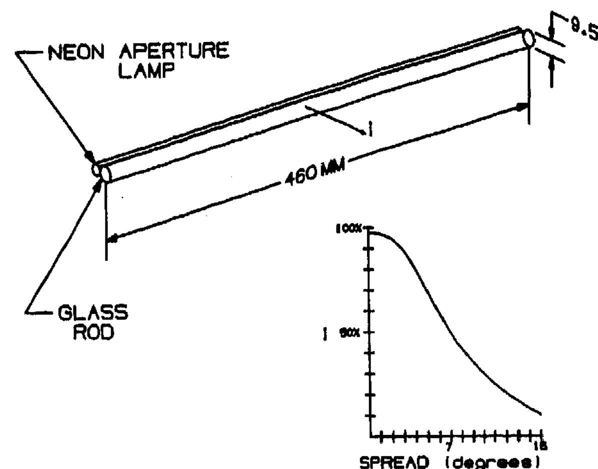


Figure 9a and 9b - Thick lens aluminized aperture lamp. The spread angle is transverse to the lamp axis.

A second similar device is depicted in Figure 10. In this case, the aperture lamp is providing light for a thick sheet optic. The rays from the light source progress through the optics as shown. This device has an added degree of design freedom, the reflective surface, which can provide package depths of 10-15mm and heights of between 25-50mm.

TRUNK SPACE AND SHEET METAL SAVINGS- Thinner lighting packages can be highly beneficial in the reduction of vehicle sheet metal surface complexities. Sheet metal fabrication techniques can be used to deform the metal into a concave section as depicted in Figure 10. A rather complex shaped tail lamp could be mounted into the depression. Thus, thin neon signal packages 25-60mm in height extending from the central part of the car to and including the corner of vehicle appear possible. (See Figure 11).

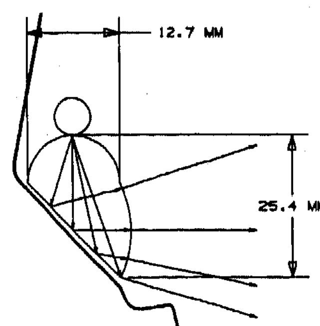


Figure 10. - Aperture source thick lens optic provides thinner "Hang-On" lighting packages for surface mounting.

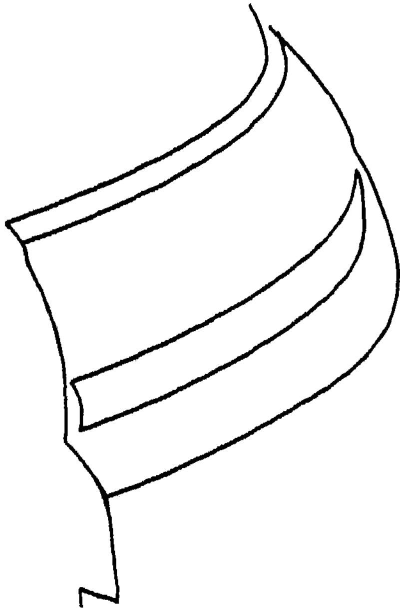


Figure 11. - Neon extended light source provides side markings as well as brake or presence lighting.

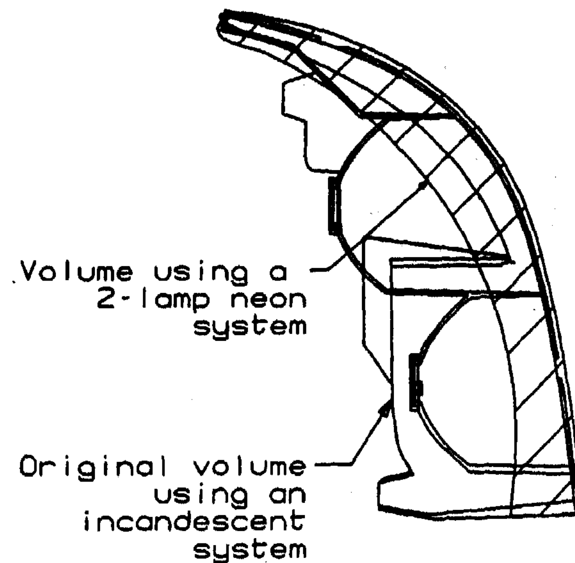


Figure 12. - Improved utilization of the trunk space for transporting large packages.

A lamp with such a thin cross section minimizes cutting and welding of sheet metal compartments on the rear section (See Figure 12). These deformations can be designed into the sheet metal to strengthen the body. While some cutting in the deformed area may be needed for electronics, wiring, and mounting, the largest hole sizes will still be small in comparison with current techniques. This should reduce or eliminate many of the water leaking problems encountered today. Since the lamp is designed to last the life of the vehicle, base-socket systems are not required. Thus, the entire package or section can be attached to the body surface and bolted into place. This packaging would reduce space requirements, adding a cubic foot or more to the truck space in some vehicles. Additionally, the lighting fixture will greatly increase the trunk utilization for storing large packages especially in small vehicles. This is due to the fact that currently the lighting fixtures project into the trunk space causing, in some cases, a high percentage of reduction in capacity for large containers.

LPD SOURCE IN HIGH AMBIENT CONDITIONS - The neon light source luminance is lower than that of either the filtered incandescent or the LED. At first, this might be considered a problem when driving in the presence of the high ambient solar radiation during the beginning and end of the daylight hours. However, as has been shown by M. Sivak, the intensity of the total luminaires plays a more dominate role than the peak source luminance. [7] It should be noted that the light intensity distribution behind the car is dependent upon the source luminance, the optical cross-sectional area and the degree of light spread by the system. Both source luminance and efficiency must be taken into account when designing optical illumination systems for daylight as well as night time driving conditions. High luminance values during night time can cause glare for trailing motorist. While a high luminance is important to achieving high peak intensities, the source must supply enough lumens to provide the desired lighting pattern. Spread lens elements and reflector diffusion effectively reduce the average luminance of the lamp. Thus, while the luminance levels of the LPD lamps are different, the efficiency values (for commonly found vehicle cross-sectional lamp areas) do provide proper rear neon signal illumination functions for turn, brake and presence lighting. This can be achieved at wattage levels from about 15-30 watts for a brake and turn, and about 5 watts for a presence lamp. Strobing or higher levels during daylight are available if desired. The optics in one case studied was about 76cm long by about 5cm in height and depth. This package provided a more uniform wider beam pattern than could be achieve by an incandescent or LED source in the given package depth.

ELECTRONICS

POWER MODES - The neon LPD source and ballast generate both radiative and conductive electromagnetic interference(EMI) when operated in an Alternating Current (AC) mode, and both lamp and ballast need to be shielded and isolated. Common shielding methods involve either a screen or a conductive coating applied to the interior of the fixture or to the lamp surface to reduce the radiated EMI to acceptable levels. The lamps can be designed to be powered in several ways.

The first method is direct current (DC), but this does not convert electricity to light energy efficiently, unless the discharge is operated at low current density levels. For most lamp geometries the power limit is less than 10 watts. However, this approach can be used in some low wattage applications such as appliqués where heat is not a problem because of the low wattage. In this case it may cost less to use DC because filters, screens, etc., are not necessary. The second method is a pure alternating current (AC) sine wave type that provides a better energy conversion for power levels as high as 50 watts. Triangular wave shapes have also been evaluated with an improved efficiency, but this method requires proper electrode design.

Investigations carried out at OSRAM SYLVANIA have shown that the highest output can be achieved by short transient pulses. This method produces a gain of about 1.7 or 1.8 times the AC approach. Briefly, the pulse mode method of driving a lamp can best be explained or summarized as follows: If a discharge tube of rare gas such as neon has a sufficiently large voltage maintained across it's electrodes, it will break down and emit light. This light is formed by the electrons within the neon atom returning to lower states after being excited by the current from the power source. If the supply of current is limited to a few microseconds, an additional gain from the neon emission can be achieved. Allowing the stimulation to decay back to nearly a complete neutral ground state (see Figure 13) provides an efficient light generating method. This decay phenomenon occurs every time the discharge is turned off. During the continuous application of power, the light source dynamics redistribute the energy in the visible, and infrared which produces heat. The upper driving limit is set by the current limitations of component dielectric and of the electrodes while the lower limit is set by the decay characteristics of the gas discharge and/or phosphor excitation as discussed below. A sequential series of pulses are used to produce continuous light. Generally, the rate is above 1,000 pulses per second.

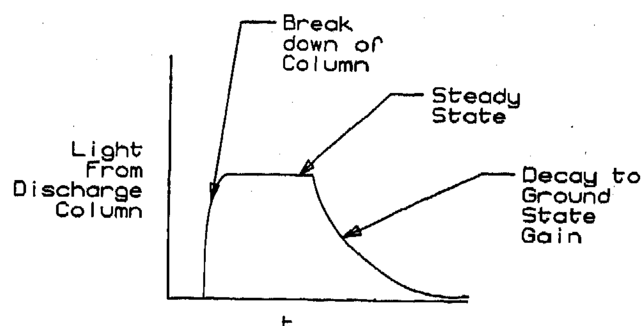


Figure 13. - Advantage of Decay Time (the Impact of Pulse Mode)

The conversion efficiency of electrical power into light will be reduced, however, if the frequency of the pulses is too high. Also, the break down voltage across the electrodes will vary depending upon the time between pulses due to residual charge density within the tube. Typical pulse widths of 1 microsecond with 2-to-3% duty cycle, depending upon gas pressure, bulb wall properties, etc., provide optimum driving conditions for lamps of reasonable lengths and at fill pressures about 100 torr of neon gas. The electron population decay characteristics provide some reduction in break down voltage after the initial pulse. The EMI produced by the pulses of current can be minimized by keeping the leads carrying the high voltage to the lamp as short as possible and by other spatial and time filtering techniques. One aspect of EMI control is to minimize the high voltage lead length by putting a step-up transformer and other electrical impedance matching components as close to the lamp end as possible. These components can serve to isolate the discharge break down from the surroundings of the fixture. For long lamps a second transformer adapter that is driven out of phase with that of the first may be needed at the second end. Figure 14 shows an example of such a configuration. The ballast could be placed remote from the lamp, but this is not advisable for long lengths of lead wire. If the ballast is mounted externally to the lamp module, the ballast needs to provide voltage pulses of less than 50 volts to the transformers (if higher voltage levels are needed, special shielding is required to comply with automotive safety requirements). Such circuits can then run any lamp length with the proper matching of the transformers characteristics. The configuration of transformers at the ends of the lamp and a driving circuit can be designed to operate with any voltage wave shape (non-DC), i.e., pulses, triangular, sine wave, etc.. Thus, a given ballast or driver circuit can be used in all applications by using a matching circuit (adapter). Also, the lamp output can be modified by changing the number of pulses and the wave shape delivered to the arc tube. This may require heavier electrodes for the high current levels, but it yields multi-level lighting mode. This mode can be used to provide a tail and brake function or a tail, brake, and turn function more easily than by amplitude modulation of the source. Modulation of the LPD source by amplitude, digital or frequency is possible. The use of such a drive circuit with the transformer-arc tube combination enables

operation of a lamp of any length in any timing or signaling mode. In addition, two or more transformer/lamp combinations can be shared by the driver circuit, as schematically depicted in Figure 15. Such approaches are necessary to optimize system performance and costs. In multi-discharge lamp systems the regulation, filtering, etc., functions are shared to achieve these goals.

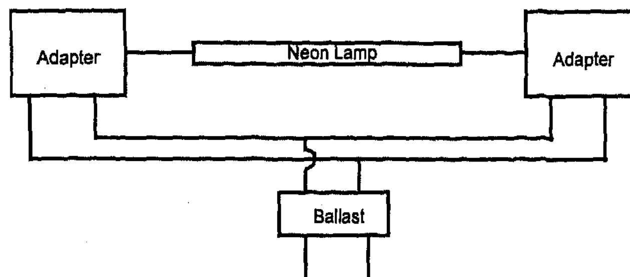


Figure 14. - Neon Pulse Ballast Wiring Diagram with Adapter Interface Circuit for Interchangeable Ballast

During daylight hours and high ambient lighting conditions more signal light can be provided using these techniques than provided during evening hours by changing the number of current pulses applied to one of the arc tubes or by changing the duration of the gating signal depicted in the timing sequence shown in Figure 15. In addition, variable lengths of pulsing of the source can be introduced as a driving aid. For example the current method of signaling a braking command could be changed by introducing the red light level with several initial, visually detectable, red pulses. Pulses of light might be used in some driving conditions to improve recognition of the brake intention.

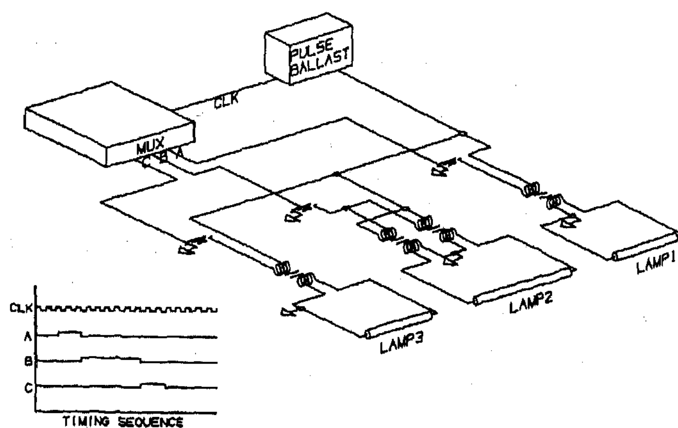


Figure 15a and 15b - Multiple light sources sharing common power source with adapters to interface various lamp loads.

SYSTEM ISSUES

HEAT LOAD AND ENERGY CONSERVATION - The total system power and efficiency needs to be considered when introducing electronic devices and other heat generators such as lamps within a closed, sealed tail lighting package. High ambient temperatures due to geographic and solar environmental considerations along with the absorption of colored lenses and the electrical to light energy conversion efficiency affect system performance. The effects of temperature rise on tungsten and neon sources are not as critical as for the ballast circuit components of a discharge source or for LED light sources themselves. As mentioned earlier, stimulating a LPD lamp with fast transient pulses provides the highest conversion of electrical energy to visible electromagnetic energy. This is a preferred method for reducing heat loads and controlling temperature rise. Concurrently, the pulse mode operation increases intensity in the red region of the spectrum and increases average source luminance for a given diameter, neon filled tube. This option provides a trade-off between a given arc tube wall temperature and arc tube output (over 60 percent gain of light for the same input wattage comparing pulse to AC). In other words, at a lower operating wattage (30% power reduction) a lower wall temperature can be achieved while maintaining luminance and lumen output. Alternatively, a higher luminance will further improve the optical efficiency (the maximum intensity achievable is obtained by the product of the luminance times the cross section of the optical area). Therefore, the pulsed discharge method provides the best optical and energy conservation performance. Balancing these factors, the operating power for the neon source and ballast can be minimized.

Optimization of the power is not only important because of the vehicle requirements, it also reduces the amount of heat energy that must be transfer from the ballast. Heat energy trapped in the ballast can increase individual electrical component temperatures to unacceptable levels before equilibrium occurs. Thus, to limit the temperature rise of the ballast, the design must be as efficient as possible (85% or above), be heat sunk, be convectively cooled, and/or be allowed to radiate the undesired heat energy (The ambient temperature, depending upon placement of the power supply within the vehicle, can reach a value between 85° and 105°C or higher). For example, even an efficient ballast when operating a lamp at 30 watts will need to dissipate about 5 watts of heat. The lamp will produce about one watt of visible light energy, generate about 9 or 10 watts of heat at the electrodes and the remaining heat energy in the gas and leads. Ballast electrical components can be obtained to operate at elevated temperatures before ballast life degradation occurs. However, if not designed properly an individual component in the circuit may be operating at a low efficiency thus generating excess heat, and it may absorb energy from adjacent components. If the ballast is required to operate in an environment of 85° or even 105°C, individual components need to have the rated temperature specification. Some

components may require enhanced energy transfer methods or higher operating temperature specifications to ensure long life and proper operating performance. Added cost, reliability issues, and system complexity, will result.

In summary, as the temperature difference between ballast and the ambient approaches zero, it becomes more difficult and costly to transfer energy to the surroundings, whether by conduction, convection, and/or radiation. The discharge device (HID ,SMFL, or LPD) and ballast must be designed to operate at the lowest power value and highest efficiency possible or it will add cost, consume space, and generate undesirable heat that must be dissipated, or the electronic elements will have a shortened life.

NEON SYSTEM APPLICATIONS

APERTURE LAMP AND VARIABLE COLOR SIGNAL AIDS - The neon LPD source luminance can be increased by the use of a reflecting coat placed on the arc tube bulb wall forming an aperture lamp. Coatings of metal or diffuse dielectric materials can be used to increase the effective luminance of the light source over a limited range of directions at a cost of the total lumen output.

Phosphor materials added to the interior of the arc tube can be used to transform the electromagnetic energy from one portion of the spectrum to another when the transient pulses of current pass through the gas column.[8] The pulse discharge currents in the neon gas stimulate the phosphor producing a light output. The resultant lamp color is determined by emitting materials and the fill gas. As one increases or decreases the number of pulses per second and the pulse duration and wave shape, the ratio of stimulated phosphor light to red light from neon can be adjusted. The light generated from the phosphor is in addition to the red light from the neon column of gas. Some of the red light passes into or through the phosphor, the remaining exiting through the aperture either directly or after multiple reflections from the phosphor. The spectral power distribution of the neon phosphor combination are shown in Figure 16. When the vehicle is moving forward, the rear of the car can display a red presence light for a tail function by applying a continuous sine wave from the ballast to each aperture lamp. When the turn signal is required, an amber color can be generated from the same lamps by changing the wave shape to that of a pulse form (i.e., one micro-second, 2-3% duty cycle). The brake function intensity can be achieved by a higher level sine wave application of power to achieve approximately ten times the intensity of the presence or tail functions. The phosphor material could be used in aperture lamps to blend better with red or other colored gas discharge characteristics. Shown in Table 3 are the possibilities of color with power wave form applications for an aperture lamp.

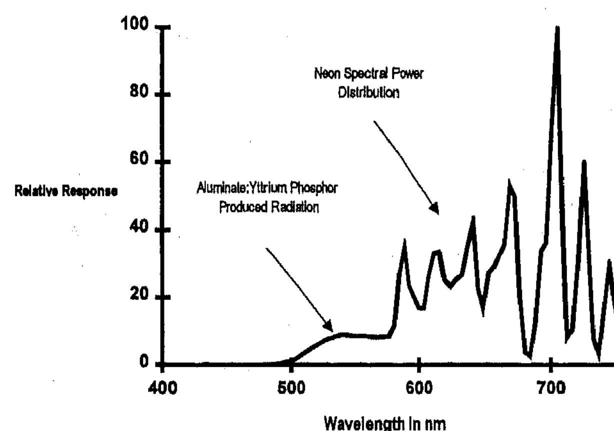


Figure 16 - Spectral intensity for a neon discharge tube coated with agreen phosphor.

ADDITIONAL SIGNAL CONSIDERATIONS - In addition to human visual communications, the LPD source with its fast response time can be used to send information from one vehicle to another, while appearing to be a visually continuous light source. Information can be transmitted optically in the visible or the infrared portion of the spectrum. The frequency is set by the maximum stimulation rate of the gas and ballast. The frequency used to date for the best visual efficiency is about 20Khz. Sending digital codes of information in the visible portion of the electro-magnetic spectrum reduces the amount of radiation being produced in other bands, assuming proper EMI shielding of the source. This is a "line-of-sight" transmission that may prove to be useful, in minimizing noise in the communications channel.

The use of a green emitting phosphor coat in the aperture lamp and a controllable power system allows for a novel signaling opportunity. The radiation is generated by driving current spikes of pulse width of less than one micro-second. The green phosphor coating when stimulated radiates at about 556nm. The stimulus produced by the use of the pulse mode provides two colors green and orange-red. The addition of these two colors provides amber light (see matrix on Table 3). A sine wave produces only orange-red light. Thus this approach can create a visual signal device that can generate a temporal color signal. The color can be switched between amber and red, or other possible colors, by means of adjusting the amount of pulsed to continuous power. It can be done continuously causing a visual "drifting" effect from one color to the other, switching discretely, or a combination of the two (see Figure 17). This also provides an option of using a standard or reference and a variable (continuous or in discrete steps) to communicate a degree of compliance.

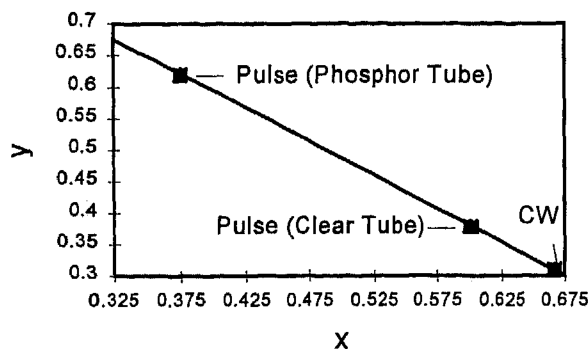


Figure 17 - Shift in the x,y chromaticities during pulse operation.

The red color can serve as the standard to indicate a stop signal or a presence lamp output. By dividing the path between amber and red into several steps one could communicate the degree of slowing of the vehicle. Schematically shown in Figure 18 is a rear lighting assembly to demonstrate the color combinations between a standard red color and a continuously varying red-to-amber source. The lower portion of the figure depicts a temporal signal switching between two color levels. Table 4 provides a list of possible combinations achievable to transmit information. The use of these novel devices must be evaluated before their significance in coding can be attained.

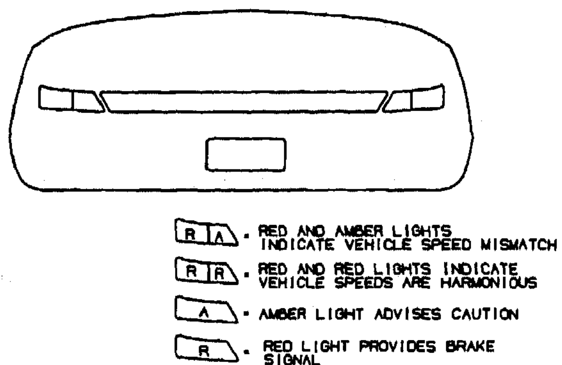


Figure 18 - Schematic of rear lighting. The color combinations for the individual lighting sections are indicated as R=red, A=amber.

VARIABLE LENGTH NEON (WALKING NEON) - By electronically adjusting the pulse height and frequency it is possible to control the discharge length within the lamp tube. While this technique is not new, it does have a good application in automotive. A neon lamp can be made to "sweep" from one side of the lamp to the other at any desired rate. For most applications the sweeping would be at the rate of a turn signal, providing an additional acknowledgment of the drivers intent to turn.

The ability to control the position of the discharge within the tube also provides a wide range of effects that can be both design aesthetic and functional.

A schematical diagram for the walking neon system is presented in Figure 19. The use of an indium tin oxide (ITO) coating as a ground plane is essential for proper operation, but also has the added advantage of reducing the required EMI shielding.

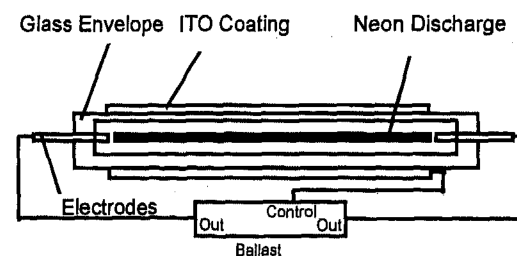


Figure 19 - Schematic drawing of the variable length "walking" neon.

CONCLUSION

The increasing demand for long life light sources to reduce replacement cost is aiding in the transition from incandescent to discharge lighting devices. Both High Intensity Discharge (HID) and Low Pressure Discharge (LPD) sources are becoming available for automotive lighting applications. The HID sources are for forward road illumination and the LPD are for signal lighting. Both sources have excellent resistance to hot-shock breakage - a formidable problem encountered by the incandescent source in the automotive environment. The LPD source has a very long life because of its low wall loading (watts per cm^2) and lamp chemistry. Thus the source and ballast can be sealed into a luminaire with a 12-volt input. This package design can last the life of the vehicle without replacement. Currently, this feature removes the need for legislative jurisdiction and yields the opportunity of harmonization of signal lamps world-wide. The long linear source provides uniformly illuminated areas that can be easily seen from much wider angles than incandescent or LED sources. The uniform illumination will reduce eye fatigue. The faster rise time of the LPD source improves recognition and, therefore, reduces the braking reaction time. The ability to switch colors (red/amber, etc...) enables new coding techniques to be developed to improve communications between drivers and vehicles. By electrically switching power levels, the intensity can be adjusted for different ambient conditions, such as night and day time driving. LPD also conserves power when used as an emergency signal aid on parked vehicles because of its strobing feature. Sheet metal savings in weight, in the component fabrication processes, and structural integrity appears to be favorable. Better utilization

of the trunk compartment is also achieved by the flatter, thinner illumination optics. The smooth continuous flow of the illuminated areas provide the stylist with new exciting tools to provide more appealing and useful vehicle designs.

The accident avoidance capability of LPD affords an additional opportunity to reduce accidents thus saving human life and eliminating suffering. It is difficult to assign a value to an accident but judicial rulings of late are monumental. It is therefore reasonable to consider the cost of new safety devices over the vehicle life cycle similar to ABS brakes and air bags. The ability to reduce or avoid an accident by the use of newly conceived lighting systems is much more economically desirable than building the structure of the car to absorb high speed impacts.

	LED	LPD (Neon)	Incandescent with red filter
Efficacy (Lumens per Watt) for Red	2 - 20*	8 - 16**	3

* Diode wattage range from a typical value of 0.05 watts per diode to a maximum of 0.15 watts. The maximum output is about 2 lumens per diode with present state-of-the-art.

** Depending on ballast characteristics.

Table 1. - Luminous efficacy comparison.

Speed - MPH	Feet/second	Distance Traveled Until driver reacts to light (in feet)		Margin of safety gained by using Neon (in feet)
		Incandescent	Neon	
65	95.3	19.067	0.018	19.049
60	88.0	17.600	0.016	17.584
55	80.7	16.133	0.015	16.118
50	73.3	14.667	0.013	14.653
45	66.0	13.200	0.012	13.188
40	58.7	11.733	0.011	11.723

Speed - KPH	Meters/second	Distance Traveled Until driver reacts to light (in meters)		Margin of safety gained by using Neon (in meters)
		Incandescent	Neon	
105	29.2	5.833	0.005	5.828
97	26.9	5.389	0.005	5.384
89	24.7	4.944	0.005	4.940
81	22.5	4.500	0.004	4.496
73	20.3	4.056	0.004	4.052
65	18.1	3.611	0.003	3.608

Table 2. - Computed travel distances based on reaction times for incandescent and neon.

Power Supplies		
Source	Continuous	Pulse Mode
Clear Neon (3 x 5mm)	Red	Red
Aperture Metalized (3 x 5mm)	Red	Red
Phosphor Coated Neon (5 x 7mm)	Red	Amber
Phosphor Coated Aperture Neon (5 x 7mm)	Red	Amber

Table 3. - Matrix of color combinations for phosphor coated neon.

	BALLAST			
	Continuous Wave		Pulse Mode	
FUNCTION	Low	High	Low	High
Constant Velocity	X			
Acceleration	X	X		
Brake Release	X		X	
Coasting				
Low Brake		X		
Moderate Brake		X	X	
Emergency		X		X

Table 4. - Information modes transmitted through neon lighting combinations.

REFERENCES

[1] Rothwell, Jr., H. L., and English, G. J., "Automotive Discharge Headlights: a New Light." Automotive Technology International 1990, Sterling Publications.

[2] Rothwell, Jr., H. L., and English, G. J., "Automotive Discharge Headlights: A Colour Study." Automotive Technology International 1991, Sterling Publications.

[3] Rothwell Jr., H. L., and English G. J., "Neon Discharge Vehicle Lighting." Automotive Technology International 1994, Sterling Publications.

[4] Sivak, M., Flannagan, M. J., Sato, T., Traube, E. C., and Aoki, M., "Reaction Times to Neon, LED, and Fast Incandescent Brake Lamps." The University of Michigan Transportation Research Institute, Report No. UMTRI-93-37, September 1993.

[5] Sivak, M., Sato, T., and Flannagan, M. J., "Effects of Voltage Drop on Rise Time and Light Output of Incandescent Brake Lamps on Trucks." The University of Michigan Transportation Research Institute, Report No. UMTRI-93-28, July 1993.

[6] Fohl, T., "Design Study of Collimating/Turning Prisms for Neon Aperture Lamps." Technology Integration Group, Inc., March 29, 1995.

[7] Sivak, M., Flannagan, M. J., Olson, P. L., Bender, M., and Conn, L. S., "Evaluation of Brake-Lamp Photometric Requirements." The University of Michigan Transportation Research Institute, Report No. UMTRI-86-28, July 1986.

[8] Jennato, S., and Rothwell, Jr., H. L., "Neon: A Mercury-Free, Multi-Colored Light Source." SAE Technical Paper Series No. 960921, February 1996.

970655

Neon Signal Lighting - An Integrated Approach

Philip Rimmer
Tunewell Technology Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Neon discharge lamps contribute four main benefits in automotive rear signalling applications, longevity, efficacy, speed of response and most importantly styling versatility. In systems demonstrated to date, the styling advantages have failed to materialize whilst the cost of implementation has been discouragingly high. Together, Corning and Tunewell seek to solve these problems through the use of sophisticated glass processing and novel power distribution technologies, and the adoption of a systems approach to design to minimise cost.

A glass molding process will be detailed permitting the integration of optimal topology discharge channels with compact optics and connection systems into any curved, laminar form required by the lamp designer. Clear aesthetic advantages over incandescent and LED technologies can be shown due to the extended, uniform light source and the precise placement of its associate optics.

A power distribution system will be demonstrated that provides a substantial cost reduction in the provision of power for multiple lamps, additionally eliminating high voltage wires and bulky, individual ballasts. The system will be shown to facilitate multiplexed control and fault reporting. Extension to encompass additional power supply tasks within the vehicle will further reduce costs.

A systems approach will be outlined demonstrating the cost benefit of reducing the power supply burden in favor of an enhanced lamp performance. It will be shown that substantial cost advantages accrue if power requirements, lamp strike requirements and radiated and conducted emission problems are dealt with at appropriate locations in the system.

1. INTRODUCTION.

Often discussed in this forum has been the issue of vehicle power supply systems. The problem of coping with the ever growing demand on the vehicle's power supply has led people to moot higher and multiple power supply rails in the pursuit of efficiency. At 48 volts it has been argued that alternator and motor efficiencies are increased by 4 and 6 percent respectively. The cost and complexity of such a supply change are, however, intimidating for these small returns.

Equally, people have been discussing the implementation of higher efficiency components, not least amongst these lighting components. Discharge lamps like HID, Cold Cathode (Neon) and Hot Cathode (Fluorescent) and LEDs all offer substantial efficiency improvements along with many functional enhancements. Again, component cost has appeared discouraging.

Fortunately, a number of interesting technologies are appearing that have a degree of synergy. To better appreciate why this may be so, it is revealing to analyse the kinds of power supply automotive components would use preferentially.

TABLE 1.1 OPTIMUM POWER SUPPLY FOR AUTOMOTIVE COMPONENTS

COMPONENT	VOLTAGE	CONSTANT CURRENT	CONSTANT VOLTAGE	REASON FOR CHOICE
Incandescent lamp (fast response)	low <12v		YES	Robust filament
Incandescent lamp (slow response)	low <12v	YES		Robust filament and soft start
NEON (cold cathode)	v.high >>12v	YES		Lamp requirement
HID	high >12v v.high >>12v	YES		Lamp requirements
Fluorescent (hot cath.)	high >12v	YES		Lamp requirement
LED	high >12v	YES		More efficient power use
ICE (stereo)	high >12v		YES	Eliminates internal PSU
IGNITION	high >12v	?	YES	Reduces IGBT loss. Increases power available for lean burn.
MOTORS	high >12v		YES	Increases efficiency
LOGIC (microcontrollers) low	<12v		YES	Circuitry requirement
COMMS.	low / 12v		YES	Circuitry requirement
HEATERS etc.	various			

Though the above is an incomplete list, very few items have as their optimum a 12 volt, constant voltage supply system. There is also a lack of bias towards any one type of supply. Moving to higher voltages may partially address our concern for improved efficiency, but it certainly raises other problems.

The unusual power supply demands of discharge lighting are, in fact, for a standard 12v power supply, the major cost element in the use of these components.

A debate is now emerging that power savings brought about by novel and improved electrical components can be set against cost savings in the vehicle power supply system. Inevitably the component supplier's view tends to be at variance with the vehicle manufacturer's. Some component suppliers anticipate a 50 cent per watt saving, manufacturers merely 5 cents. The truth,as ever, will lie somewhere between those two points. The maximum theoretical saving will only come about when the power supply system components are fully optimised for the reduced power demand, and as most supply components only come in discrete sizes, a lot of accumulated power saving may be necessary to trigger a real economy.

The car manufacturers dilemma is that he cannot specify a more efficient electrical component if it is more expensive, and does not immediately pay for itself in savings elsewhere. Incorporating many such devices might unlock the power system savings that would pay for their inclusion, but this is a scenario with considerable risks. There is clearly something of a log jam here, locking out improvements such as discharge lighting.

The question must be asked - Is there a universal power supply system that allows us to optimise all or most functions? Is there a power supply system that is more accommodating to change, reducing the cost of implementing components with novel power demands?

2.0 WHAT IS THE POWER SUPPLY SYSTEM?

Current thinking has it that the "power supply" system consists of all those components and wires carrying the battery voltage. The complement of this is that the "powered components" are those that receive this battery voltage and produce a useful function in response. A growing number of components, however, contain their own power supply sub systems. Ignition has always needed to reprocess the battery voltage to derive the very high voltages needed at the spark plugs. In-Car Entertainment now makes such power demands that it contains a voltage boosting power supply of its own. Similarly, logic and micro-processor control units managing disparate auto functions require their own voltage-converting sub-systems. HID headlamps have massively sophisticated voltage boosting and regulating power supplies.

Every one of these distributed power supply sub systems individually require further complex and expensive protection circuitry, because, not only is the battery voltage inappropriate for the job but it is highly variable and noisy. It would seem that any new electrical component with a degree of electronics involved will bring with it another expensive bit of power supply system, discouraging its implementation. Neon signalling lamps are a case in point.

3.0 NEON DISCHARGE LIGHTING.

3.1 Discharge lighting has a great many attractions amongst them, efficacy, longevity and robustness. For the lamp designer two of the greatest opportunities are the ability to work with a light source that is uniform and compact facilitating very effective optics (HID and Neon) and highly conformable (Neon) facilitating styling.

A Neon lamp, consisting as it does of, glass, iron (typically) and a small amount of gas has negligable material cost. Such significant cost as do exist reside in the processes that form the lamp into useful shapes. For this reason the first Neon lamps used on cars were CHMSLs exploiting the simplest straight tube style lamp.

Turning such tubes into complex shapes is entirely possible. Compact multi-tube fluorescents are cheap but limited in the varieties of shapes that are possible, none particularly useful. The neon sign industry produces most complex shaped lamps, but even with semi-automated bending equipment the unit cost is very high.

3.2 CORNING'S MOLDING PROCESS- The need for a low cost forming process was siezed upon by Corning who adapted an existing vacuum and pressure forming process to create two plates simultaneously, bearing the complementary shapes of the lamp channel. The process continues with the fusing together of the two, still softened, halves of the lamp. A sophistication of the process allows the automated introduction of the electrodes whilst the lamp is still hot. The whole operation is complete in a short space of time yielding a complete lamp, needing only to be cleaned and gassed.

Whilst the cost advantages of such an operation are obvious, the full technical implications which can lead to further cost savings are not obvious. For many reasons the pursuit of efficacy (i.e. the lumens per watt) is highly desirable. Reduced power demand may have its rewards, but so to does the resultant reduced temperature rise. Lower temperatures translate to improved reliability, but also, if sufficient gains are made, a lowering of enclosure costs. The molding process makes two contributions to reduced power consumption.

First, it can facilitate an optimum efficacy tube by permitting highly constricted discharge channels at relatively low gas fill pressures without compromising tube life. Tube life is limited by the amount of gas in the tube, the electrode area and the electrode current. Low volume, small electrode lamps have short lives. The molding process can easily accomodate the inclusion of gas reservoirs providing the necessary life span. By using a constricted channel at low pressure an optimum, discharge-region efficacy can be achieved at the same time as producing a high voltage gradient in the column. This in its turn permits the use of relatively low electrode currents to achieve the desired power input into the discharge region. This low current is desirable as it minimises electrode losses, losses that make no contribution to the lumen output.

TABLE 3.1 10 WATT NEON LAMP WITH DIFFERENT ELECTRODE CURRENTS

TUBE POWER	TUBE CURRENT	ELECTRODE POWER	DISCHARGE POWER	LIGHT (REL.)
10 watts	30mA	4.5 watts	5.5 watts	65%
10 watts	10ma	1.5 watts	8.5 watts	100%

The second advantage, stems from the narrowness of the channel and a further potential of the molding process. Stop, tail and turn lamps have small vertical dispersion angle requirements. A narrow line of light is easy to focus to this small dispersion angle without requiring a very deep reflector. Such a reflector (unsilvered) can be fabricated that is integral to the lamp. This reflector will always be perfectly sited with respect to the tube and therefore can utilise greater degrees of focussing without risking the misalignment of a discrete lamp and reflector.

The design of an integral reflector is quite an interesting task as it involves the combination of reflector optics and light guide optics. The extent and coverage of a silvered coat is subject to factors too numerous to explore here, but suffice it to say that much optical and electrical advantage can be gained from this simple step.

Many other lamp structures can be produced in similar fashion. Front refractor optics can provide advantage as can glass tinting. In summary, very efficient light sources can be created by this process at low unit cost with the singular advantages of covering large, compound-surfaced areas, yet remain very low profile.

4.0 THE COST OF USING NEON LAMPS.

4.1 Lamps produced by the above molding process are low cost compared with bent tube equivalents, but in integrating other functions and minimising package height even greater cost savings are realisable. Thin, high efficacy lamps will also possess lower enclosure costs due to the reduced amount of material required as well as the lower temperature rating mentioned earlier. Indeed, in future, it can be envisaged that complete rear lamp assemblies can be fabricated in a thin, ready-wired package that can be attached to the outer surface of the vehicle bodywork. Such pre-tested packages could well incorporate in a single component all rear signalling and include such components as fenders, spoilers and license plate and will have a major impact on vehicle assembly costs. Such superficial lamps will in addition maximise vehicle internal space, simplify body panel press tools and increase body panel strength.

Now we must address the major cost penalty of using any discharge lamp. This is the cost of converting power at the battery voltage to the high voltage, high impedance type required to safely drive the lamp. Typically a ballast to do this task will cost some 5 to ten times the cost of the lamp itself. Ballast costs are unlikely to drop below $7 even for the lowest power lamps, and current designs are in the region of $10. In consequence a typical five lamp rear-end arrangement with four Stop/Turn/Tail lamps and one CHMSL,would cost $50 for the ballasts alone.

4.2 "COMPLEXITY COST"- One of the reasons for this floor to the cost of a ballast, is the necessary screening, filtering and protection circuitry required by each. There is, if you like, a fixed "complexity cost"associated with each ballast whatever its rating. This complexity is related to the variety of disparate functions in the ballast, by which is meant the number of different sorts of things it has to do. This is reflected not only in the number of components required but also the number of functions to be tested at manufacture. The great majority of these functions are included for safety reasons, protecting the unit from outside abuse and vice versa.

5.0 A NOVEL POWER DISTRIBUTION SYSTEM.

5.1 AC POWER DISTRIBUTION SYSTEMS- One of the lesser known power distribution systems proposed for vehicles is the use of a High Frequency Alternating Current (HFAC) power bus. The major advantage of this system is the ease of changing between different voltage levels. Indeed, one hundred years on from the commercial tussle between Tesla and Edison the arguments for AC over DC are identical. The arguments can be summarised as follows-

1) Power is transmitted more efficiently at higher voltages because ohmic losses are lower.
2) Unfortunately, high voltages do not suit all / many loads.
3) Fortunately, AC can be transformed easily to differnt voltages.

To Edison, the champion of DC power distribution systems, Tesla's AC system, generating and distributing AC at high voltages, then transforming it to lower voltages locally, seemed complex and potentially expensive. What he overlooked was-

- The cost of power from a single large power plant was less than from many smaller ones. (There is a complexity cost associated with each power plant.)
- Many electrical appliances worked advantageously with AC and most didn't care.

Indeed Tesla offered not just a competing system of AC transmission, but also designs for reliable AC powered machines including induction motors. The flexibility of AC as a power source facilitated the introduction of discharge lighting of all kinds, by providing appropriate running voltages and lossless current regulation using inductors.

5.2 THE HIGH FREQUENCY ADVANTAGE- The debate for HFAC is almost identical with one major addition- the high frequency.

For the most part the range of frequencies being considered is in the range 20 - 200kHz, in other words, the first decade of frequencies above the range of hearing. Whilst not being very high it is still high enough to bring about a huge reduction in the size and cost of the transformers required to implement a working system.

Basically, magnetic flux density in transformer cores is inversely proportional to frequency, so we might expect that we could reduce the volume of a 60kHz transformer to one thousandth that of a 60Hz transformer. In fact, the scale of difference is ten times less due to the maximum flux desity allowed in high frequency transformer materials.

TABLE 5.1 TRANSFORMER VOLUMES AGAINST FREQUENCY

Frequency	Transformer Power	Transformer Volume	Cost
60Hz	100W	100 cu in	$5
60kHz	100W	1 cu in	$1-50

A fairly good assessment of the cost of a 60kHz transformer can be made using the following expression-

Total Cost = Complexity Cost + Material Cost

where Complexity Cost = 50 cents
and Material Cost = 1 cent per watt

Given the small physical volume of the transformers likely to be needed for each function in a vehicle it is clear that they would not be dissimilar in size to the conventional connector used in a 12V DC system.

Indeed, building such a transformer into a connector could be the lowest cost way of housing such a component.

Whilst, these costs are significant, I believe they are less than some might expect. Certainly, when viewed against other system savings they are not discouraging.

The argument regarding power production costs carries over very well into HFAC. If we look at the problem of powering 5 Neon lamps, for instance, running at 10watts each-

TABLE 5.2 COST OF POWERING FIVE, 10 WATT NEON LAMPS

Lamps	Ballasts	Ballast Cost	Transformers	Transformer Cost	Total Cost
5 X 10W	5 X 10W	$50	0	0	$50
5 X 10W	1 X 50W	$30	5	$3	$33

Ballast costs are computed with the same expression as for 60kHz transformers, but with the following values-

Complexity Cost = $5
Material Cost = 50 cents per watt

(The Material Cost figure holds good 10W to 100W but becomes depressed above this range.)

As mentioned earlier Edison's final oversight was not realising some electrical tasks worked better with AC. Reviewing our list of loads we can see that all those component / functions preferring voltages other than 12 volts or preferring current mode delivery would use AC with advantage. Clearly transformers can supply the most appropriate voltages but where current mode delivery is preferrable, this can now be achieved by lossless ballasting with inductors or capacitors. LEDs are a case in point.

From the foregoing arguments, regarding the cost of powering discharge lamps, it would seem that these might serve the same facilitating function that Tesla's induction motors did.

6.0 THE PROBLEMS OF HFAC

To date an HFAC power distribution system has not been implemented due to a number of problems. These

are, electromagnetic interference, HFAC switching issues and the daunting number of affected components.

6.1 EMI- Broadcast interference from the distribution cable, mainly in the form of a high frequency magnetic field (H-field), can cause serious interference to other electrical apparatus. It can also lead to serious power loss into adjacent metal surfaces. The total amount of magnetic flux generated is dependent on the magnitude of the current and on the area of the loop around which the current flows. Apart from altering the current, two approaches are taken to minimise the H-field. First, the loop area is made as small as possible and second the conductors are screened. Some distribution cable types are listed below. The peak H-field flux levels are noted in the region close to the cable (0.5mm to 5mm) and are assessed at a nominal 1 amp current. In appendix B the flux plots in this close region are shown.

6.1.1 Wire Pair.

Construction:	Wires kept adjacent to each other by their insulation. This proximity minimises loop area and helps reduce H field.
Cost:	Very cheap.
Distant H-field:	Medium
Local H-field:	Very High. 0.19 milliTeslas.Fig. B1
Flexibility:	Good

6.1.2 Twisted Wire Pair.

Construction:	Wires kept adjacent by twisting together. Rotating field cancels at a distance.
Cost:	Very cheap.
Distant H-field:	Low
Local H-field:	Very high. 0.19milliTeslas. Fig. B2 (Flux plot shows adjacent twists side by side not in line)
Flexibility:	Good

6.1.3 Screened Wires

Construction:	Wires, probably twisted, with a braid outer screen.
Cost:	Medium
Distant field:	Low
Local H-field:	Med./ low. 0.02 milliTeslas
Flexibility:	Poor

6.1.4 Coaxial Cable.

Construction:	Inner conductor with return path conductor as outer sheath.
Cost:	Expensive
Distant H-field:	Very low
Local H-field:	Very low 0.09 microTeslas. Fig. B3
Flexibility:	Poor. Very difficult to break in to and difficult to terminate.

6.1.5 In an automotive application we need to be particularly concerned over the issue of H-field in the local region (i.e. next to the cable). Such a cable will undoubtedly share some of its path with other cables, perhaps sensor and control cables. Also it will lie against the body of the vehicle at some point. It can be seen from the above that only coaxial cable provides the very low field that we desire to assure ourselves of satisfactory results.

6.1.6 DOUBLE SIDED FLEX CABLE- Another cable structure exists that may prove particularly appropriate for automotive use. Following are three different structures, the latter two incorporating amorphous or nano-crystalline metal screens.

6.1.7 Two layer "Flex"

Construction:	Two flat strip conductors separated by a very thin insulator. (Cu, I, Cu)
Cost:	Cheap
Distant H-field:	Low
Local H-field:	Low. 2 microTeslas. Fig. B4
Flexibility:	V.good flexibility.Easy to break in to and terminate.

6.1.8 Two layer "Flex" plus back screen.

Construction:	Two flat strip conductors separated by a very thin insulator plus a back screen of high permeability (Hi-u) foil. (Cu, I, Cu, (I), Hi-u)
Cost:	Med/low.
Distant H-field:	Low/very low.
Local H-field:	Low/very low. 0.24 microTeslas Fig. B5
Flexibility:	V.good. Easy to break in to and terminate

6.1.9 Two layer "Flex" plus back and front screen.

Construction:	Two flat strip conductors separated by a very thin screen of high permeability foil. (Hi-u, (I) Cu, I, Cu (I), Hi-u)
Cost:	Medium
Distant H-field:	Very low.
Local H-field:	Very low. 0.002microTeslas. Fig B6
Flexibility:	V. good flexibility. Easy to break in to and terminate.

Comparing unscreened twisted pair (6.1.2) with unscreened flex (6.1.7) there is a 100 fold flux reduction for the flex, the reason being the dramatically reduced loop area. Remarkably the flex can be screened by placing high permeability foil on one side only (6.1.8). The result is a further ten fold reduction in flux on the non-foil side and a one hundred fold reduction on the foil side. The flux reduction on the non-foil side is due entirely to the close positioning of the flux generator to the low reluctance path of the foil.

Placing screening on both sides of the flex (6.1.9) reduces flux levels to one hudred thousandth of the unscreened twisted pair, substantially better even than the coaxial cable. The screening foil can be electrically connected to the conductor on each side, allowing power removal from the surface of the cable without disturbing its very low H-field structure.

Much development work has been done on this low H-field cable minimising various small loss mechanisms. The end result is a distribution cable whose reduced copper content, when used at 84 volts rms at 60kHz, will more than compensate for the added cost of its novel structure.

6.2 CONTROL ISSUES- HFAC systems are advantageously implemeted with remote switching. Such switching may be carried out in a number of ways.

6.2.1 HFAC constant voltage systems can use semiconductors to interrupt the power on both the primary and secondary sides of any coupling transformer. Because the power is processed and regulated these semiconductors can be considerably lower rated in their voltage specs as transients are prevented entry by the central power supply. Further, by employing zero-crossing switching techniques, generation of any further transients will be avoided. Semiconductor switching costs will be substantially reduced.

6.2.2 Much work has been done on constant current systems. In these the power is provided either by using a constant current output power supply or by inserting a current regulator in series with a feed from an HFAC constant voltage bus. The former would be used where a current loop was required on its own, the latter where a mixed voltage and current system was required.

See Appendix A Fig.A1

A constant current power supply utilising the same flex structure would permit a very cost effective power delivery and control method for discharge lighting. The regulated current in the loop would be transformed down to the current level required by the lamp. The lamp current would be defined by the loop current and the turns ratio of the transformer. Control of each discharge lamp is simply and cheaply achieved by causing the transformer itself to saturate and thereby extinguishing the lamp. Such saturation can be achieved with the application of approximately 100mW of DC power to a control winding.

These saturable coupling transformers constitute massively robust switches that have a very useful analog control region and can be used to moderate the voltage peaks present in the turn-on cycle of cold cathode discharge lamps. The degree of control can eliminate transients in the current loop when switching a wide variety of loads and can also switch sufficiently fast to permit power level control through pulse width modulation.

Special techniques have been developed to power HID lamps, and some work has been done to use a current loop power supply for ignition systems.

6.3 IMPLEMENTATION- Implementing an HFAC bus system throughout a vehicle is a daunting task, not least because it would affect so many components and systems, all of which would require modification. Fortunately, the task can be broken down into smaller parts. It is envisaged, for instance, that a stand alone constant current loop system could ideally power a complete Neon, rear-end signalling arrangement. Provided a reasonable number of lamps need powering, the constant current loop approach will confer a considerable cost benefit over discrete ballasts.

Once the step to put some HFAC on a vehicle has been taken any further tasks for it will simply drive down the effective cost of the system and its components. HID added to a neon equipped car will effectively drop power supply costs for both Neon and HID as a further ballast "complexity cost" is omitted. A progressive growth incorporating interior fluorescent lighting, car stereo, ignition systems etc. will come to displace further amounts of noisy, unregulated battery voltage wiring.

Another driver for such a system will be the desire to implement multiplexed control of the vehicle's electrical functions. At the moment, much effort to achieve such a system is partially thwarted by the problems of dealing with the 12 volt bus that supplies the power to the remote functions. Such a bus carries very high currents to many switched loads. It is unwieldy and is subject to unprecedented levels of transient noise and voltage variation. A regulated HFAC bus is entirely complementary to multiplexed control. Indeed it offers the high reliability, "soft" switching needed as part of the system.

As a final note on this, techniques have been devised to provided back signalling to the central power supply for fault diagnosis. These are rapid poling techniques that require little extra hardware and permit a multiplicity of functions to be controlled via a single signal bus (e.g. CANBUS) node at the power supply or regulator.

7.0 CURRENT LOOP POWERING OF DISCHARGE LAMPS

7.1 The constant current loop approach to powering discharge lamps confers control benefits to the lamps as well as reducing total system cost. These benefits, however can only be exploited if the power supply provider works closely with the lamp maker. With an understanding that the cost of the discharge lighting system resides mainly in the power supply itself, the two groups need to solve problems at a system level if lowest overall costs are to be achieved.

7.1.1 Power supply requirements are by far and away the greatest cost incurring parameter. At 50 cents per watt, the saving of 4 watts could well pay for the lamp.

An efficient optical system is the first priority. A light source that is readily focussed, using a high quality reflector helps here. Also recognising that a neon light source needs only very slight filtering to meet colour specification. Using standard lenses is throwing power away.

Achieving maximum positive column efficacy can be problematic, but if the lampmaker can extend tube life by artificial means (e.g. gas reservoirs) then optimum gas pressures can be chosen without worry.

7.1.2 Cost resides in the dealing with the high voltages that occur in Neon tubes, particularly during the tube ignition period when voltages may climb as high as five times above the running voltage. These peaks require considerable upgrading of the transformer and connector insulation and may even force expensive manufacturing and testing techniques. Here, the lamp maker can aid the situation by aiding a process known as "flow-on". This phenomenon occurs when high frequencies are used to power a Neon tube.

If a high frequency voltage, applied to a tube, is ramped up over a period (several milliseconds) the tube will partially ignite at the electrodes even when the voltage is quite low. Numerous factors are involved in this, but essentially what is happening is that the discharge is capacitively coupling into the environment. As the voltage rises the glow extends out from the electrodes until until they meet in the middle of the tube. Because the tube has been partially alight all the time the voltage never goes far above its normal running value.

The lamp maker can help by utilising the reflector metalisation to enhance the capacitive coupling. This technique can specifically use these external conductive areas to shunt current from the end regions of the tube towards the centre. With care the result is, not only the stimulation of "flow-on", but also a lower peak voltage due to earlier full-tube conduction. A further benefit accrues in a balancing-up of tube brightness.
This same capacitive coupling can bleed current away from the discharge, particularly at the higher voltage ends. The result is the loss of current and therefore brightness in the central region. Low current, high voltage tubes can be noticably non-uniform in brightness. The reflector metalisation, with appropriate electrical discontinuities can restore discharge current uniformity.

Such techniques can only be utilised cost effectively where the reflector and lamp channel are of unitary construction. Molded glass technology is particularly appropriate here.

7.1.3 RADIATED EMISSIONS 1.- It is not often appreciated but the biggest generator of interfering harmonics is usually the discharge tube itself, and it is these harmonics that are the major nuisance. It is comparatively easy to make a power supply produce a low distortion sinusoidal current. Such a current at 60kHz would produce little concern fed to a resistive load. Discharge tubes, however, can be far from resistive in their electrical characteristics.

At 60 Hz a Neon tube will extinguish and re-ignite every 8 milliseconds. Each ignition constitutes a step, a sharp corner, appearing in the tube voltage even though it is fed from a sinusoidal current generator. A considerable high frequency component exists in this step and it is not unusual to find line powered neon signs interfering with AM radio, security tag systems and the like. Increasing the frequency of supply to the tube will result in a softening of this step as some of the neon remains ionised through to the next half cycle. By about 20 kHz most will remain ionised and the step may be no longer visible as such. There is, however, a broad region of operation, depending on the lamp pressure,

diameter etc., when noticable "conductivity modulation" occurs, That is, the conductivity (and therefore resistance) of the lamp varies throughout the sinusoidal current cycle. In short, the voltage waveform will not itself be sinusoidal, but will contain harmonics. By about 200kHz these harmonics have normally disappeared, and the radiated emissions will be comparatively simple to deal with, but at the lower frequencies we want to use, to control capacitive loss from the discharge, we may generate a plethora of much higher frequency components.

The lamp maker can do most to control radiated emissions. Indeed, implementing the start -voltage-limitation-strategy outlined above will help considerably. Channel widths and gas pressures are the main parameters to vary. The power factor of the tube is directly related to harmonic generation, and higher power factors (less harmonic generation) correlate to lower pressures. Some lamp makers have been forced to adopt very high pressures to meet lifespan requirements with consequently poor radiated emissions performance.

The physical layout of the channel has a profound effect on broadcast interference. Because, the voltages are high and the currents low, the major problem is one of E-field generation. The connecting wires and the tube ends, where the voltages are highest, constitute the antenna system broadcasting the interference. The lamp maker can reduce radiated emissions by reducing the efficiency of this antenna. This is achieved be reducing the extent and physical separation of the highest voltage parts of the circuit. For, example a U-tube is greatly to be prefered over a single straight tube. Long high voltage leads are eliminated and throughout the antenna system points of equal and opposite voltage are in close proximity.

The molded lamp can achieve the most desirable structures without incurring any additional fabrication cost. Electrodes can be presented in close proximity, and interference cancelling layouts adopted. The compact nature of the coupling transformers used by the current loop system permit very close siting to the electrodes obviating the need for high voltage cable.

Some have reasonably attempted to use DC to power neon lamps to reduce radiated interference. The results are fairly good, but limit maximum tube length due to ignition problems deriving from the great reduction of flow-on effect. The constant current loop system can help here, also. By using one transformer to supply the highly filtered DC (via rectifiers etc.) and another to provide an AC trigger stimulus, any reasonable length of tube can be powered by DC cost effectively.

7.1.4 RADIATED EMISSIONS 2.-The emission mechanism of conductivity modulation can be described as differential in that the interfering signal is manifest as a voltage difference between the ends of the tube. Another mechanism exists that is best described as "common-mode". The signal produced by this mechanism is essentially a voltage of between zero and 150 volts consisting of even harmonics of the tube current and generated between the whole tube and its power supply. The source of this signal is the steep cathode voltage drop region changing ends every half cycle. Any asymmetry in the capacitive coupling to these regions results in the generation of a common mode signal.

Discrete tubes and reflectors suffer capacitive assymetry due to dimensional tolerances and so need extra circuitry to filter out the resultant common mode noise. Molded glass lamps can be manufactured with much tighter tolerances and so that the amplitude of the signal can be reduced. As a single component it offers the possibility of being "trimmed" to null the effect completely.

8.0 CONCLUSION

The growth of electronics installed on a vehicle is one of the few things we can be sure of in future car design. The demand for power is growing in both quantity and varieties of supply. Useful new technologies like discharge lighting, that save substantial amounts of power have cost penalties related to power supply. Viewing the totality of a vehicle electrical system, i.e. switches, looms and loads we can see that the more "electronic" loads, like HID, ignition systems and car stereos have a substantial "complexity cost".

A system of power distribution has been proposed that offers the possibility of absorbing much of this complexity into a single unit and thereby significantly reducing overall vehicle costs. The stable power delivery and noise-free switching will simplify the performance requirements for all loads. The option of delivering power at any voltage or current will permit the eventual optimisation of all existing loads, and remove substantial barriers to the incorporation of new ones.

Such a comprehensive system could never be introduced onto a vehicle in a single step as every electrical component would need redesign. It is envisaged, however, that sub-systems can be introduced with immediate cost benefit. Specifically, neon rear signal lighting, where a new low cost lamp manufacturing process can supply designers with the shapes and appearance they need, is just such a sub-system opportunity. Other light sources, HID, LED, hot cathode and incandescent provide other sub-system opportunities, likewise ignition, car stereo and motor control. Multiple sub-systems on one car may, one day, lead to the adoption of a pan-vehicle HFAC power delivery system, the ultimate complexity reduction.

APPENDIX A

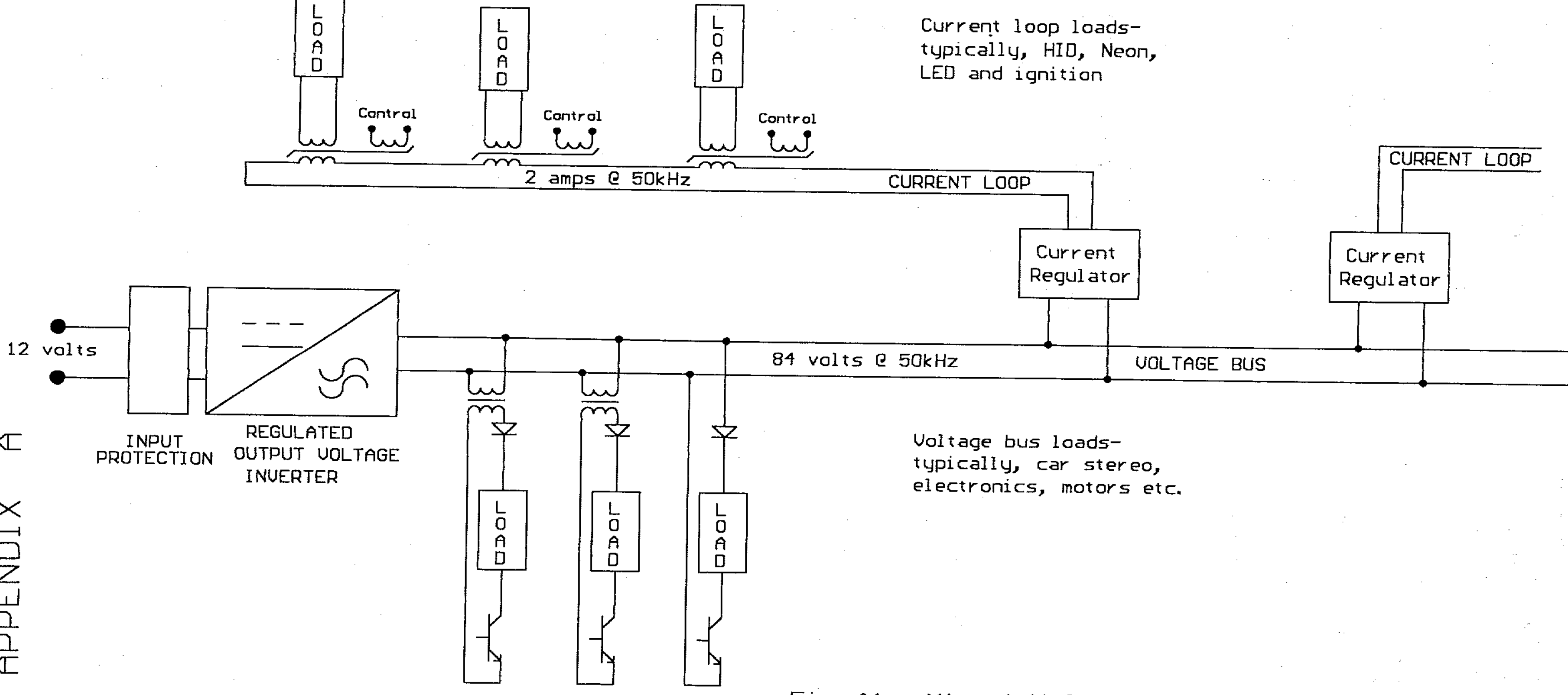


Fig. A1 Mixed Voltage and Current Power Distribution System For Vehicular Use.

Appendix B - Flux Plots

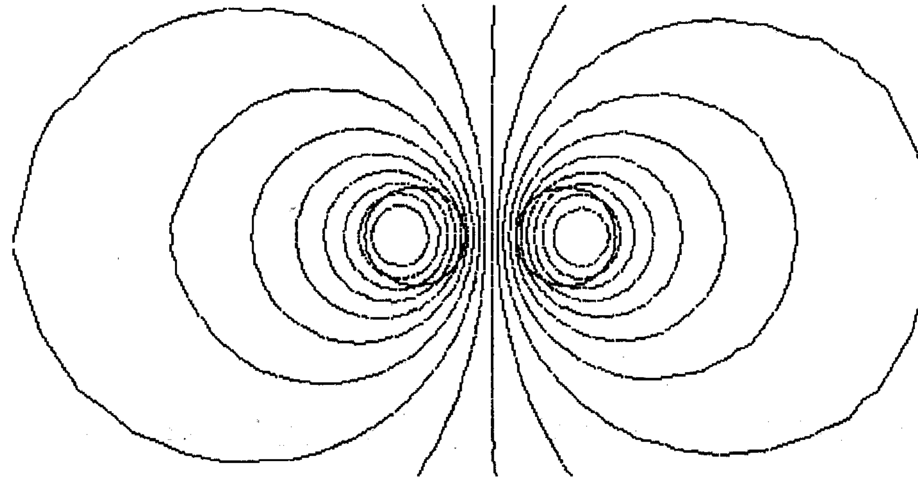


Fig B1 Wire Pair

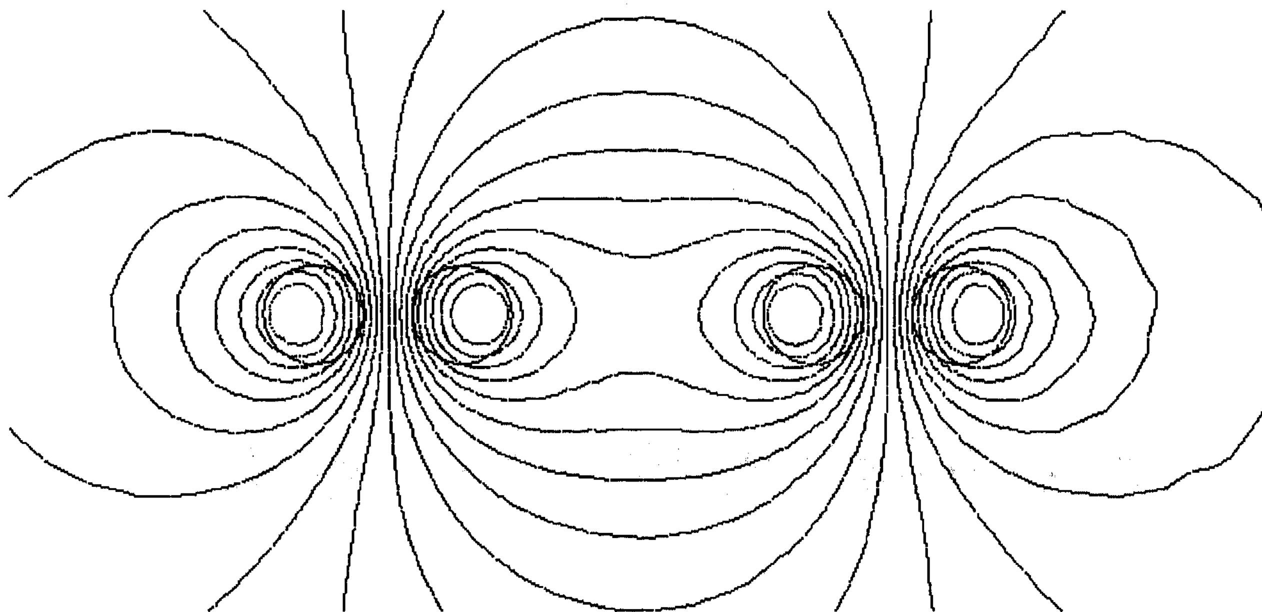


Fig B2 Twisted Wire Pair

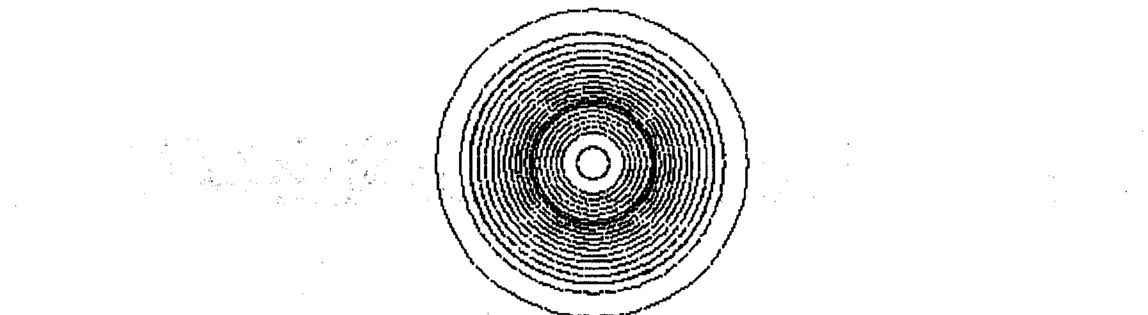


Fig B3 Coaxial Cable

Appendix B cont. - Flux Plots

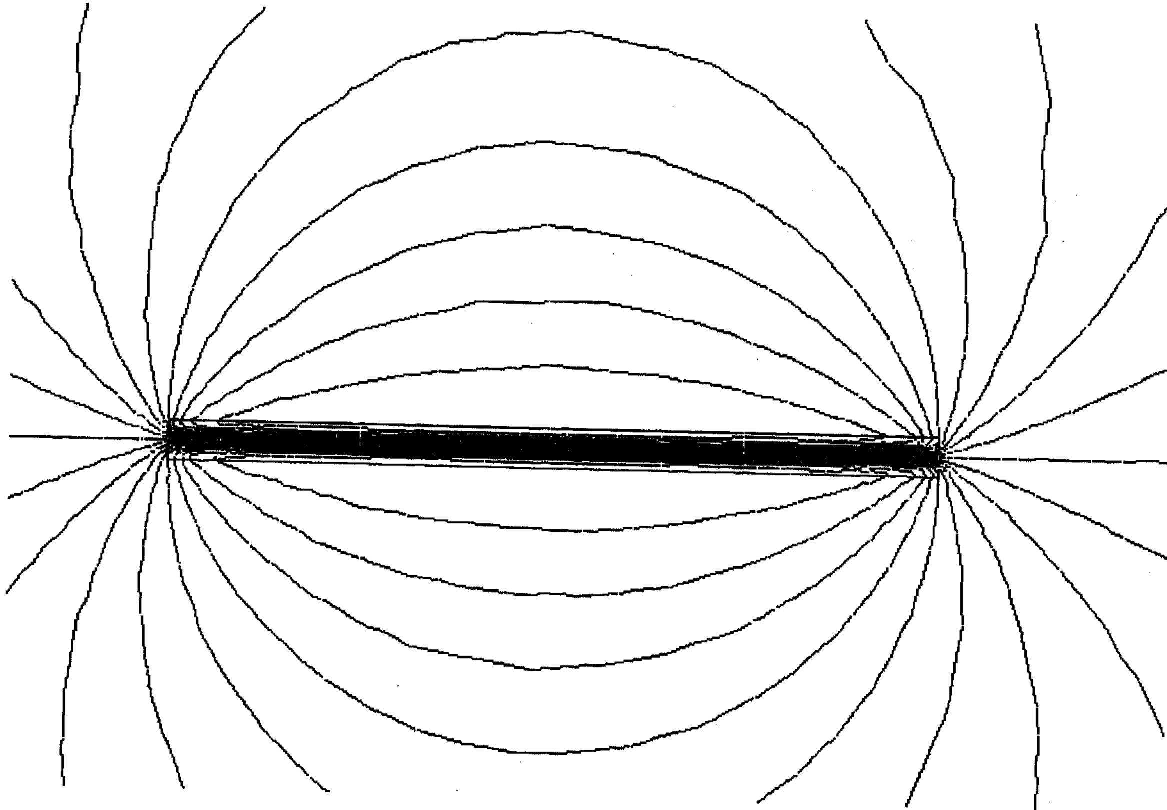


Fig B4 Two Layer Flex

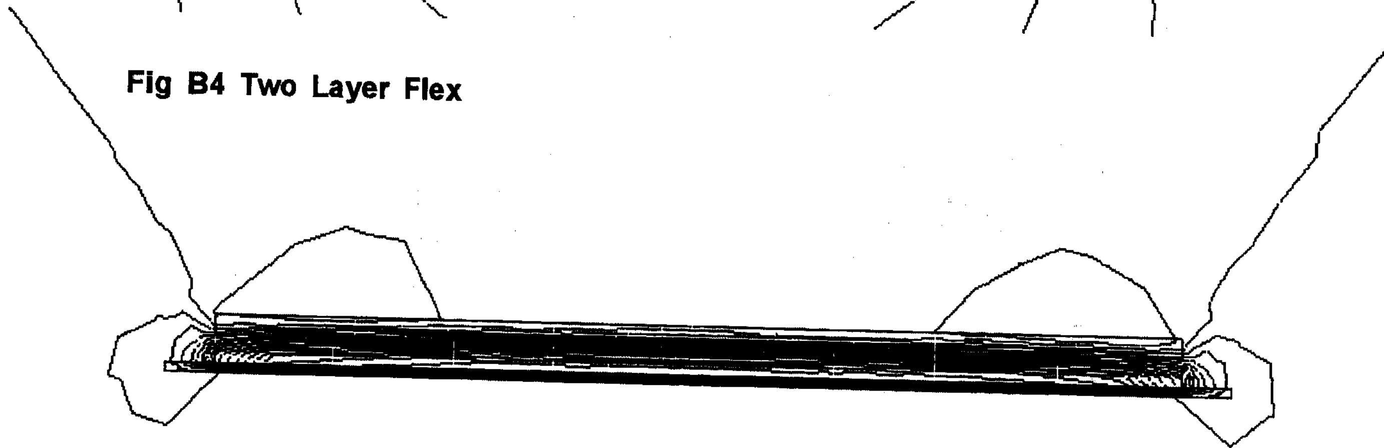


Fig B5 Two Layer Flex plus Back Screen

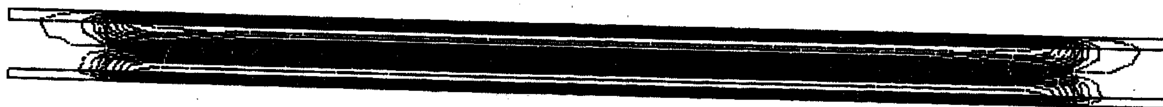


Fig B6 Two Layer Flex Plus Back and Front Screen

970679

Slipping Torque Converter Clutch Interface Temperature, Pressure and Torque Measurements Using Inductively Powered Radiotelemetry

M. J. Throop and D. G. McWatt
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Torque converter clutch friction interface and automatic transmission fluid (ATF) temperatures, pressure difference across the clutch piston, flow through the friction material grooves, and engine crankshaft dynamic torque were measured for typical operating conditions on a running transmission. The friction coefficient, clutch unit pressure, fraction of heat rejected to ATF flowing through the grooves, and time dependent thermal response were determined. Simplified heat transfer calculations were compared with thermal data. Clutch interface temperatures were assessed as they relate to the process of friction material and ATF degradation. The inductively powered radiotelemetry system was found to be a robust and powerful tool for investigating continuously slipping clutch system performance.

INTRODUCTION

In order to gain improved fuel economy, automatic transmission torque converter clutches are designed to bypass the torque path through the fluid coupled impeller-turbine system in favor of a parallel path from the engine directly to the gearbox. When the clutch is employed in an on-off mode of operation, significant improvements in fuel economy can be realized and issues of clutch durability and performance are not much different from those encountered with shifting clutches. On the other hand, even greater overall fuel economy improvements as well as better engine torsional vibration isolation can be realized with a continuously slipping torque converter clutch. Unfortunately, experience has shown that as operating hours are accumulated, continuously slipping clutch systems present a durability challenge because the clutch friction material-ATF pair can acquire a negative friction coefficient slope with respect to slip speed. Once the friction coefficient slope becomes sufficiently negative, the powertrain is likely to exhibit undesirable surge and/or shudder phenomena [1]. The exact means by which a particular friction material-ATF pair degrades (i.e. acquires a negative coefficient of friction slope with respect to slip speed) over time is unknown, but there is evidence that friction material interface temperature plays an important role. Attention has been focused on friction material [2] and ATF compositions [3,4]. The effectiveness of some friction material groove designs has been assessed [5]. It is evident that the slipping speed strategy and clutch system thermal design must have some bearing on the degradation process as well, since these determine the rate of heat release and strongly influence the temperature at the clutch friction material interface.

This paper presents a set of clutch interface temperature and related measurements for a continuously slipping torque converter clutch operating in a realistic powertrain environment. Simplified heat transfer calculations were compared with the results. Friction coefficient slopes with respect to slip speed were examined. Friction interface temperature data for a set of typical operating conditions were used to assess the relative potential for friction material degradation. No attempt was made to degrade the clutch as the measurement process progressed.

This paper also describes the inductively powered radiotelemetry system and its associated instrumentation. The system was a powerful tool for making reliable measurements in a running torque converter without compromising normal powertrain functionality. Its use can be extended to vehicle measurements as well.

EXPERIMENTAL LAYOUT

THE POWERTRAIN consisted of a production 3.8L V-6 engine and four speed automatic overdrive transaxle (Figure 1). The transaxle was modified for dynamometer cell use by eliminating the final drive and

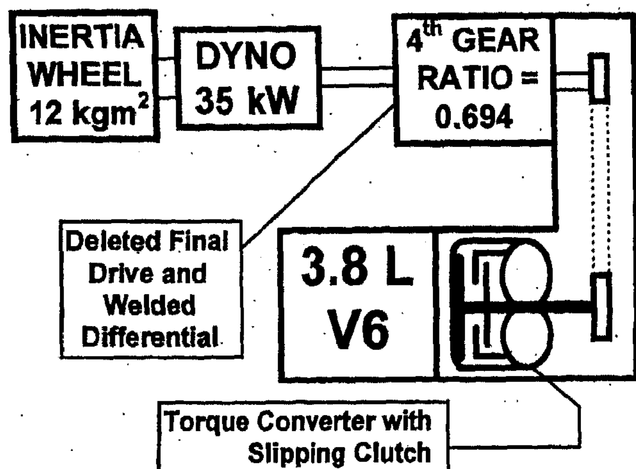


Figure 1 Powertrain layout

locking the output shaft to the final drive input. Following this modification, restoration of the vehicle-equivalent driveline dynamics and the capability of observing shudder required the installation of a special driveshaft with appropriate compliance. Since the work described here was not directed towards observing shudder, a much stiffer conventional driveshaft was used. The driveshaft was connected to an in-line torque transducer (300 Hz bandwidth) which in turn was connected to a 35 kW dynamometer capable of sustained torque absorption at 136 Nm. A 12.4 $kg{\cdot}m^2$ inertia wheel was connected to the other side of the dynamometer by a torsionally stiff coupling. The combined effect of all these modifications made it possible to reasonably simulate the equivalent vehicle speed, load, and inertia.

THE TORQUE CONVERTER ASSEMBLY was a take-apart converter design which was necessary to accommodate pressure and temperature sensor installations (Figure 2). The continuously slipping clutch

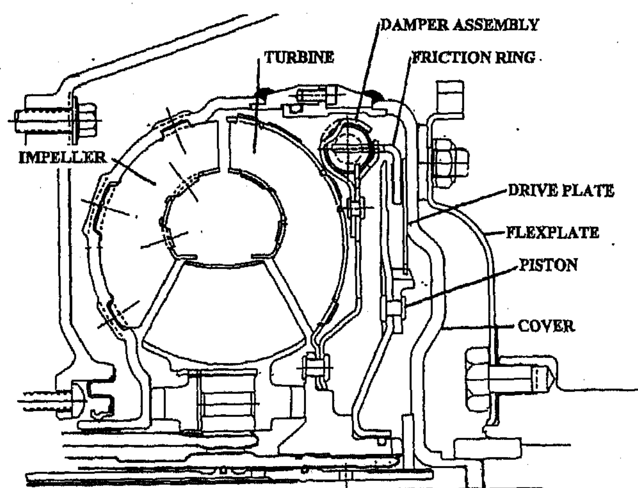


Figure 2 Torque converter cross section

system consisted of a spring damper assembly connected to the turbine hub, which improved engine torsional attenuation during clutch lockup, a friction material ring with grooved friction material on both sides and four tabs that inserted into the spring damper assembly, and a piston assembly piloted on the turbine hub and splined to a drive ring fastened to the converter cover. When the piston pressed on the rear side of the friction material ring which in turn was pressed against the drive ring, the turbine was mechanically coupled to the converter cover. The clutch was operated for locked, unlocked, or modulated slip speed mode by controlling the pressure on the piston release side. The clutch, cover and damper assemblies used in the experiments described here were experimental parts and although similar do not correspond to any production converter clutch system.

THE INDUCTIVELY POWERED TELEMETRY SYSTEM was a joint design and installation with Wireless Data Corp. [6]. It consisted of both stationary and rotating elements (Figure 3). The stationary portion of the system

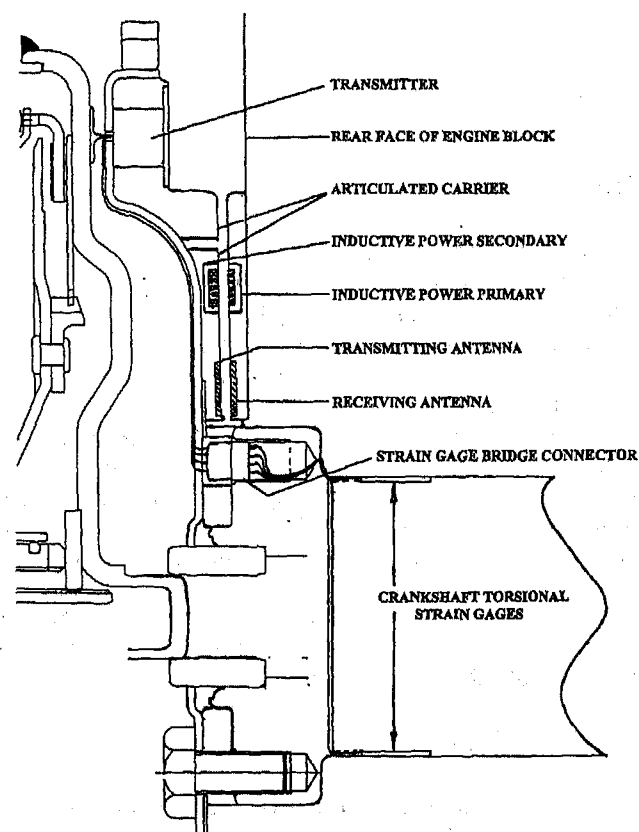


Figure 3 Inductive power and telemetry sensor cross sectional layout

was secured to the rear face of the engine block and included the primary inductive power coil driven at 160 kHz and the radio frequency receiving antenna. The rotating portion of the system was mounted on the torque converter flexplate and consisted of a two piece articulated carrier assembly designed to leave the axial stiffness of the flexplate unaltered. Near its outer perimeter, the rotating carrier assembly contained two pairs of pockets for housing transmitter modules, each 180 degrees opposed to maintain balance. Three pockets contained FM transmitters, two static strain gage type for converter clutch Δ pressure and engine crankshaft torque and one six-channel type for ATF and converter clutch interface

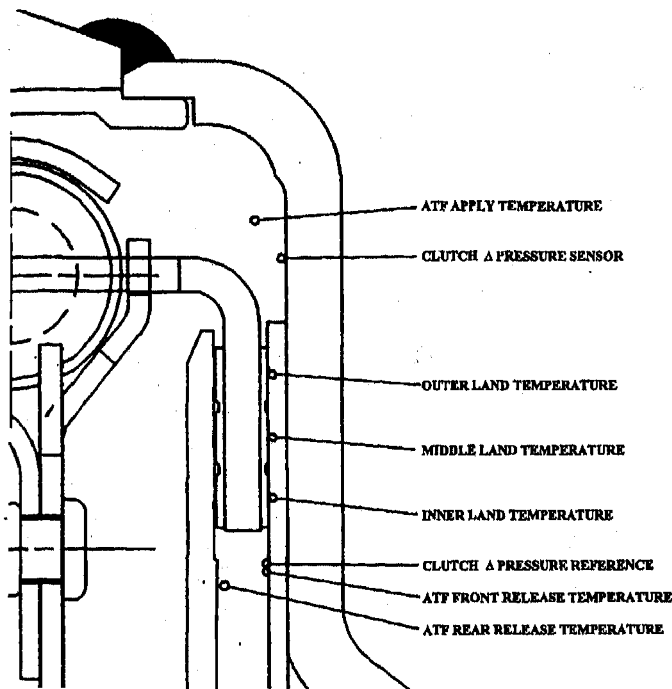


Figure 4 Torque converter cover and drive plate telemetry instrumentation - axial and radial layout

temperature measurements. A counterweight was installed in the fourth pocket. Included on the rotating carrier assembly was the inductive power secondary coil, energized across a small air gap by the inductive power primary coil and in turn powering the transmitter modules and strain gage sensors, and a four piece transmitting antenna (one for each transmitter) relaying sensor information to the stationary receivers via the receiving antenna.

ENGINE CRANKSHAFT TORQUE WAS MEASURED by instrumenting the engine crankshaft rear main bearing journal (Figure 3). Two EDM (Electro Discharge Machined) flats were machined on the rear main bearing journal. The flats were located 180 degrees opposed with the rear main bearing journal oil hole located 90 degrees midway between them. Each flat was connected to the crankshaft rear flange by an EDM groove along the journal surface. One of the grooves continued as a hole through the crankshaft flange to an embedded torque sensor wiring connector. A four active arm torsional strain gage bridge configuration, consisting of two-element high temperature 45° torsional strain gages bonded to each flat with a high temperature epoxy, was employed. Teflon insulated 36 AWG wire was attached to each strain gage set and laid in the journal grooves. From there, the wiring from one gage set was fastened with Nichrome spot welded tabs along the outer edge of the bearing journal to the hole in the rear flange at the end of the other journal groove. At this point all crankshaft torque strain gage sensor wiring continued to the embedded connector in the rear flange. A mating connector continued this wiring on to the flexplate mounted telemetry system carrier assembly and torque sensor transmitter. The strain gage sets, journal grooves, and rear flange wiring hole were coated with a high temperature oil resistant epoxy to assure retention during rotation, restore original oil flow pattern as much as possible, and prevent oil leakage through the rear flange hole.

CONVERTER CLUTCH INTERFACE AND ATF TEMPERATURES were obtained by installing six Type K thermocouples inside the torque converter at appropriate locations in and around the torque converter clutch (Figures 4 and 5). The thermocouple assemblies were 0.75 mm

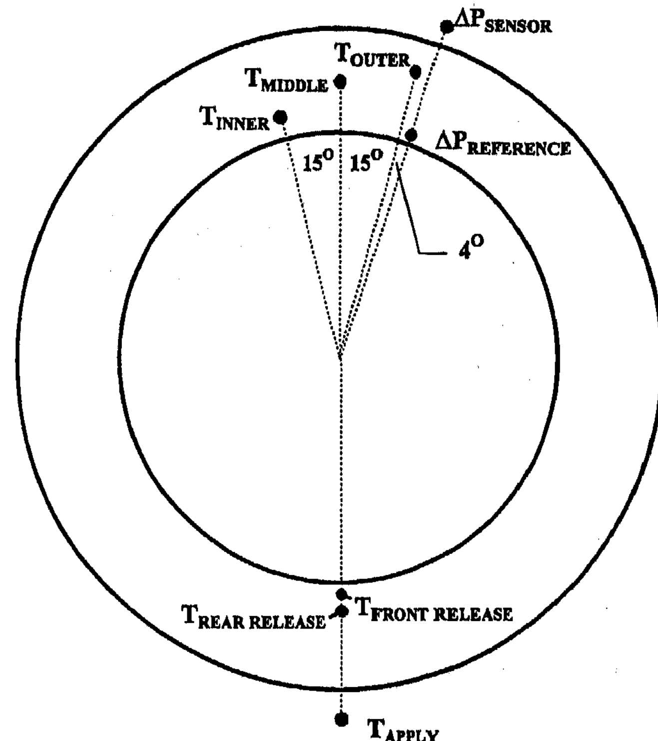


Figure 5 Torque converter clutch and drive plate telemetry instrumentation - circumferential layout viewed from the cover side

diameter stainless steel tubes containing thermocouple wires surrounded by MgO insulating material. The clutch ATF apply temperature thermocouple was mounted in the converter cover and extended into the ATF flow, just before it entered the clutch groove system on the apply side. The converter clutch ATF front and rear release temperature thermocouples extended into the ATF flow above the surface of the drive ring on the release side of the clutch, each sampling ATF temperature as it exited the front and rear friction material groove systems, respectively. The three ATF temperature thermocouple shells were insulated from the drive plate and cover for a distance of at least 10 shell diameters. The friction material-drive plate interface or land thermocouples were routed through EDM grooves in the rear face of the drive plate, entering from the outer edge, and continuing to locations approximately 0.075 mm below the drive plate surface directly underneath the center line of the annular area contacted by each land. After the land thermocouples were fixed in place, the groove volume not taken up by the thermocouple assemblies was filled by welding small sections of fine wire to the thermocouple shell and groove walls. The rear surface of the drive plate

was then restored to its original flatness, the thermocouples were routed through cover penetrations and sealed with epoxy, and the drive plate was bolted to the cover. The three land thermocouple locations were 15 degrees apart (Figure 5).

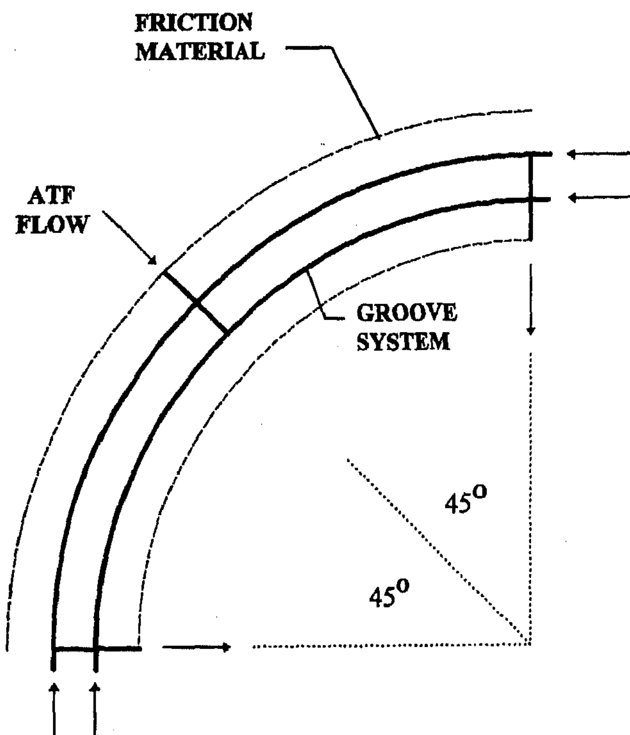


Figure 6 Friction material groove pattern (90° sector)

All thermocouple wiring was bundled and secured to the outside of the torque converter cover with Nichrome spot welded tabs and then guided through the flexplate to the six-channel multiplexed thermocouple transmitter mounted on the flexplate carrier assembly. How much do observed drive plate thermocouple temperatures differ from true drive plate surface temperatures? Since the land thermocouples were located below the drive plate surface, the observed temperatures must be somewhat low. Given that each thermocouple sensing element was located 0.37 mm below the surface of the 1.75 mm thick drive plate, for a worst case power dissipation of 1.8 kW, a friction material contact area of 81 cm^2, and all heat flowing perpendicular to the drive plate surface, the thermocouple was expected to indicate a temperature 1 °C lower than the actual drive plate surface temperature. This is not significant compared to other sources of error.

THE CONVERTER CLUTCH Δ PRESSURE MEASUREMENT was obtained by installing a miniature Kulite strain gage differential pressure transducer in the converter cover radially outward from the ATF apply side entry point into the clutch groove system (Figures 4,5,6). The transducer pressure reference tube was secured along the outside of the converter cover and routed to a location corresponding to the release side of the clutch. A hole was drilled through the cover and drive ring and the reference tube was installed and sealed. The pressure sensor wiring was secured to the outside of the torque converter cover with Nichrome spot welded tabs and then guided through the flexplate to the strain gage pressure sensor transmitter mounted on the telemetry system carrier assembly. The measurement obtained was the pressure difference between the apply and release sides of the torque converter clutch system.

THE CONVERTER CLUTCH BYPASS FLOW MEASUREMENT was obtained by installing a specially modified transaxle valve body that isolated bypass ATF flow after it exited the friction material groove system, directed it to an external Micro Motion flow meter for measurement and then returned it to the transaxle sump. With the modified valve body, clutch slip speed was controllable by manual valve adjustments only and it was not possible to perform transient experiments.

THE DATA ACQUISITION AND CONTROL SYSTEM was utilized to acquire, display, and record all pertinent parameter information and control the engine, transaxle, and dynamometer functions. Steady state data were acquired at 200 Hz, averaged and displayed at 1 Hz, and collected into time averaged data sets for 60 second intervals. A total of 17 temperature, 4 RPM, 3 torque, 15 pressure, 3 flow, and 3 position data channels were monitored, recorded, and displayed. In the dynamic experiments, 16 channels of data were collected at a 5 kHz rate for 60 seconds.

EXPERIMENTS

Time averaged steady-state data were taken at 1200 and 1700 RPM turbine speed and 70 Nm output shaft torque, and at 1500 RPM turbine speed and 40, 70 and 90 Nm output shaft torque. These operating points spanned a representative range of operation for the slipping clutch system. At each operating point, the time averaged torque converter slip speed was varied from zero to 100 RPM in discrete steps, constant output shaft torque was maintained by adjusting the engine throttle, and turbine speed was kept constant by operating the dynamometer in speed control. For constant output shaft torque, the slipping clutch torque decreased with increasing slip speed because the turbine took an increasing portion of the torque split between the clutch and turbine. The clutch torque was derived from output shaft torque taking into account the gear and turbine chain ratios, the gear and chain losses, and then subtracting the turbine torque determined from torque converter performance data. Temperatures were allowed to settle to their steady-state values over a period of several minutes before any data were taken. The transaxle ATF sump temperature was controlled at 70 or 93 °C. At selected operating points, high speed data were acquired for engine and turbine speeds and crankshaft and output shaft torque in order to assess the engine torsional attenuation performance of the slipping clutch

system. The torsional attenuation results are not reported here.

In a second set of measurements, dynamic thermal performance data were taken at one operating point - 1500 turbine RPM and 70 Nm output shaft torque. A step command was input to the pulse width modulated solenoid that controlled the continuously slipping clutch release pressure. Two sets of measurements were made, the first during a step command that corresponded to a zero to 100 RPM slip speed change, and the second during a 100 RPM to zero slip speed change. Data were logged at a 5 kHz rate and included land, apply and release temperatures, engine and turbine speeds, crankshaft and output shaft torque, and Δ pressure across the slipping clutch. The sump temperature was maintained at 93 °C.

In a third set of measurements, time averaged converter clutch release side flow data were taken at all the operating conditions of the first set of measurements. Data were taken for the full range of differential pressures across the clutch, corresponding to both continuously slipping and locked conditions.

RESULTS

A TYPICAL SET OF TIME-AVERAGED LAND TEMPERATURE data is shown in Figure 7 for 1700 RPM turbine speed, 67 Nm output shaft torque, and 93 °C sump temperature. All the clutch land and ATF temperatures were within ± 2 °C of the sump temperature at zero slip speed, then increased as slip speed was increased. The minimum slip speed at which best torsional attenuation was obtained was 20 RPM. ATF exiting the clutch grooves had a higher temperature than that entering the grooves, consistent with a process of heat removal from the drive plate. Assuming equal heat transfer to the ATF from each land, the land temperatures were expected to increase from outer (lowest) to middle (intermediate) to inner (highest), as ATF flowed through the clutch groove system (Figure 6), removing heat and increasing in temperature from one land to the next. However, Figure 7 shows that the outer land temperature was the highest and the inner land the lowest. This result was duplicated at all operating conditions except one high torque condition where the outer and middle land temperatures were the same. Figure 7 also shows the inner land temperature was nearly the same as the ATF front release temperature, indicating heat rejection from this land to the ATF was minimal.

The most likely explanation for the observed land temperature distribution is unequal friction element contact pressures on each land, with the pressure highest at the outer land and lowest at the inner land. No data were available for running land-to-land contact pressure differences in the system reported on here. It is

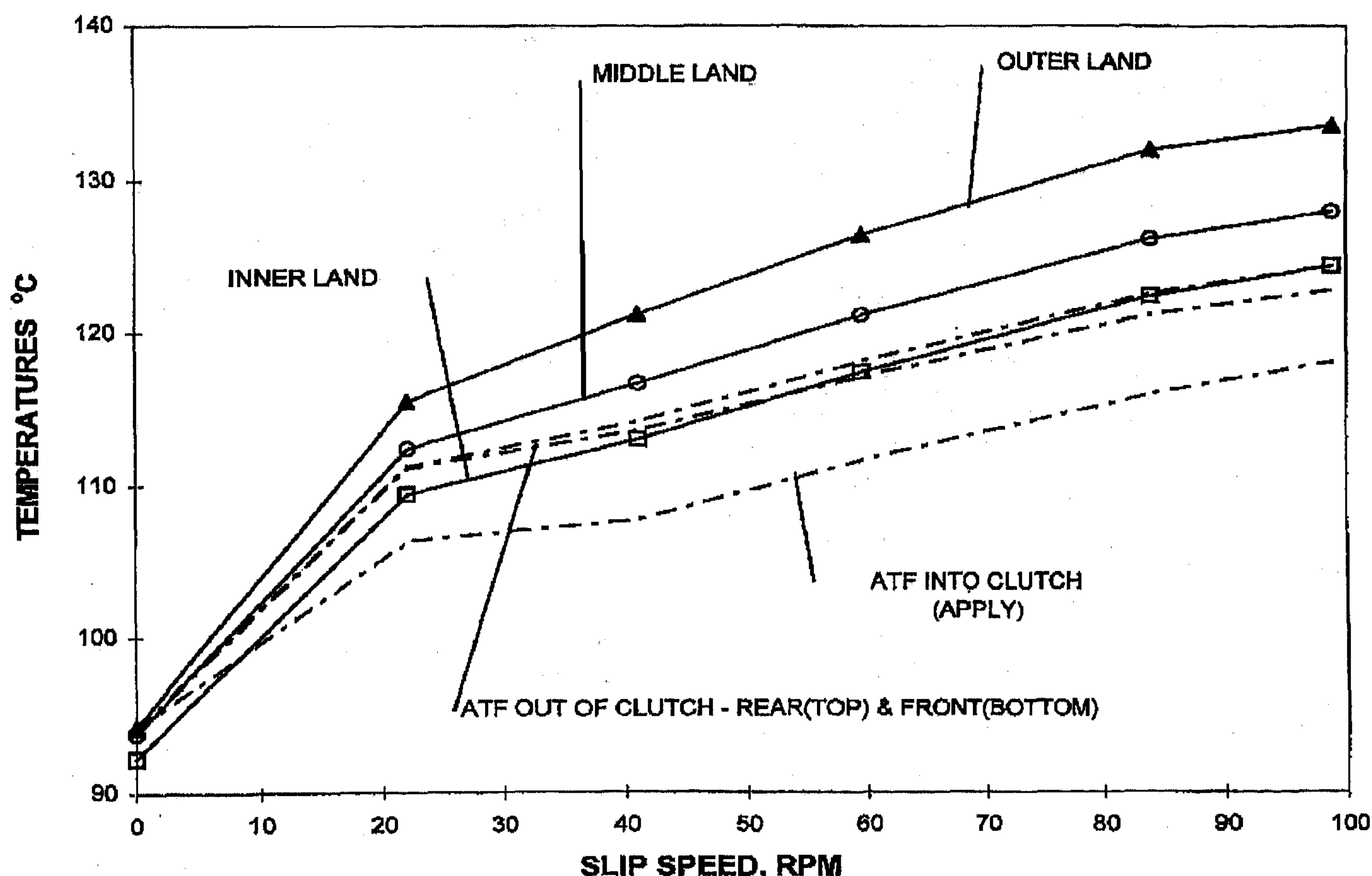


Figure 7 Land and ATF temperatures vs. slip speed

known that the converter housing changes shape as RPM and other operating conditions change, so the drive plate, bolted to the front cover, must change shape as well. Similarly, the piston deforms as a function of the pressure difference across it. These two effects can create a geometry in which the piston and drive surfaces opposite the facing friction surfaces are not coplanar, leading to uneven contact pressures at some operating conditions. Uneven contact pressures at each land imply the heat generated will not be the same, and assuming a constant friction coefficient at each land, it follows that the land temperatures will not be equal. Slipping clutch designs can attempt to remedy this situation by tapering the drive plate and piston surfaces such that they will be coplanar, but only at a chosen operating condition, for example where maximum clutch torque capacity is desired. Inspection of the drive plate and piston surfaces after 125 hours of operation revealed a progression (on both piston and drive plate) from least wear at the inner land to greatest wear at the outer land. This radial wear pattern suggested that average clutch operation must have occurred at conditions where the friction material contact pressure increased from inner to outer land, and the land-to-land temperature differences observed, as in Figure 7 for example, are consistent with this. The inspection also showed that the 12 bolts attaching the drive plate to the cover slightly distorted it such that the drive plate had a vertical variation of about ± 20 microns at 30 degree intervals on all three lands. A circumferential wear pattern, consistent with the drive plate vertical variation, was also observed, although it was not as strong as the radial wear pattern. It is possible that some portion of the land-to-land temperature differences were due to thermocouple locations at drive plate peaks or valleys. As installed, the outer land thermocouple was located near a 20 micron high spot, the inner land near a -20 micron low spot, and the middle land near a spot of average height.

INCREASED APPLY ATF TEMPERATURE was another feature of Figure 7 that was observed at all operating conditions as the slip speed increased from 0 to 100 RPM. Was ATF heated in the impeller-turbine part of the torque converter before some of it migrated to the slipping clutch and entered the groove system? This question can be answered by measuring the rise in converter-out temperature, defined as the ATF temperature exiting the impeller/turbine system referenced to the sump temperature, as a function of slip speed. At 1500 RPM and 70 Nm output shaft torque, Figure 8 shows that the converter-out minus sump temperature rise was constant within ± 2.5 °C over a 0 to 100 RPM slip speed range, compared to the observed apply minus sump ATF rise of 24 °C (Figure 7). The converter-out temperature rise was also calculated, using known converter torque performance and flow data, to be 1.1 °C. Thus an explanation for the ATF rise with slip speed must be sought elsewhere.

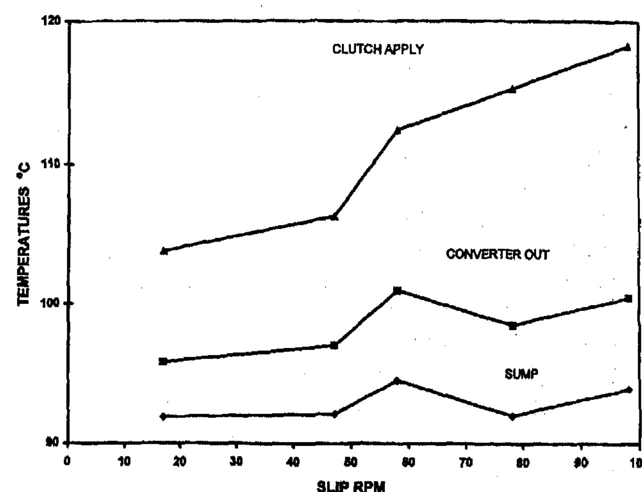


Figure 8 Apply, converter out and sump ATF temperatures vs. slip speed

Another possible cause of apply ATF temperature rise is heat transfer from the torque converter cover and the piston rear face. Apply ATF originated from the primary converter flow circuit via leakage past the impeller - turbine gap and leakage through gaps in the blade retention slots in the turbine shell (Figure 2). The geometry involved in possible apply ATF heating is shown in Figure 9, which indicates how heating could occur as apply ATF flowed past the inside surface of the cover toward the entrance of the front friction material groove system. A similar ATF flow past the rear face of the piston prior to entry into the rear friction material groove system could also take place. Land temperatures were not measured at the interface of the piston and friction material rear face. The actual apply ATF flow paths were not known. The piston and cover contributions to apply ATF heating may differ due to geometry and flow path differences.

An estimate of apply ATF heating by the cover and piston can be made for the case of Figure 7, assuming some reasonable areas, flow rates and flat plate heat transfer coefficients. The estimated apply ATF heat removal rate from the cover and piston was

$$Q = F \rho C_P (T_{APPLY} - T_{SUMP}) \quad \text{Watts}$$

where

F = ATF flow rate through friction material groove system (0.5 l/min.)

ρ = ATF density (8 g/cm^3)

C_P = ATF specific heat capacity (2160 J/kg/K)

to solve for $T_{APPLY} - T_{SUMP}$ required a heat transfer calculation for

$$Q = \alpha A (T_{COVER} - (T_{APPLY} - T_{RELEASE}) / 2) \quad \text{Watts}$$

where

$\alpha = 2 K (0.332 Pr^{0.33} Re^{0.5}) / L$

K = ATF thermal conductivity, 0.14 W/m/K

Pr = the Prandtl number, 93 for ATF at 93 °C

Re = the Reynolds number, 1.86E4 for ATF at 93 °C, estimating the ATF velocity to be

the cover heating area speed with respect to the turbine outside diameter at 100 RPM slip

L = characteristic length, chosen to be 4 cm

A = area which heated apply ATF (a 26 cm diameter, 4 cm wide ring on the cover, plus a piston area of 415 cm^2 or 742cm^2 total area)

T_{COVER} was taken to be the average land temperature

$T_{RELEASE}$ was taken to be the average ATF release temperature

For α = 571 W/m^2/K, the calculated temperature rise ΔT (apply minus sump) was 44 °C compared to the measured temperature rise of 23 °C, which is acceptable agreement considering the difficulty in estimating the Reynolds number and heating area. The heat transfer calculation lends support to the idea that the front cover and rear surface of the clutch piston were responsible for heating the ATF prior to its entry into the friction interface groove system. This heating process had the undesirable effect of raising the land temperatures and the potential for accelerating friction material-ATF pair degradation.

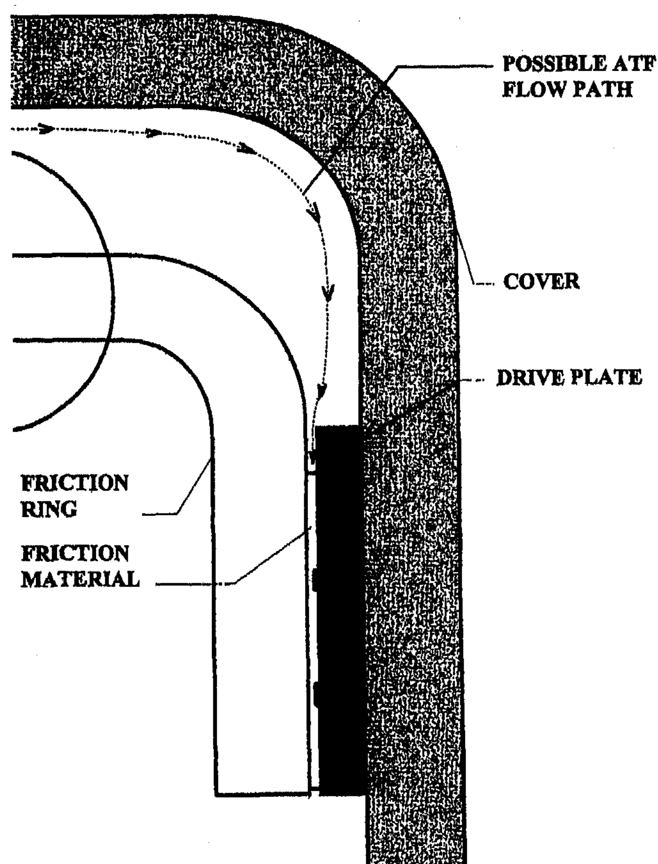


Figure 9 Geometry for apply ATF heating

HEAT TRANSFER TO ATF FLOWING IN THE GROOVES was qualitatively estimated and compared with experiment. At the conditions of Figure 7, the total ATF flow rate was 0.5 l/min. The mean ATF speed in a single groove sector (in Figure 6, a 45° sector of a circle with 215 mm mean radius) was 0.9 m/s. The flow was laminar (Re = 51), but it was not fully developed (in addition to entrance, exit and bend/tee effects, the drive plate wall was moving, the groove was curved, and centrifugal forces acted on the ATF). For fully developed laminar flow, the heat transfer coefficient α depends only on the ATF thermal conductivity K and the mean hydraulic diameter D

$$\alpha = 4.12\, K / D \qquad W/m^2/K$$

Using the groove dimensions given in a following section on groove flow, α was 1377 W/m^2/K. The calculated ATF heat removal rate was

$$Q = \alpha A\, (T_{LAND} - (T_{APPLY} + T_{RELEASE}) / 2) \quad Watts$$

where

A = heat transfer area, taken to be that area of the drive plate and piston exposed to groove ATF fluid flow (27.5 cm^2)

T_{LAND} was taken to be the average land temperature

In the case of the Figure 7 data at 100 RPM slip, the calculated ATF heat removal rate was 30 W, compared to the experimental heat removal rate

$$Q = F \rho C_P\, (T_{RELEASE} - T_{APPLY}) \qquad Watts$$

which was 74 W. It appears that the fully developed laminar flow assumption leads to a significant underestimate of groove ATF heat removal.

HEAT REMOVAL AS A FUNCTION OF CLUTCH POWER for groove ATF flow is plotted in Figure 10 as a percentage of total clutch power dissipation as slip RPM was swept for three output shaft torques at 1500 RPM. The percent power carried away by the grooves decreased significantly as slip speed increased, so the groove system became less effective in cooling at high slip speeds. In contrast, as the clutch torque increased at a fixed slip speed, the percent power carried away by the grooves also increased, so the groove system became more effective in cooling at high torques. Directionally, decreasing slip speed and increasing torque are both associated with increased ΔP across the clutch and therefore increased flow though the clutch grooves and improved heat transfer. The proportion of heat rejected to the ATF before it entered the groove system (apply ATF heating) generally followed the groove ATF trends. These two heat removal processes combined removed less than half the clutch power at 100 RPM slip. Heat removal by cover conduction is a likely explanation for the remainder. Since cover conduction increases with increasing land temperature and slip speed, it removes a larger proportion of clutch power at the conditions where groove flow and cover heating of ATF are less effective.

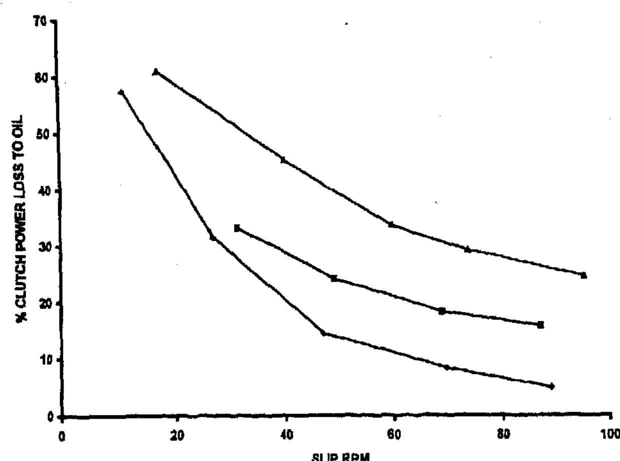


Figure 10 Heat rejected to ATF flowing through the grooves vs. power dissipated in clutch at 1500 RPM, 55 Nm output shaft torque (diamonds), 95 Nm (squares), 125 Nm (triangles)

Outer land temperatures as a function of clutch power (slip RPM times clutch torque) for all 93 °C sump temperature operating conditions are shown in Figure 11. The outer land temperature was always the highest land temperature. The highest temperature observed in the experiments reported here was 163 °C at a clutch power of 1.63 kW. The outer land temperature data showed a trend of increasing temperature with increasing clutch power. Figure 11 is a useful way to assess the susceptibility of a particular clutch design and friction material to temperature degradation. Assuming the land temperature is determined by three heat removal processes - ATF flowing through the grooves, conduction through the cover to other cooler parts of the transaxle, and convective heat loss from the cover and piston rear face to the apply ATF, a generic relationship for land minus sump temperature ΔT can be written

$$\Delta T = Q / (\alpha_G A_G + k A_C / t + \alpha_{CV} A_{CV})$$

The first two terms in the denominator are groove flow and cover conduction terms which are fixed by the ATF properties and the geometry. If only these terms were important, a linear relationship between land temperature and clutch power Q would be expected. The third term is a convective term which has a Reynolds number dependence, implying a correlation with slip speed. Since higher levels of clutch power dissipation are also correlated with slip speed, the presence of a significant convective term in the denominator would reveal itself as a less than linear dependence of land temperature on clutch power.

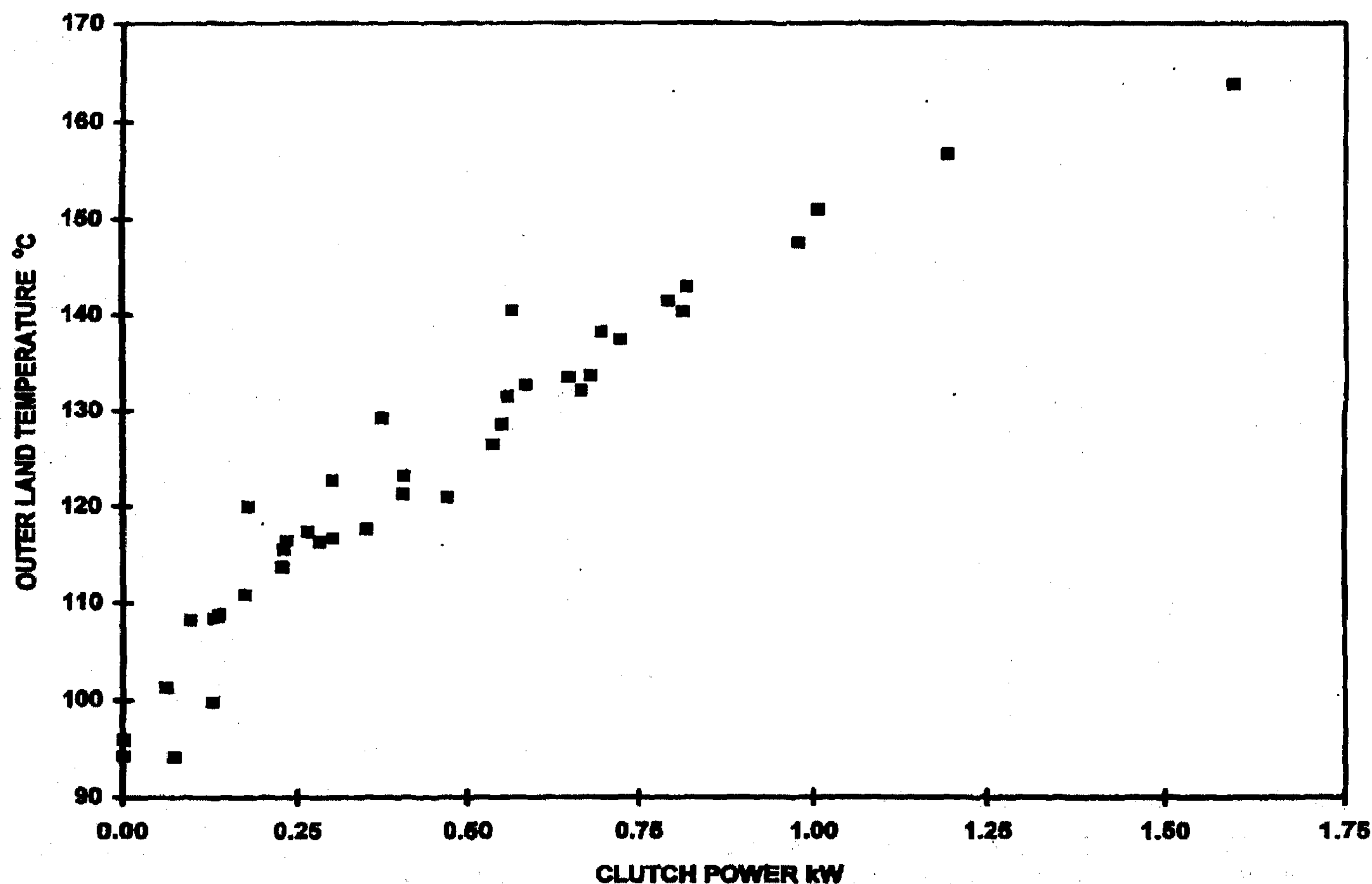


Figure 11 Outer land temperature vs. clutch power, 93 °C sump

Figure 11 does indicate a less than linear trend, consistent with the expected influence of convective heat removal processes. The effect of heat removal by ATF flowing through the grooves alone can be isolated as shown in Figure 12, where the outer land temperature less ATF apply temperature is plotted as a function of clutch power; data for 73 and 93 °C sump temperatures are included. As with Figure 11, there is some evidence for convective heat removal processes, but sump temperature does not seem to have an effect.

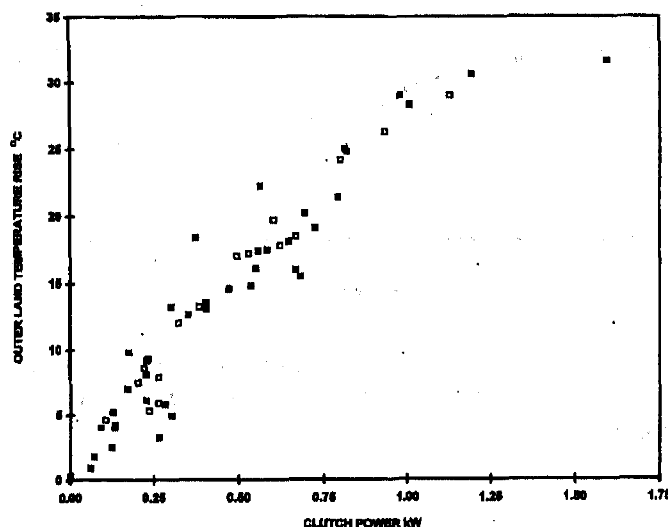


Figure 12 Outer land less apply temperature vs. clutch power, dark squares 93 °C sump, open squares 73 °C sump

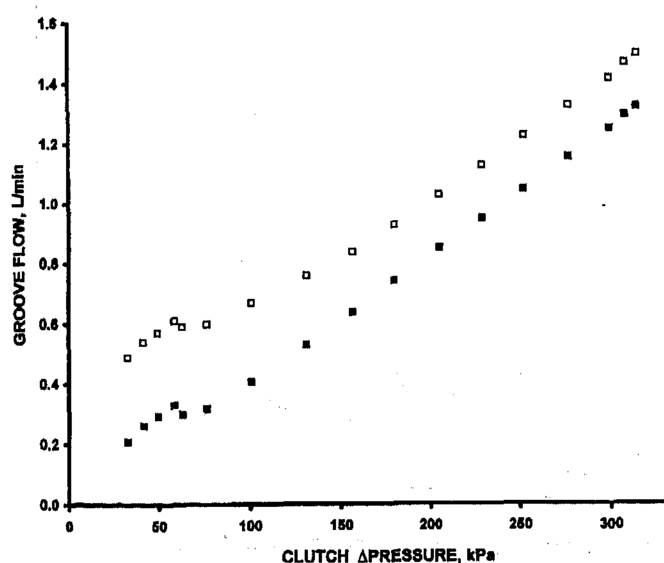


Figure 13 Clutch groove flow vs. clutch Δpressure. Open squares are observed flow, filled squares are calculated flow.

FLOW MEASUREMENTS - Time averaged groove system ATF flow data taken at 1500 RPM for 70 Nm output shaft torque at 93 °C sump temperature are shown in Figure 13. The ATF flow rate through the friction material grooves was estimated using several simplifying assumptions - fully developed laminar flow, neglect of minor losses such as entrance and exit effects, and ignoring the small effect of bends and short radial sections. The groove flow system (Figure 6) was thus reduced to 8 quadrants (4 on the forward friction material surface and 4 on the rear facing surface) containing 4 grooves, two in which ATF flowed in the same direction as the slipping drive plate and two in which ATF flowed in the opposite direction. Each groove was approximated as a straight passage of length equal to one eighth of the average groove circumference. The calculated [9] flow rate F was

$$F = [\, wh^3 / (12\mu L) \{ 1 - 192h / (\pi^5 w) \tanh(\pi w/2h) \} \Delta P \pm whV/2 \,] N_G$$

where

- w = groove width
- h = groove depth
- μ = ATF viscosity
- L = groove length
- ΔP = clutch apply - release pressure
- V = average drive plate velocity with respect to groove flow
- N_G = number of grooves with flow in the same direction

and the ± sign indicates whether the drive plate velocity is in the same or the opposite direction to the flow driven by ΔP. The groove pattern of Figure 6 had equal numbers of grooves for the +V and -V cases, so the V dependent flow contributions canceled, leaving no net contribution to groove flow. Therefore, using the approximations described here, the groove flow depends, apart from geometrical factors, only on the ATF viscosity (a function of temperature) and the apply minus release pressure ΔP.

The groove width was 1 mm, the depth was 0.265 mm, the viscosity varied from 7 cP at 90 °C to 4.6 cP at 114 °C, the average groove length was 84.4 mm, the ΔP ranged from 20 to 375 kPa, and the total number of grooves was 32. The calculated result for a 93 °C sump temperature is shown in Figure 13 along with the measured flow data. At ΔP > 75 kPa the land temperatures were the same as the sump temperature because the slip speed was zero, and for ΔP < 75 kPa, where the clutch was slipping and the ATF was heated as it flowed through the grooves, the flow rate increased due to a reduction in viscosity. The calculated and observed flows are in qualitative agreement for ΔP > 75 kPa. A good flow prediction capability is desirable when the clutch is slipping, but Figure 13 shows the calculated and observed flows differed by about a factor of 2 at ΔP < 75 kPa. Qualitatively, however, there is a clear trend of flow proportional to ΔP and flow increase due to reduced viscosity when the clutch was slipping.

FRICTION COEFFICIENTS - Friction coefficients for various slip speed and clutch torque conditions were calculated according to the relation defined in [7]. The results are shown in two ways, friction coefficient as a function of average load pressure on the friction material (Figure14) and as a function of slip speed (Figure 15). However, as discussed above, wear evidence and land temperature differences showed that the loading of the drive plate and piston were not uniform, so the load pressure referred to here can only be understood in an average sense. In Figure 14 the average load pressure data show that the coefficient of friction increased by about 20% as average load decreased from 650 to 200 kPa. At lower load pressures, clutch ΔP and torque were small and therefore more difficult to measure, causing more scatter in the calculated friction coefficient. The dependence of friction coefficient on load pressure shown in Figure 14 is similar to what is seen for many sliding materials [8].

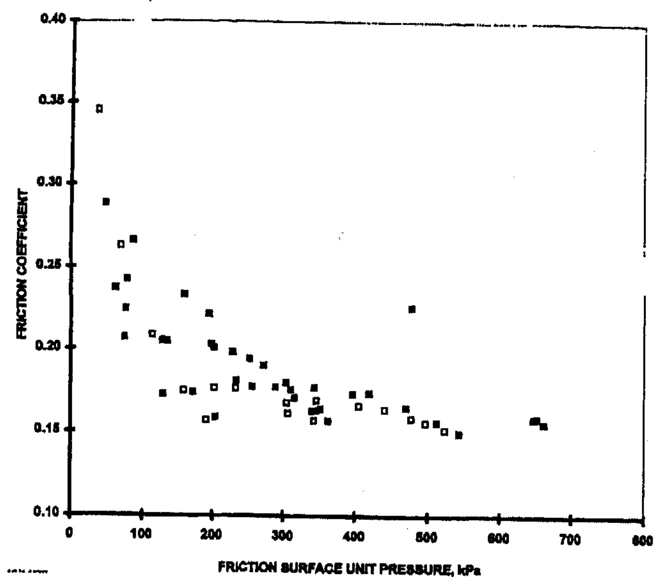


Figure 14 Friction coefficient vs. average load pressure. 73 °C sump (open squares), 93 °C (filled squares)

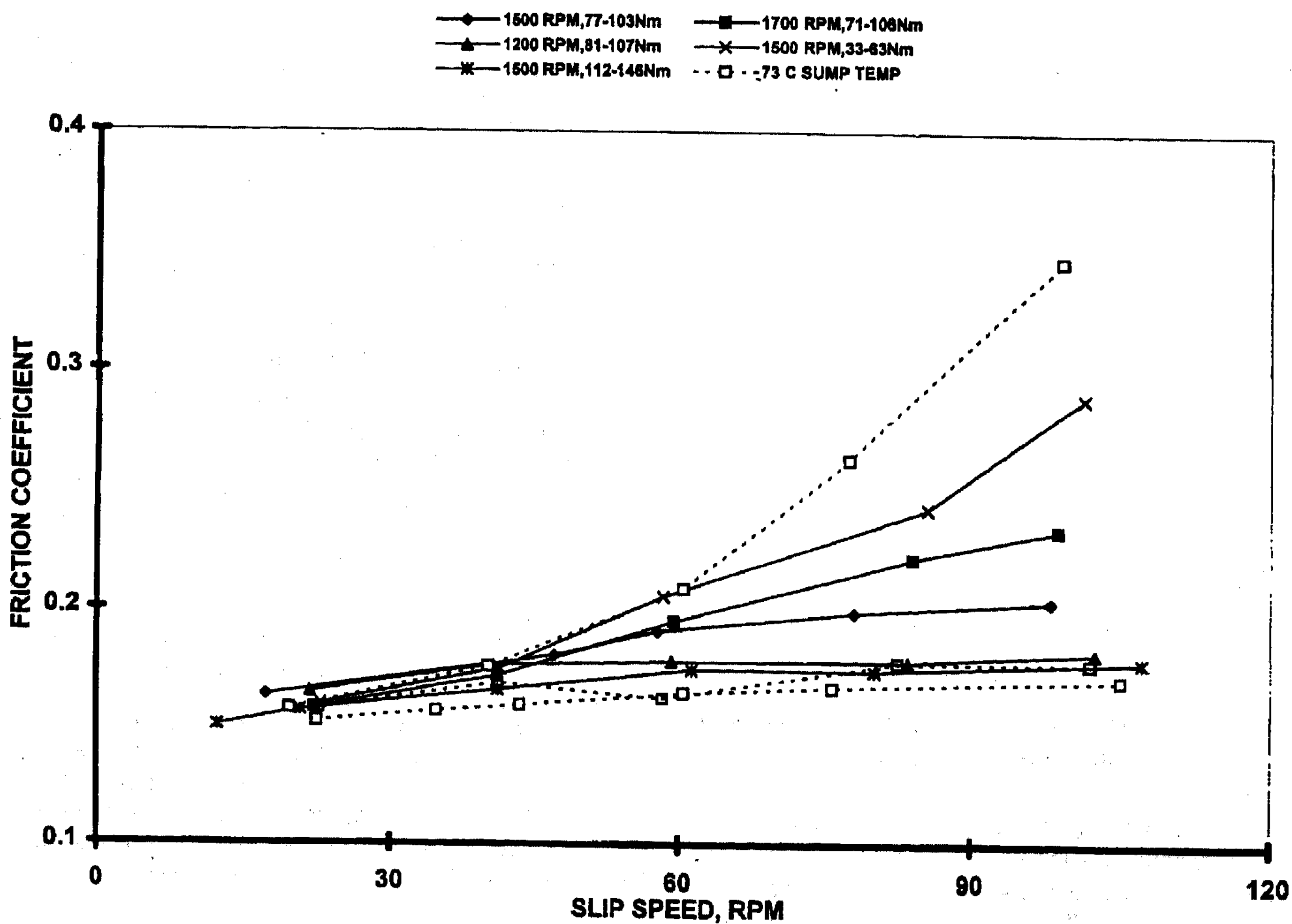


Figure 15 Friction coefficient vs. slip speed

Figure 15 shows friction coefficients for slip speed sweeps at constant output shaft torque and at 73 and 93 °C sump temperatures. The most important feature of Figure 15 is that the slope of the friction coefficient as a function of slip speed was everywhere flat or positive, as was expected for an undegraded friction material - ATF pair. Any friction material degradation leading to a negative friction coefficient slope was not expected for the duration of the experimental conditions reported here. Figure 15 also shows the friction coefficient decreased with increasing clutch torque (or average load pressure as in Figure 14) at a fixed slip speed, but the decrease was less evident at lower slip speeds. At 73 °C sump (for 1500 RPM only) the friction coefficient decreased even more rapidly with decreasing slip speed at a low output shaft torque (40 Nm) but was nearly flat at higher torques (70 and 90 Nm). The results of Figure 14 and 15 show that it is not possible to specify a universal friction coefficient value for all operating conditions.

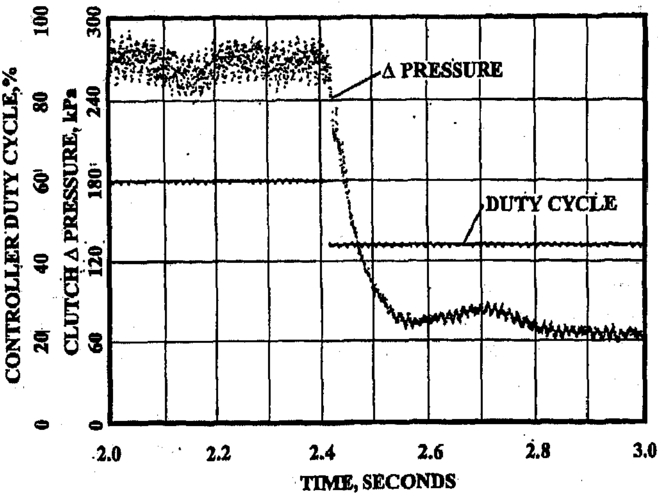


Figure 16 Clutch delta pressure following a step change in controller duty cycle.

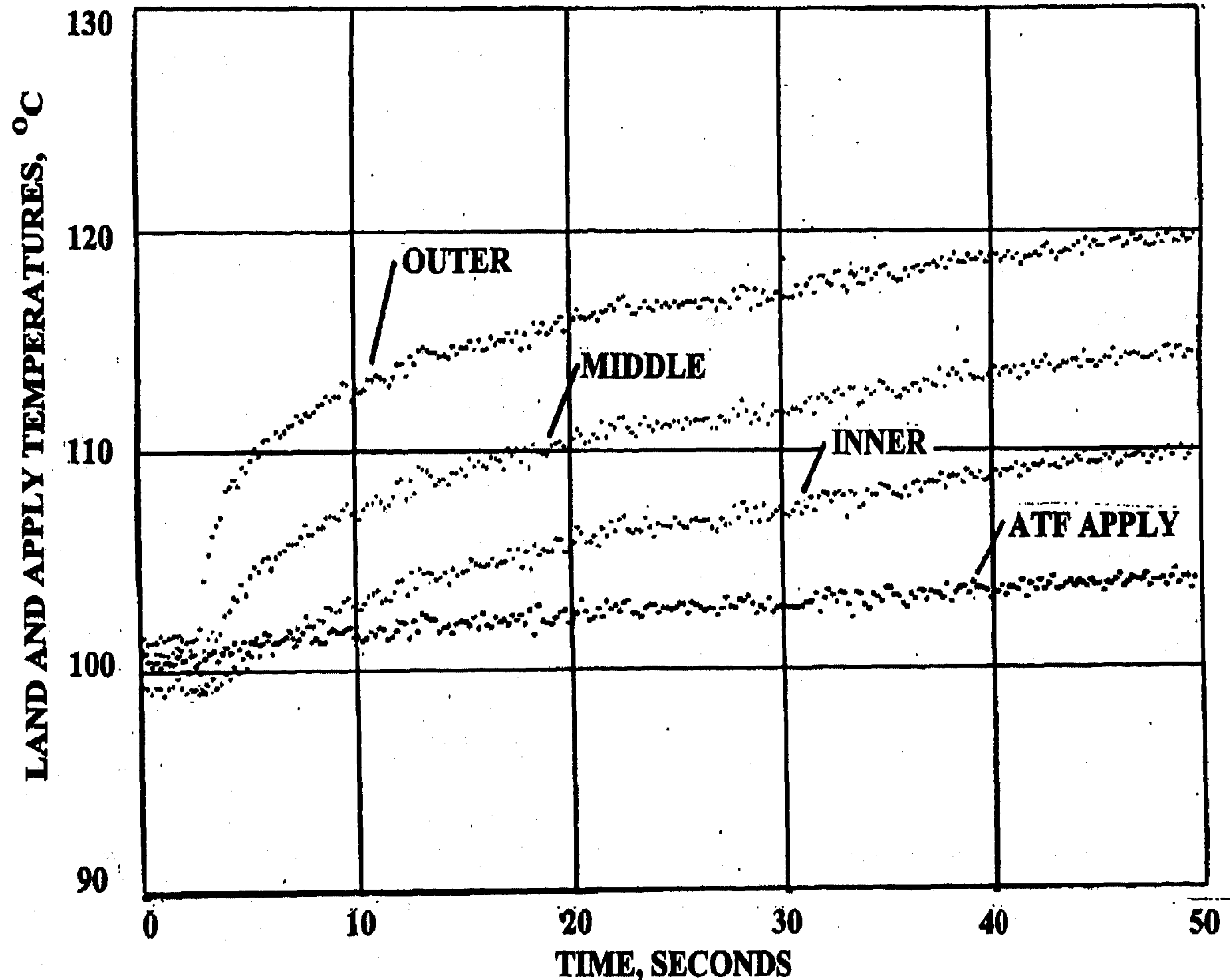


Figure 17 ATF apply and land temperatures following a step change in clutch Δ pressure

DYNAMIC TEMPERATURE MEASUREMENTS - An insight into mechanisms that contribute significantly to land temperature cooling can be gained by injecting a step change in heat flow into the system. An ideal heat flow step change was approximated by commanding a step change in the pulse width modulation of a solenoid controlling the ΔP across the continuously slipping clutch, which had the effect of stepping between an initial and final slip RPM. A step change also took place in the clutch torque, since the torque split between the clutch and the turbine depends on slip speed. The commanded percent pulse width modulation of the solenoid and the clutch ΔP response are shown in Figure 16 as a function of time. The clutch ΔP has a response close enough to a step for the purpose of understanding what happens to land and ATF temperatures following this step. The three land temperature responses are overlaid and shown with the apply temperature as a function of time in Figure 17, where the step command took place at 2.5 seconds. The three lands responded with a rapid temperature rise for a few seconds after the step, then continued to rise at a progressively slower rate for several minutes. After thermal equilibrium was reached, the final outer land temperature, for example, was 137 °C.

The temperature-time dependence was symmetric with respect to sign of the step change. As with the steady-state land temperature data, the transient land temperatures increased in the order inner, middle and outer. For the inner land, the rapid rise following the step input was significantly reduced compared to the other two lands. The rapid rise part of the temperature response is attributed to heating of that part of the drive plate between the friction material and the thermocouple

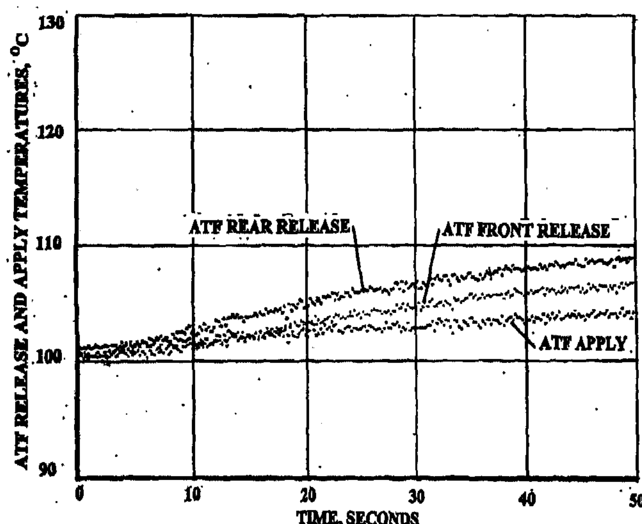


Figure 18 ATF apply and release temperatures following a step change in clutch Δ pressure

sensing element. The ATF apply and front release temperatures are shown in Figure 18. Neither shows the fast temperature rise component evident in the land temperature data. The slow rate of increase in apply temperature is consistent with the process of cover heating by the drive plate, followed by ATF heating by the rear of the piston and the cover upstream of the groove entrances. The process of cover heating is slow, because heat must be stored in the cover mass to raise its temperature. Assuming the front half of the torque converter cover plus piston weighed 5 kg, was isothermal, and using steady-state land temperature and clutch power data for the initial and final conditions of the step, the exponential time constant of the temperature rise was estimated to be 110 s. This is a lower limit, because not all the clutch power was available for cover and piston heating; at this operating condition, a substantial amount was carried away by apply and groove ATF flows. This time constant estimate was qualitatively consistent with the observed ATF temperature time constant of about 300 s.

CONCLUSIONS

- The inductively powered telemetry system was a powerful and reliable tool for measuring torque converter clutch performance under realistic operating conditions.
- Continuously slipping clutch interface temperatures were measured for each land over the clutch operating range on a running transaxle.
- Land temperatures differed significantly, decreasing in temperature from the outer to middle to inner land. The degree of wear observed on the piston and drive plate correlated with land temperature. When there are significant differences in land temperatures, it is not meaningful to use a single interface temperature to characterize the clutch thermal behavior.
- The highest land temperature observed was 163 °C at a clutch power dissipation of 1.63 kW. Friction material-ATF combinations exposed to the high end of the temperatures observed will degrade (i.e. acquire a negative slope coefficient of friction with respect to slip speed) over time.
- Land and land minus apply ATF temperatures increased nearly linearly with increase in clutch power dissipation. This was consistent with heat being removed from the lands primarily by conduction and laminar groove flow. The data were also consistent with some lesser heat removal by convection.
- Laminar heat transfer estimates for the groove and drive plate system were lower than experiment by a factor of roughly two.
- ATF was heated before entering the groove system. Estimates of cover and piston heating effects could account for the apply ATF temperature rise.
- Friction coefficient data were consistent with positive or flat slope with respect to slip speed. An increase in friction coefficient with decrease in average load pressure was observed.
- Friction material groove flow rate was generally proportional to ΔP and increased as clutch interface temperature or sump temperature increased. The

flows measured were consistent with a laminar flow model estimate.
Dynamic land temperature data were consistent with heating of the torque converter cover plus piston, with a fast temperature rise followed by a much slower temperature rise. The time constant for cover heating could be qualitatively reproduced.

ACKNOWLEDGMENTS

The authors thank V. Lambson, now of Lambson Engineering, for his skill and dedication in designing, building and installing the telemetry system in 120 days, and S. Bill of Wireless Data for expediting this process. F. MacDonald, S. Rosen, and J. Ward supplied the specially modified transaxle hardware. J. Kovach assisted with data acquisition and analysis. We also thank W.E. Tobler for his unwavering advocacy of and support for advanced torque converter instrumentation over the years.

REFERENCES

1. T. Hiramatsu, T. Akagi, H. Yoneda, "Control Technology of Minimal Slip Torque Converter Clutch", SAE paper 850460 (1986).
2. R.C. Lam and Y. Chen, "Friction Material for Continuous Slip Torque Converter Applications: Anti-Shudder Considerations". SAE paper 941031 (1994).
3. R.F.Watts, R.K. Nibert and M.Tandon, "Anti-Shudder Durability of Automatic Transmission Fluids: Mechanism of the Loss of Shudder Control", 10th International Colloquium, Technische Akademie Esslingen, Vol 2, p 1341, (1996).
4. H. Ohtani, R.J. Hartley, D.W. Stinnett, "Prediction of Anti-Shudder Properties of Automatic Transmission Fluids Using a Modified SAE No. 2 Machine". SAE paper 940821 (1994).
5. H. Ohnuma and K. Kato, "The Effect of Groove Pattern of Paper Friction Plate on Its Life". SAE paper 910804 (1991).
6. Wireless Data Corporation, 620 Clyde Ave.,Mountain View CA 94043.
7. G.R. Smith, W.D. Ross, P.L. Silbert and W.B. Herndon, "Putting Automatic Transmission Clutch Friction Researchers on Speaking Terms", Design Practices - Passenger Car Automatic Transmissions, p. 125 (1988) SAE Publication.
8. K.C. Ludema, Tribology - University of Michigan Engineering Short Course notes (1991).
9. G.R. Keller, Hydraulic System Analysis, 4^{th} Ed., LCCN 78-52991, and H.E. Merritt, Hydraulic Control Systems, John Wiley and Sons, New York (1967) LCCN 66-28759.

970680

Flow Visualization and Measurement of Torque Converter Stator Blades Using a Laser Sheet Lighting Method and a Laser Doppler Velocimeter

Hisashi Watanabe and Tetsuo Kurahashi
Toyota Central R&D Labs., Inc.

Masahiro Kojima
Toyota Motor Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A new experimental apparatus to visualize and measure the flow in the stator of a torque converter is proposed. A one-sided coaxial shaft constructed of an input shaft and an output shaft provides an open space inside the stator shaft for measurement. Through the window on the stator shaft, the flow in the stator can be directly observed. We also improved the laser sheet lighting method into the blade passage by using a mirror inside the blade.

By visualizing the flow with the laser sheet lighting method, we found that the flow around the leading edge has different separation regions along the blade span. Furthermore, by using a laser doppler velocimeter, velocity vectors and turbulence intensities were measured in three stator blades of different thicknesses with the same camber line. The thickness of the stator blades affects the flow patterns. Both the outlet flow rate and the average outlet flow angle show similar characteristics to the torque capacity coefficient when the stator blade thickness is altered. We have verified that the general flow characteristics are determined by the outlet flow rate, and are emphasized by the average outlet flow angle.

INTRODUCTION

In recent years, there has been a strong requirement for improvement in both fuel economy and acceleration performances of automatic transmissions. It is well known that a torque converter's characteristics significantly affect such performances. There are also constraints on the torque converter such as axial length due to the powertrain packaging of front-wheel-drive vehicles and lock-up clutch. To improve the torque converter's characteristics, it is necessary to understand in detail the fluid flow phenomena that occurs in a torque converter. However, it is very difficult to visualize the flow in a torque converter because closed-loop flow occurs in a very complicated geometric space where three coupled blade cascades rotate at different speeds (Fig. 1).

Some research on measuring flow in a torque converter have recently been documented:

1. Flow measurements of a stator using a hot-film anemometer with an X-type probe [1,2]
2. Flow measurements of a stator [3], pump [4] and turbine [5] of a torque converter constructed from Plexiglas using a laser doppler velocimeter
3. Exit flow measurements of a pump, a turbine, and a stator using 5-hole pitot tubes [6,7,8]
4. Measurements of pressure distribution on stator blades [9,10]
5. Flow visualizations of a pump, a turbine, and a stator using the oil-film method [11]
6. Flow visualizations of two-dimensional blade models of a stator [10].

There is no known published case of applying a flow visualization technique to a real torque converter because of its complex construction. We therefore proposed a new experimental apparatus to visualize and measure the flow in a stator of a torque converter[12]. With this visualized torque converter, the flow in a stator has been measured and analyzed using the tuft method, oil-film method and laser doppler velocimeter.

In this study, we further propose a new visualization technique applying the laser sheet lighting method to the visualized torque converter. By this method, the distribution of separation region along the stator blade span can be observed. Furthermore, using a laser doppler velocimeter, velocity vectors and turbulence intensities can be measured in three stator blades of different thicknesses with the same camber line. Finally, the effect of the thickness of the stator blades on the flow patterns can be measured and analyzed concerning the torque capacity coefficient.

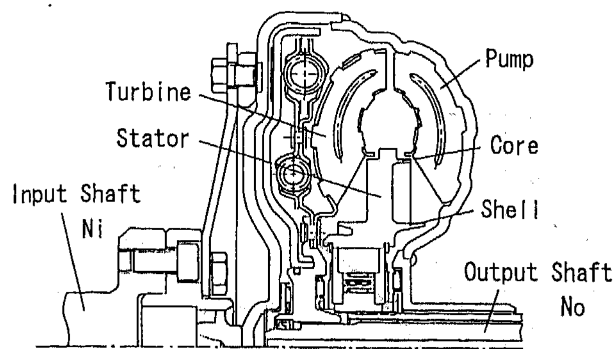


Fig. 1 Torque Converter

VISUALIZED TORQUE CONVERTER FOR OBSERVATION AND MEASUREMENT OF A STATOR FLOW

The construction of an apparatus for visualization of a stator is shown in Fig. 2. A conventional torque converter (shown in Fig. 1) has an input shaft (pump shaft) and an output shaft (turbine shaft) on each side. As shown in Fig. 2, a one-sided coaxial shaft consisting of an input shaft and an output shaft provides an open space inside the stator shaft for measurement. Through an acrylic window on the stator shaft and a mirror in the open space, the stator blades can be directly observed from inside of the stator shaft. With this construction, the whole image of the flow in the stator blade passage along the blade span can be observed. The laser sheet lighting method was applied to locate an observation section across the span. The laser sheet lighting method into the blade passage is shown in Fig. 3. A section of the metallic blade was replaced with an acrylic section with a mirror inside inclined to 45 degrees to the blade span. A laser light sheet was aimed at the mirror inside the stator shaft, and was reflected into the acrylic blade, and then reflected again off the mirror inside the blade. The laser light sheet then entered into the blade passage perpendicular to the span. The location of the reflecting mirror inside the blade determined the observation position of the spanwise direction. The span was divided into quarters and a reflecting mirror is placed at each of the three inside-points. The three blades, each with one mirror inside, can be removed and replaced as an attachment. Small bubbles were used as tracer particles for flow visualization, and were stirred with each rotating element.

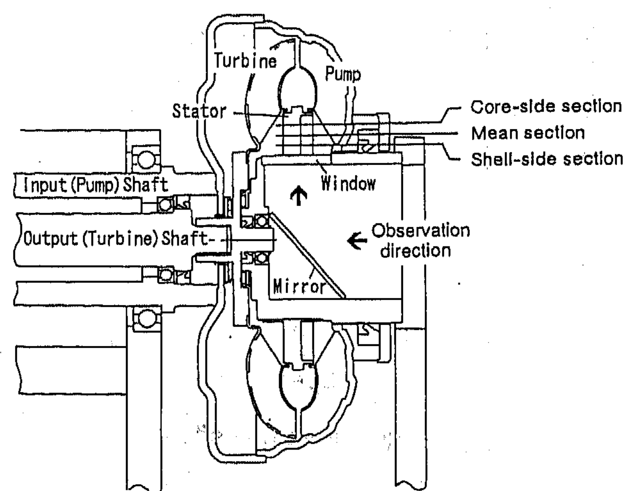


Fig. 2 Visualized torque converter for measurement of stator flow

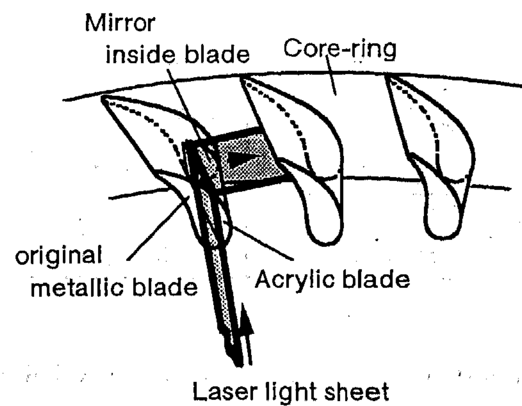


Fig. 3 Laser sheet lighting method

The experimental speed ratio (e=No/Ni; where Ni and No are the input shaft and output shaft speeds, respectively) is limited within the torque conversion range. A motor drives the input and output shafts connected to a continuous variable transmission (Fig. 4). Transparent, low viscous spindle oil was used in place of the red copper colored, high viscous torque converter oil. Experimental conditions are presented in Table 1.

When the scale of the torque converter and the order of the meridional velocity under the operating conditions are considered, the Reynolds number, Re($=uc/\nu$), is about 2×10^4, of which the reference length c is 0.02m, the reference velocity u is 2m/s and the kinematic viscosity ν is about 2.5×10^{-6}m²/s.

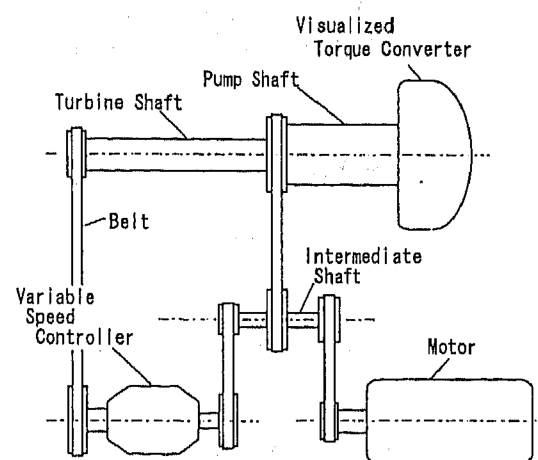


Fig. 4 Experimental apparatus

Table 1 Experimental conditions

Speed ratio	0, 0.2, 0.4, 0.6, 0.8
Pump speed	350 rpm
Stator speed	0 rpm
Oil temperature	30 degrees

TEST STATOR BLADES OF VARIOUS THICKNESS

The test stator blade shape is shown in Fig. 5. The blades had the same camber lines but three different thicknesses. The thin blade was 0.635 times as thick as the thick blade, and the plate blade was 2mm thick. Performance curves for the three torque converters showing efficiencies, torque ratios (output torque/input torque) and torque capacity coefficients (input torque/input speed^2) versus speed ratio are shown in Fig. 6. The torque converter pump, turbine and stator consisted of 31, 29 and 20 blades respectively. The inlet and outlet angles of the stator blade at the mean section were 0° and -49.5° respectively. The torque ratio decreases slightly as the blade thickness decreases. On the other hand, the torque capacity coefficient has characteristic curves according to the blade thickness. Thus we have found that the blade thickness of the stator affects the torque capacity coefficient.

FLOW VISUALIZATION IN THE STATOR BLADE USING LASER SHEET LIGHTING METHOD

Figure 7 shows flow observation photographs using the laser sheet lighting method. This figure shows visualization photographs at the shell-side, mean and core-side sections respectively at a speed ratio of e=0. Observation regions are within the limits of the leading edge area of the blade's suction surface at the location of the reflecting mirror. The separation regions at each section of the span are clearly observed as bright areas.

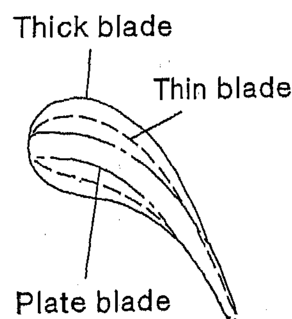
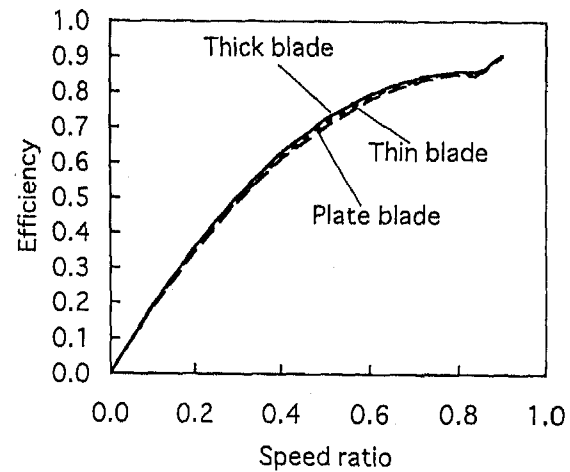
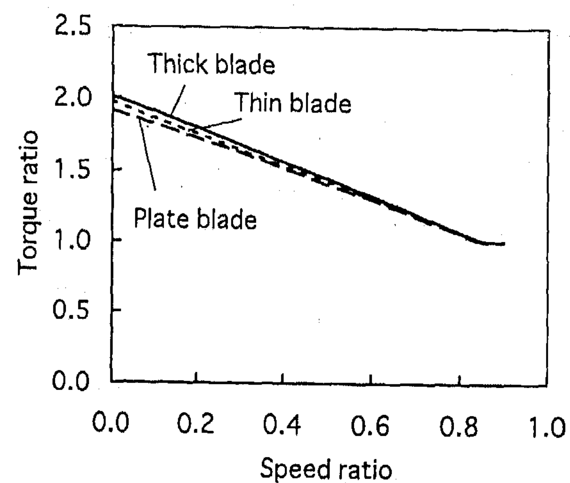


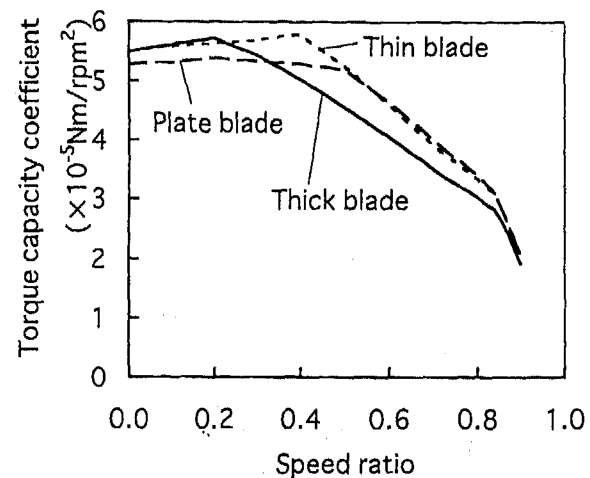
Fig. 5 Test stator blades



(a) Efficiency



(b) Torque ratio



(c) Torque capacity coefficient

Fig. 6 Performance characteristics of the experimental torque converter

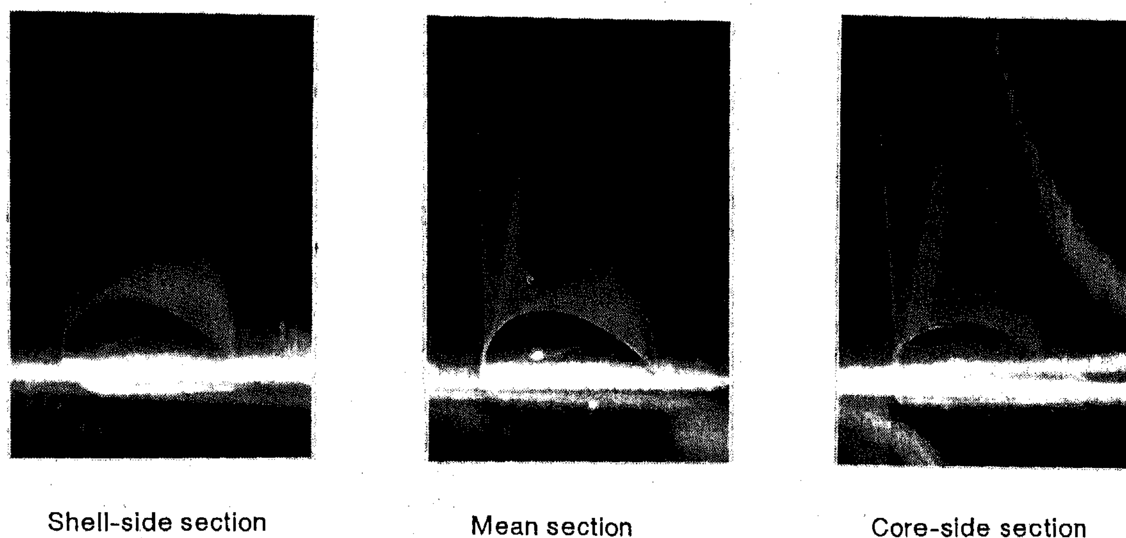


Fig. 7 Photographs using laser sheet lighting method, Thick blade , e=0.2

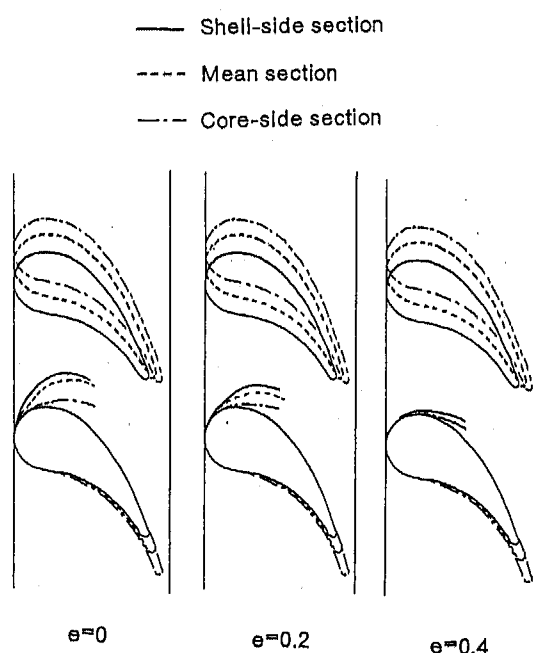


Fig. 8 Spanwise distribution of separation region

Figure 8 shows the separation regions at the shell-side, mean and core-side sections at e=0, 0.2 and 0.4. At e=0 and 0.2, separation regions at the shell-side and mean sections are larger than that of the core-side section. In addition, the area of separation regions decreases as the speed ratio increases. In other words, a large separation region is distributed uniformly from the shell-side to the mean section, as opposed to the small separation region of the core-side section at low speed ratios. The separation region area decreases as the speed ratio increases, and has uniform distribution along the span.

FLOW MEASUREMENT IN THE THREE STATOR BLADES OF DIFFERENT THICKNESS USING LASER DOPPLER VELOCIMETER

MEASUREMENT PROCEDURE - In this section, we quantify the flow in the three kinds of stators using a laser doppler velocimeter (LDV). The LDV system was a one-directional frequency shifted backscatter dual beam LDV (TSI) with a 15mW He-Ne laser. The system had a probe volume of 0.1mm in diameter and 1.4mm in length. The laser and optics mounted on a three-dimensional traversing table were operated by remote control. Flow velocity measurement sections were selected at the shell-side, mean and core-side sections across the span (shown in Fig. 2). Flow vectors were obtained by measuring the velocities of the flow in each direction crossing at right angles at the same points. A cylindrical acrylic window on the stator shaft refracts laser beams which enter toward the blade passages. The refractive angle and the crossing point of the two laser beams change according to the distance between the axis and the measuring point. The angle and crossing position of the intersecting two beams were therefore modified according to this distance. The signal processor was a digital burst correlator. A personal computer calculated the time-averaged velocities and turbulence intensity distributions from a data group of 1024 sampling points. To obtain high quality LDV signals, appropriate particle seeding in the flow must be present. In the present work, small bubbles were used into the flow and were stirred with each rotating element.

DISTRIBUTION OF VELOCITY VECTOR AND TURBULENCE INTENSITY - Figure 9 shows the time-averaged velocity vectors $\overline{U}$ and turbulence intensities σ ($=\sqrt{(\sigma_x^2+\sigma_y^2)/2}$), distribution at e=0.2 measured by LDV, where σ_x and σ_y are the root mean square of velocity fluctuations at right angles. The reason why some data for the core-side measurement are not shown is because of the removal of raw data. Larger refraction angles at the cylindrical acrylic window results in large distances between two intersection points consisting of two laser beams at right angles.

At e=0.2, flow patterns at the shell-side section are similar to those at the mean section as opposed to those at the core-side section for all three blades. In other words, at the shell-side and the mean sections, turbulence intensity becomes large in the region after the leading edge of the suction surface. This region corresponds to the separation region of the thick blade observed by the laser sheet lighting method (shown in Fig. 7). Thus the separation regions can be estimated from the scale of the turbulence intensity. The separation region at the shell-side section is almost the same as that at the mean section for thick and thin blades. But the separation region at the shell-side section of the plate blade is larger than that of the other sections. On the other hand, no separation regions are observed at the core-side section. We concluded that the separation significantly increases the turbulence intensity in the blade passage and diffusion of turbulence, due to the fact that the separation increases the turbulence intensity more at the outlet than at the inlet. The separation on the suction surface affects the outlet flow angle greatly. There are two types of effects. First, in the case of the shell-side section of the thick blade, continuous separation from the leading edge to the trailing edge increases the outlet flow angle compared to the outlet blade angle (considering the sign, positive or negative). Second, in the case of the shell-side section of the plate blade, re-attachment of separation decreases the outlet flow angle compared to the outlet blade angle. From these results, we understood that blade thickness changes the separation and thus affects the flow at the blade and outlet passage. In particular, the flow velocity and the angle at the inlet and outlet passages are important as it pertains to angular momentum change.

FLOW VELOCITY AND ANGLE DISTRIBUTION AT INLET PASSAGE - Figure 10 shows the circumferential (blade pitch direction) distribution of the meridional velocity and the flow angle at the front line on the mean section of the inlet passage. Meridional velocity distributions versus speed ratio change are basically the same for each blade. In other words, the flow is biased to the suction side at e=0 to 0.6, and becomes uniform with the increase of speed ratio, and is slightly biased to the pressure side at e=0.8. At the same speed ratio, however, the incline of the change in meridional velocity decreases in the order of thick, thin and plate blade, as their curves become as their curves flatten. This is because the shapes of the leading edge of the three stator blades affects the inlet flow patterns. The flow angle shows a decrease at e=0 to 0.8. The flow angle curves for the thin blade are similar to those for the plate blade. In the thick blade, there are a few changes versus speed ratio between the central part to the suction side. The influence of the blade thickness at the inlet passage is small as mentioned above.

FLOW VELOCITY AND ANGLE DISTRIBUTION AT OUTLET PASSAGE - Figure 11 shows the circumferential (blade pitch direction) distribution of the meridional velocity and flow angle at an end line on the mean section of the outlet passage. The thick blade has a non-uniform meridional velocity distribution at low to middle speed ratios, and the distribution becomes more uniform in the order of thin, and plate blade. Thick, thin and plate blades have the maximum meridional velocity at e=0.2, 0.4 and 0, respectively. The blade thickness greatly affects the outlet flow angle distribution. In the case of the thick blade, the flow angle changes slightly for each speed ratio, but shows a large increase at e=0.2 and 0.4 at the central part of the outlet passage. We concluded that the increase is caused by the wake at the trailing edge of the blade. The flow angle distributions for the thin and plate blades are more uniform than that for the thick blade. But the thin and plate blades have large flow angle changes compared to the thick blade. Thin and plate blades have the maximum flow angle at e=0.4 and 0.6, respectively. As a result, blade thickness affects the outlet flow. However it is difficult to distinguish the characteristics of blade thickness influence on the outlet flow angle clearly.

SPANWISE DISTRIBUTIONS OF OUTLET FLOW RATE AND ANGLE - Flow rate and flow angle at the outlet end line of the three sections across the span are calculated from the data of a group as shown in Fig. 11. We describe the flow rate at the outlet end line Q'(m^2/s) of three sections by the following expression.

Q'=(average meridional velocity) $\times$ (blade pitch length)

We also define the average outlet flow angle as the angle of a triangle consisting of the average meridional flow velocity component and the average circumferential flow velocity component.

Figure 12 shows the distribution of the flow rate and the average flow angle at the outlet end line for three blades. The flow rate and the average flow angle at the outlet end line for each section show the same tendency versus speed ratio, regardless of the blade shape. In other words, quantity relation of the flow rate is similar to the average flow angle for each blade at a certain speed ratio. The flow rate at each section becomes equal as the speed ratio increases. At the core-side section, flow rate decreases linearly for the speed ratio except at e=0 of the plate blade. In the case of the thick blade, the flow rates at shell-side and mean section are smaller than that at the

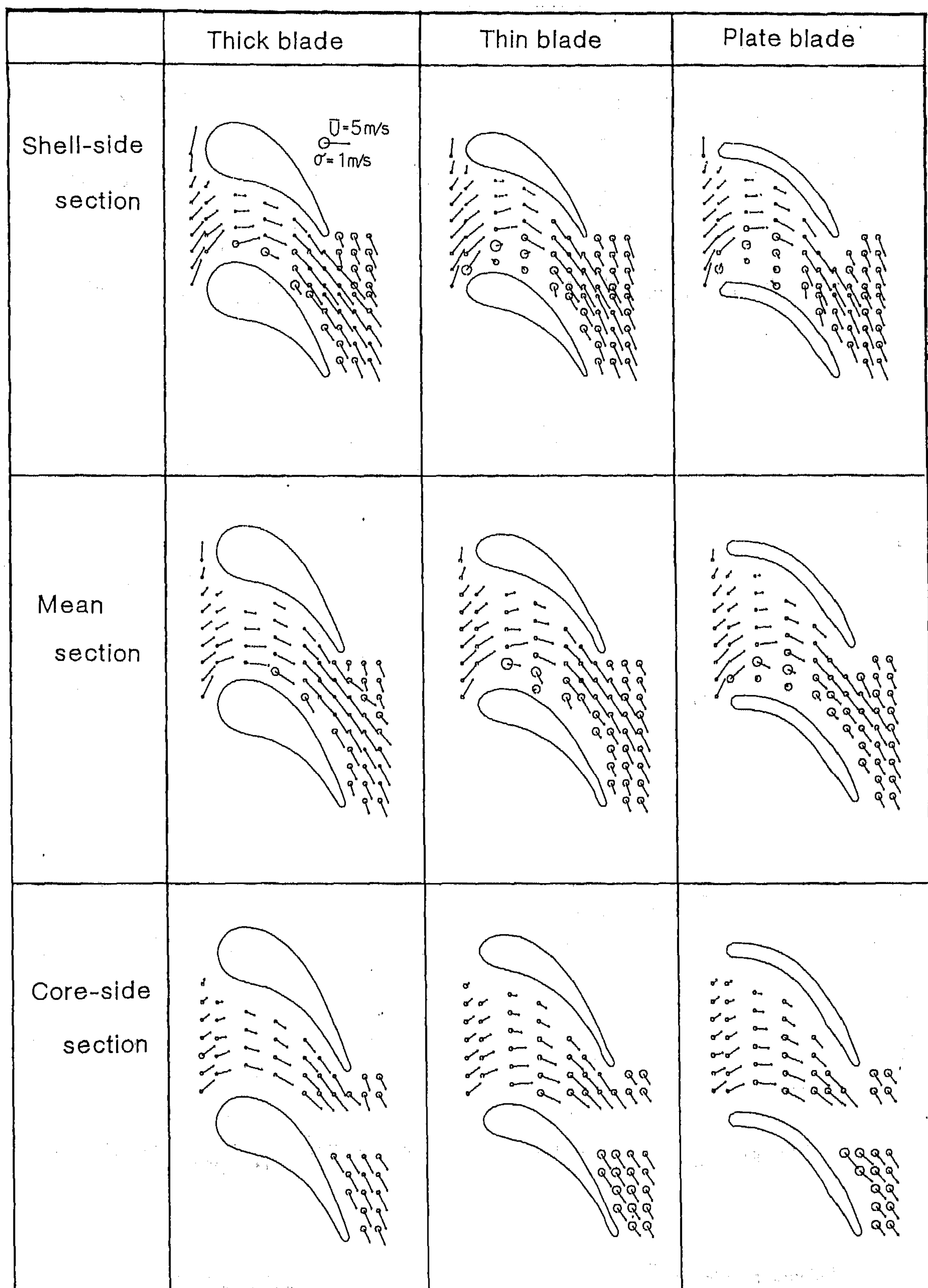
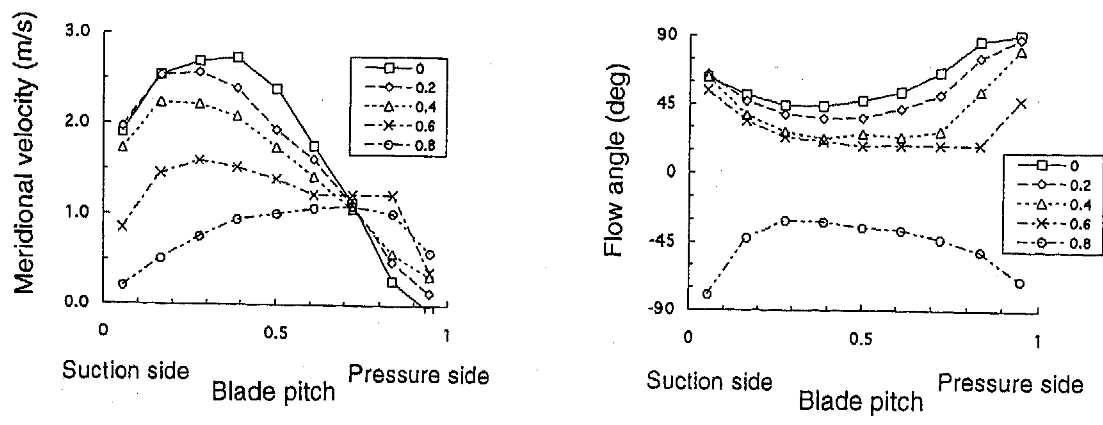
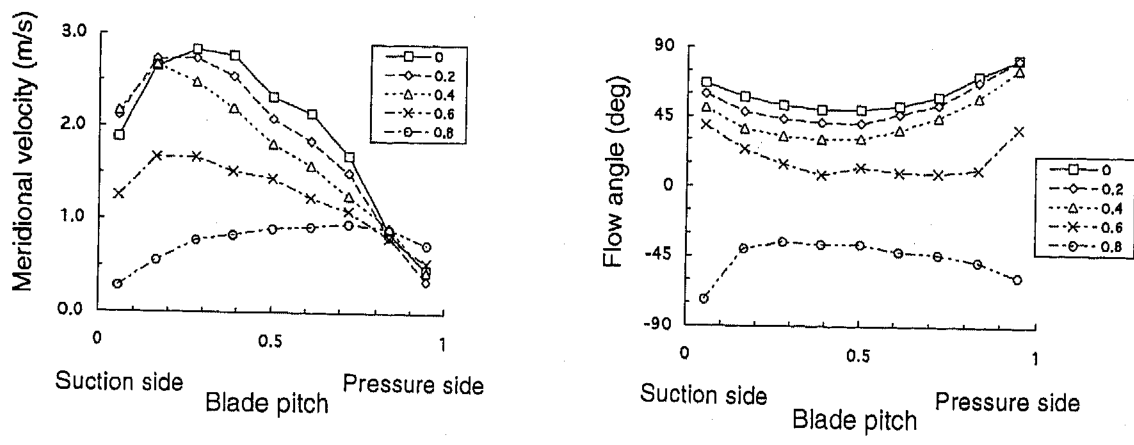


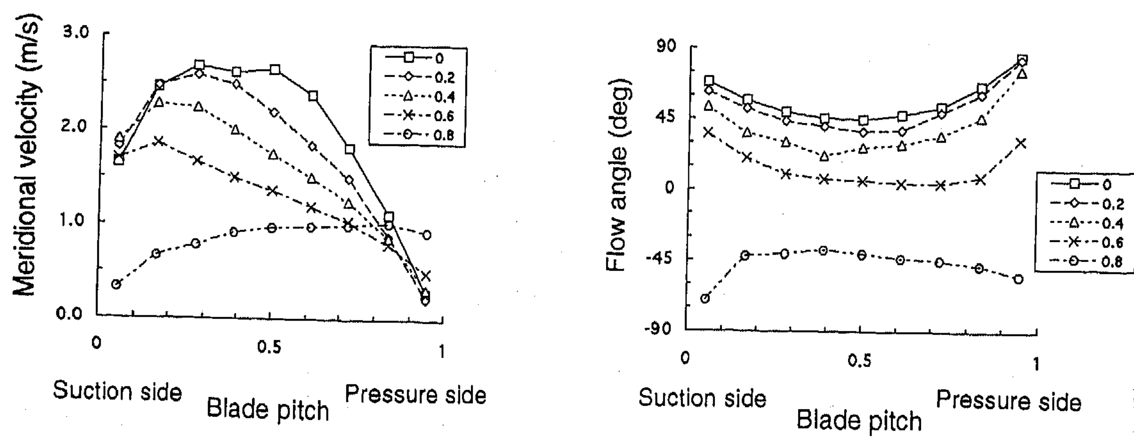
Fig. 9 Distribution of velocity vector and turbulence intensity, e=0.2



(a) Thick blade



(b) Thin blade

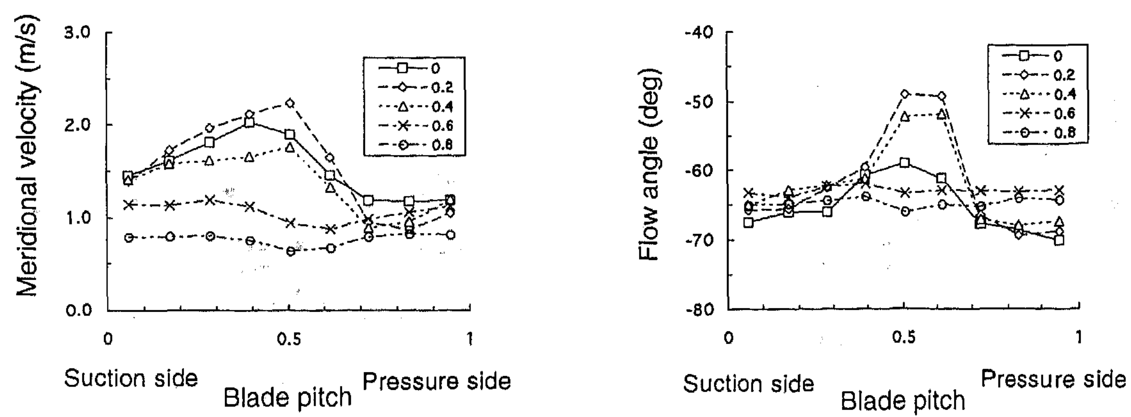


(c) Plate blade

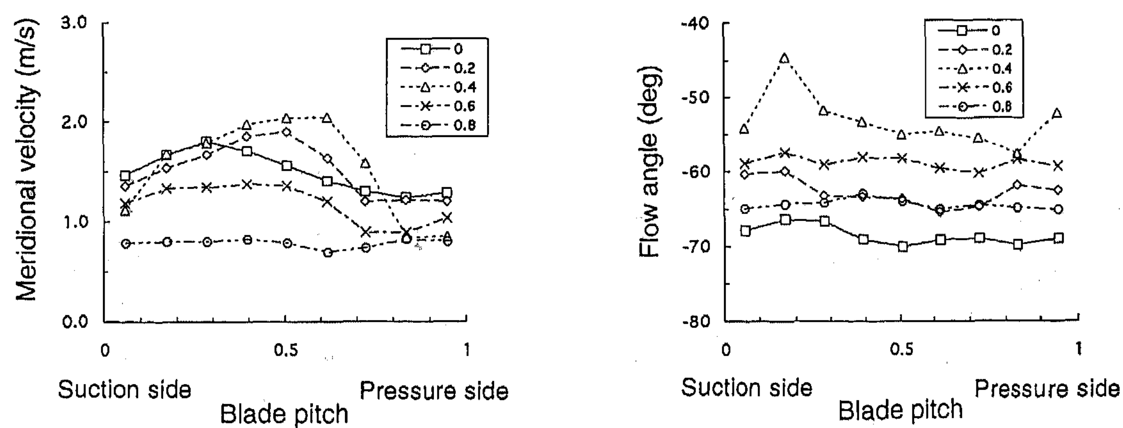
Meridional velocity

Flow angle

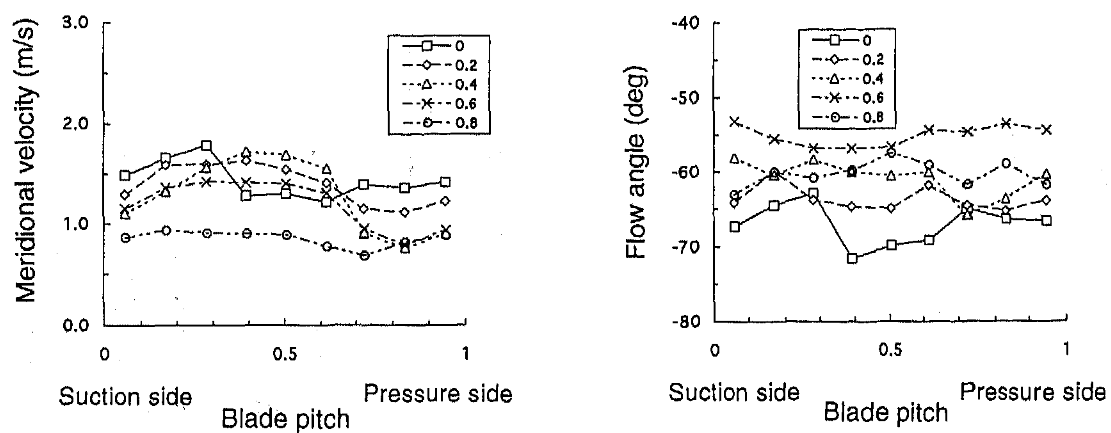
Fig. 10 Flow velocity and angle distributions at inlet passage, Mean section



(a) Thick blade



(b) Thin blade

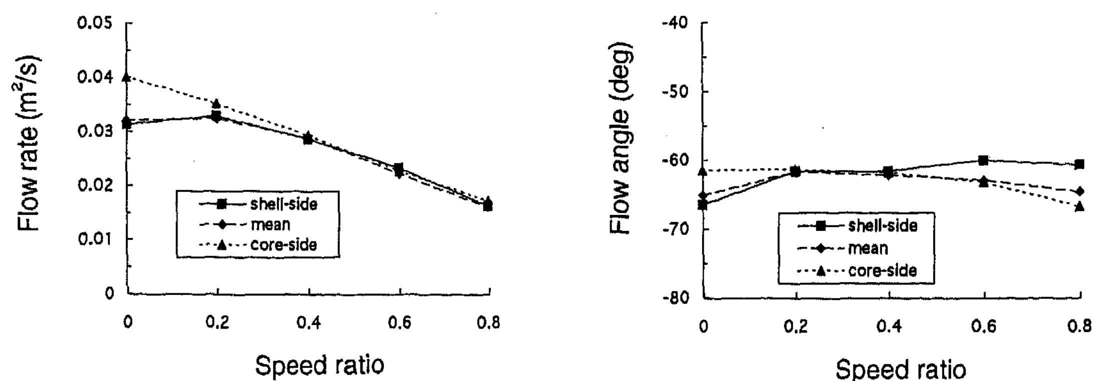


(c) Plate blade

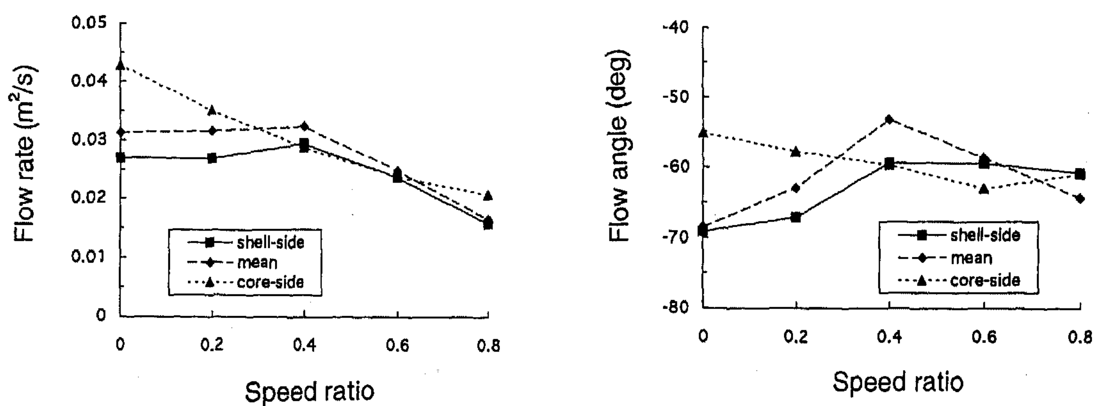
Meridional velocity

Flow angle

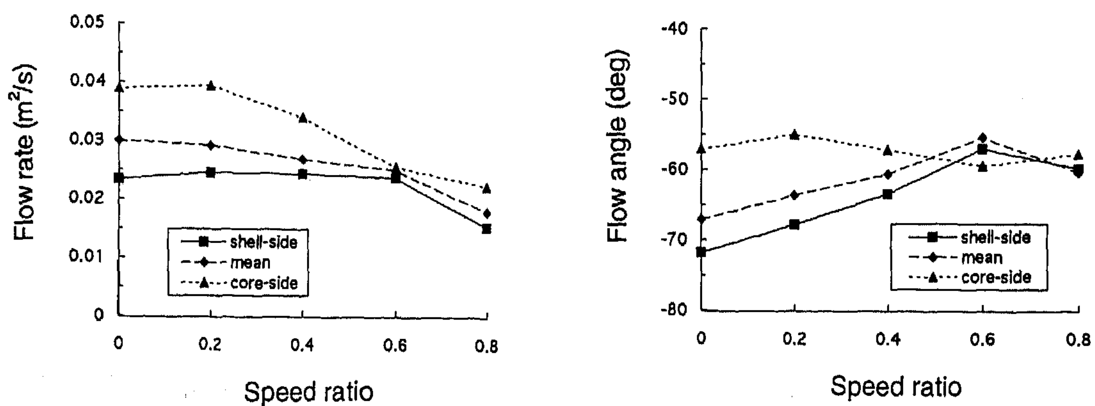
Fig. 11 Flow velocity and angle distributions at outlet passage, Mean section



(a) Thick blade



(b) Thin blade



(c) Plate blade

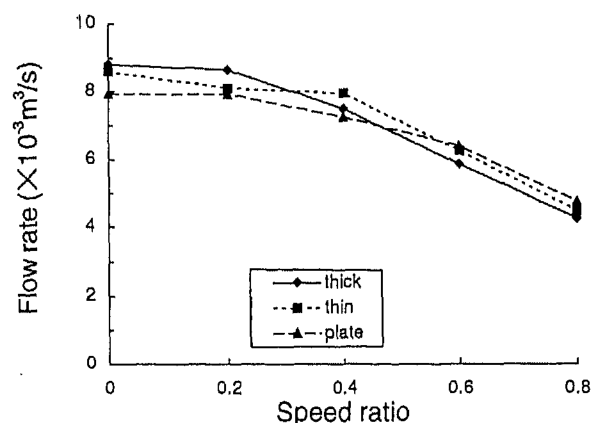
Flow rate

Flow angle

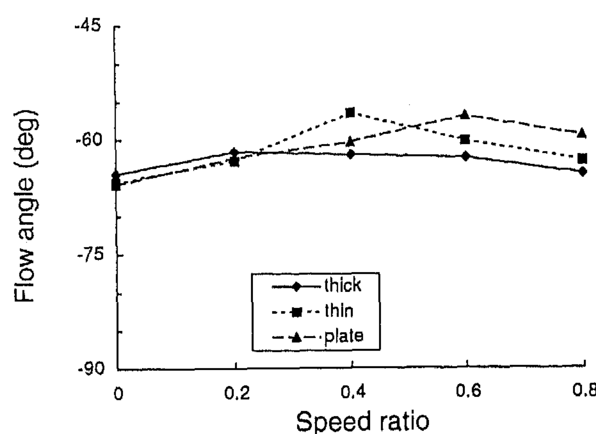
Fig. 12 Spanwise distributions of outlet flow rate and flow angle

core-side section at e=0 and show the maximum at e=0.2 and decreases linearly as the speed ratio increases. In the case of the thin blade, the flow velocity has a non-uniform distribution at the shell-side and mean sections at e=0 and 0.2 and becomes equal for the above e=0.4 at each section. In the case of the plate blade, the flow velocity has a large non-uniform distribution at e=0, 0.2 and 0.4 but becomes equal at e=0.6.

On the other hand, the average flow angle at the core-side section shows the same value for any speed ratio for each blade. The average flow angles at the shell-side and mean sections increase as the speed ratio increases and becomes equal to that of the core-side section. The speed ratio at which the average flow angle becomes the same value are 0.2, 0.4 and 0.6 at thick, thin and plate blades respectively. This characteristic is similar to that of the flow rate. In other words, the characteristics of the flow rate and the average flow angle at each section reveal that, for smaller blade thicknesses, sectional distributions along the blade span at the same speed ratio become more non-uniform and the non-uniformity lasts until a high speed ratio is attained..



(a) Flow rate



(b) Flow angle

Fig. 13 Distributions of outlet flow rate and flow angle in various thickness blades

DISTRIBUTION OF OUTLET FLOW RATE AND ANGLE IN THE VARIOUS THICKNESS BLADES - Meridional flow rate between the core-side section and the shell-side section along the blade span is integrated from the data of Fig. 12 with a trapezoidal rule. Also, from the data of the circumferential flow rate at each section along the span, the circumferential flow rate between the core-side section and the shell-side section along the span is integrated using the same trapezoidal rule. Finally, we define the average outlet flow angle between the core-side section and the shell-side section as the angle of a triangle consisting of meridional and circumferential flow rates.

Figure 13 shows the outlet meridional flow rate and average outlet flow angle between the shell-side and core-side sections for the three kinds of blades. Generally, meridional flow rates decrease as the speed ratio increases for each blade. It was observed that the meridional flow rate characteristics are very similar to those of the torque capacity coefficient as shown in Fig. 6. The condition of the general curves and value size relation of each blade is similar except at e=0.4. On the other hand, the average flow angle has the maximum at middle speed ratios. Also, the average flow angle characteristics are very similar to those in Fig. 6 with regard to the value relation except for the thin blade's angle at low speed ratios. We ascertained experimentally that the outlet flow rate and the average flow angle greatly influence the torque capacity coefficient. In other words, the general characteristics are decided by the outlet flow rate, and then the characteristics are emphasized by the average outlet flow angle.

CONCLUSIONS

This paper describes the blade thickness effects to the flow in the stator by flow visualization and measurement with a new visualized torque converter.

1. The thickness of the stator blade affects the torque capacity coefficient characteristics more than the efficiency and torque ratio characteristics.

2. At low speed ratios, the flow around the leading edge has different separation regions along the blade span.

3. The thickness of the stator blades affects the flow patterns.

4. Both the average meridional flow velocity and average flow angle at the outlet line have non-uniform distributions along the blade span.

5. The thickness of the stator blade affects the outlet flow rate and the average outlet flow angle significantly.

6. The outlet flow rate and the average outlet flow angle show similar characteristics to the torque capacity coefficient.

7. The general characteristics of the torque capacity coefficient are decided by the outlet flow rate, and the characteristics are emphasized by the average outlet flow angle.

REFERENCES

1. Yamada, M., Imai, K., Yasuda, E., Yamaguchi, H., An Analysis for Internal Flow in a Torque Converter, 1st report, A Measurement of Inlet Flow of a Stator by Hot-film Anemometer(in Japanese), Proceedings of Spring Convention of the Japan Hydraulics & Pneumatics Society, 73,1984.

2. Volker Browarzik, Experimental Investigation of rotor/rotor interaction in a Hydrodynamic Torque Converter Using Hot-Film Anemometry, ASME Paper No.94-GT-246, 1994.

3. Bahr, H. M., Flack, R. D., By, R. R., Zhanb, J. J., Laser Velocimeter Measurements in the Stator of a Torque Converter, SAE Paper No.901769,1990.

4. Gruver, J. K., Flack, R. D., Brun, K., Laser Velocimeter Measurements in the Pump of a Torque Converter Part I -Average Measurements, ASME Paper No.94-GT-47, 1994.

5. Brun, K., Flack, R. D., Laser Doppler Measurements in the Turbine of an Automotive Torque Converter Part I -Average Measurements, ASME Paper No.95-GT-292, 1995.

6. Kubo, M., Ejiri, E., Okada, K., Ishii, Y., Improvement of Accuracy in Torque Converter Performance Prediction by Internal Flow Measurement(in Japanese), Proceedings of Autumn Convention of JSAE, 912, 1991, 3.169.

7. Marathe, B. V., Lakshminarayana, B., Dong, Y., Experimental and Numerical Investigation of Stator Exit Flow Field of an Automotive Torque Converter, ASME Paper No.94-GT-32, 1994.

8. Marathe, B. V., Lakshminarayana, B., Maddock, D. G., Experimental Investigation of Steady and Unsteady Flow Field Downstream of an Automotive Torque Converter Turbine and Inside the Stator Part I :Flow at the Exit of Turbine, ASME Paper No.95-GT-231, 1995.

9. By, R. R., Lakshminarayana, B., Static Pressure Measurement in a Torque Converter Stator, SAE Paper No.911934, 1991.

10. Watanabe, H., Yamada, M., Kimura, M., Fukumura, K., Kojima, M., An Analysis for Internal Flow in a Torque Converter, 3rd report, Flow Measurements around Stator Blades(in Japanese), Proceedings of Autumn Convention of the Japan Hydraulics & Pneumatics Society, 145, 1992.

11. Numazawa, A., Ushijima, F., Fukumura, K., An Experimental Analysis of Fluid Flow in a Torque Converter, SAE Paper No.830571, 1983.

12. Watanabe, H., Yoshida, K., Yamada, M., Kojima, M., Flow Visualization and Measurement in the Stator of a Torque Converter, JSAE Review 17, 1996, 25.

970682

Generalized Equations for Roller One-Way Clutch Analysis and Design

David R. Chesney
General Motors Corp.

John M. Kremer
Borg-Warner Automotive

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

In 1973, Sauzedde [1] wrote a paper that is commonly considered the benchmark document on roller One-Way Clutch (OWC) analysis and design. Among other topics, Sauzedde led the reader through analyses on contact stress, outer ring hoop stress, and roller centrifugal force computations. Following this seminal paper, several enhancements were presented, further refining the original equations [2] and better describing roller mechanics [3]. At the time these papers were written, the automotive industry in the U.S. predominantly used the English system of measure. The use of English units often obscured the original derivations of constants used in the equations. The purpose of this paper is to present the equations used in roller OWC analysis and design in terms of dimensionless units.

INTRODUCTION

There are two types of the roller One-Way Clutch (OWC) commonly used in the automotive industry. The first has the roller cams located in the outer race, and is referred to as an outer cam design. The second has the roller cams located in the inner race, and is referred to as an inner cam design. Each design has its own advantages and disadvantages, the discussion of which is outside the scope of this paper. However, the analytical techniques developed for one design can easily be applied to the other.

The OWC has two modes of operation: 1. The freewheel mode, and 2. The engaged mode. Each mode has its unique requirements and related analytical techniques. In this paper we will present various techniques commonly used in the design and analysis of the roller OWC.

While in the engaged mode of operation, the OWC transmits a torque from the inner race to the outer race (or *vice-versa*) through the rollers. Significant contact forces are generated at the contact points between the roller and the races. These loads are known as the *strut forces*. Contact stresses and hoop stresses are the primary concerns in the engaged mode of operation. The standard analytical techniques for the computation of these stresses are presented in the CONTACT STRESS and HOOP STRESS Sections.

During the freewheel mode, there is relative motion between the inner and outer races. If the cammed race is spinning, the rollers revolve with it and, with the aid of an energizing spring, stay in contact with the *rollerway*[1] of the other race. The main concerns in this mode of operation are:

1. Wear of the rollers and races,
2. Proper positioning of the roller, and
3. Fatigue of the energizing spring.

Because of the multitude of parameters that influence it, wear is more of an empirical problem. On the other hand, the positioning of the roller and spring fatigue are primarily a function of the roller inertia forces and the spring force, and can be approached analytically. The standard analytical techniques for the prediction of these forces are presented in the ROLLER DYNAMICS Section.

EQUILIBRIUM

The first step in understanding the stresses acting on a roller OWC is to understand the balance of forces and torques acting upon an engaged roller. Figure 1 and Figure 2 show the free body diagram of a roller in an outer cam and inner cam OWC, respectively. As shown, the geometry is slightly different because of the cam location, however, free body analysis is fundamentally the same. The contact force between a roller and the races is on the order of several thousand Newtons. The spring force acting on the roller is generally less than 10 N. Therefore, without loss of accuracy, we can

[1]The rollerway is the smooth cylindrical surface of the non-cammed race that contacts the roller (see Figure 5).

neglect the spring load in our equilibrium analysis. This is the underlying assumption used in the following analysis.

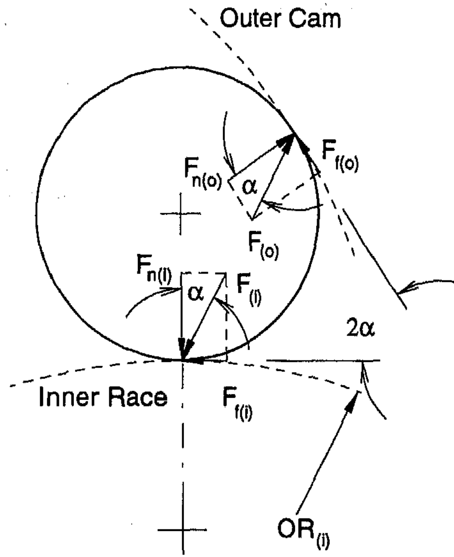


Figure 1: Free body diagram of the roller in an outer cam OWC.

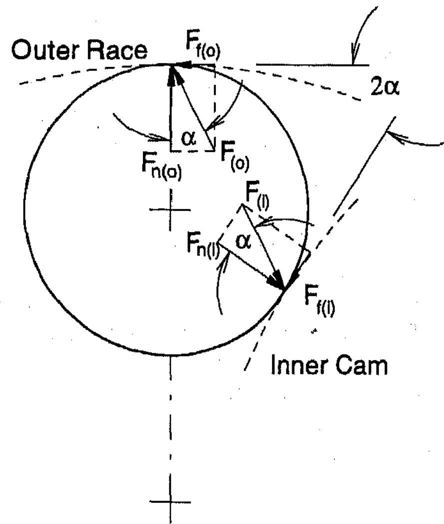


Figure 2: Free body diagram of the roller in an inner cam OWC.

The roller is in equilibrium, and thus the total force acting between the inner race and roller ($F_{(i)}$) is equivalent to the total force acting between the outer race and roller ($F_{(o)}$). That is,

$$F_{(i)} = F_{(o)} \tag{1}$$

The *tangential* force acting between the inner race and roller ($F_{f(i)}$) is a trigonometric function of the total force between the inner race and roller ($F_{(i)}$) and the strut angle (α)

$$F_{f(i)} = F_{(i)} \cdot \sin\alpha \tag{2}$$

The final step in balancing the forces on a roller is to derive the *normal* force ($F_{n(i)}$) acting on the roller.

$$F_{n(i)} = F_{(i)} \cdot \cos\alpha \tag{3}$$

The *total* torque between the inner race and rollers ($T_{(i)}$) for an outer cam roller clutch is the product of the tangential force between the inner race and a roller ($F_{f(i)}$), the moment arm ($OR_{(i)}$), and the number of rollers ($N_{(r)}$).

$$T_{(i)} = F_{f(i)} \cdot OR_{(i)} \cdot N_{(r)} \tag{4}$$

Equation (2) showed that the tangential force between the inner race and a roller ($F_{f(i)}$) is a trigonometric function of the total force between the inner race and a roller ($F_{(i)}$). That is, $F_{f(i)} = F_{(i)} \sin\alpha$, and EQ (4) becomes

$$T_{(i)} = F_{(i)} \cdot \sin\alpha \cdot OR_{(i)} \cdot N_{(r)} \tag{5}$$

Solving EQ (5) for total force per roller ($F_{(i)}$), rather than torque, results in

$$F_{(i)} = \frac{T_{(i)}}{N_{(r)} \cdot OR_{(i)} \cdot \sin\alpha} \tag{6}$$

STRUT ANGLE LIMITS

In the previous section it was demonstrated that the strut forces $F_{(i)}$ and $F_{(o)}$ can be decomposed into their tangential and normal components, as is shown in EQ (2) and EQ (3), respectively. The basic relationship between the tangential and normal forces and the strut angle is obtained by dividing EQ (2) by EQ (3)

$$\frac{F_{f(i)}}{F_{n(i)}} = \frac{F_{(i)} \cdot \sin\alpha}{F_{(i)} \cdot \cos\alpha} = \tan\alpha \tag{7}$$

The tangential force $F_{f(i)}$ is also the frictional force of the system. Its magnitude is dependent on the normal force $F_{n(i)}$ and is limited by the static coefficient of friction μ. That is,

$$F_{f(i)} \le \mu F_{n(i)} \tag{8}$$

Equations (7) and (8) define the same system and, therefore, are interdependent. Substituting EQ (7) into EQ (8), and simplifying, we get

$$\tan\alpha \le \mu \qquad (9)$$

This expression sets the upper limit for the value of the strut angle α.

The coefficient of static friction μ of the clutch environment is not a single-valued variable, but has a range of values. It depends upon factors which can change during the life of the clutch. These factors include:

1. The race and roller surface finish
2. The lubrication fluid composition
3. The lubrication fluid temperature

The value of μ for hardened steel on hardened steel bathed in automatic transmission fluid is generally between 0.080 and 0.120. Using the lower limit of 0.080 as our value for μ, EQ (9) yields a maximum strut angle value of

$$\alpha \le 4.574^{\circ} \qquad (10)$$

During operation of the clutch, the strut angle must never exceed this limit or the rollers may slip when the clutch engages. Dimensional variations during component manufacture, or wear as a result of clutch operation, can cause the strut angle to deviate from the desired value. Therefore, it is prudent to specify a strut angle which is less than the maximum.

Equation (10) defines the upper limit of the strut angle. There are practical limits on the minimum strut angle, also. For example, a strut angle of 1° is feasible; however, the release of the elastic strain energy in the clutch system (when the transmitted torque drops to zero) tends to trap the roller and prevent a smooth disengagement of the clutch. In addition, a greater normal force is required to transmit a given operating torque as the strut angle decreases and this, in turn, increases the contact and hoop stresses generated in the races. Figure 3 exemplifies this relationship, wherein we have plotted the strut component force ratio versus the strut angle.

CONTACT STRESS

Spalling and brinelling are two possible failure modes of a roller OWC. Both are contact stress related. Computation of the contact stress and how it relates to OWC design is discussed in this section.

We assume that the roller OWC has multiple rollers, each carrying an equal portion of the load. Therefore, examining one roller is representative of all rollers within the clutch. The race/roller contact in a roller OWC is either concave/convex, as between the outer race and roller; or convex/convex, as between the inner race and roller. The former is referred to as *conformal* contact, whereas the latter is referred to as *non-conformal* contact. Non-conformal contact loads lead to higher contact stresses, and therefore, we are only concerned with inner race/roller contact stresses. The variables used in the following discussion are defined in Figure 4.

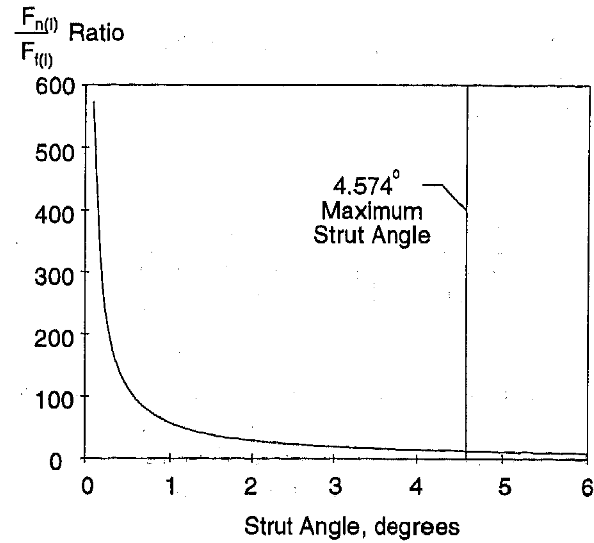


Figure 3: Ratio of the strut force components as a function of strut angle.

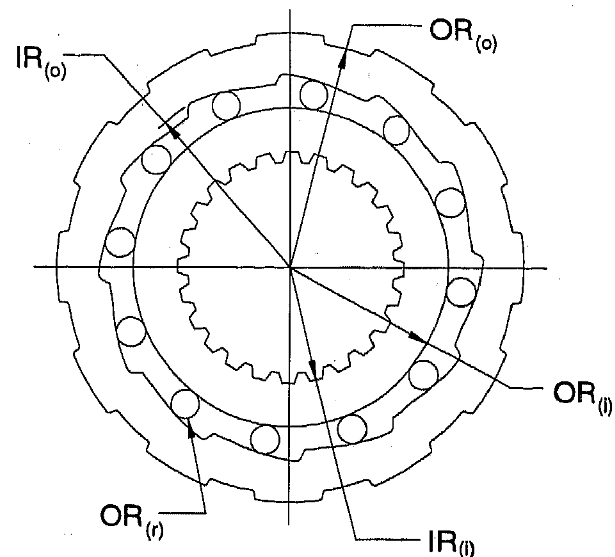


Figure 4: Geometry definition of outer cam roller clutch.

The equation to determine maximum hertz stress ($\sigma_{c(max)}$) between two convex surfaces is defined by Young [4] in terms of the resultant force between the inner race and roller ($F_{(l)}$), outer radius of the inner race ($OR_{(l)}$), roller radius ($OR_{(r)}$), length of the roller ($L_{(r)}$), and material properties of the inner race and roller ($v_{(l)}$, $E_{(l)}$, $v_{(r)}$, $E_{(r)}$):

$$\sigma_{c(max)} = \sqrt{\frac{2\cdot F_{(l)}\cdot\left(\frac{1}{OR_{(l)}}+\frac{1}{OR_{(r)}}\right)}{\pi\cdot L_{(r)}\cdot 4\cdot\left(\frac{(1-\upsilon_{(l)}^2)}{E_{(l)}}+\frac{(1-\upsilon_{(r)}^2)}{E_{(r)}}\right)}} \qquad (11)$$

A commonly used derivation of EQ (11) defines hertz stress in terms of its mean value ($\sigma_{c(mean)}$), rather than maximum value [5,6]:

$$\sigma_{c(mean)} = \frac{1}{2}\cdot\sqrt{\frac{F_{(i)}\cdot\pi\cdot\left(\frac{1}{OR_{(i)}}+\frac{1}{OR_{(r)}}\right)}{L_{(r)}\cdot 4\cdot\left(\frac{(1-\upsilon_{(i)}^2)}{E_{(i)}}+\frac{(1-\upsilon_{(r)}^2)}{E_{(r)}}\right)}} \qquad (12)$$

We use the equation for mean contact stress to derive an additional term for the total force per roller ($F_{(i)}$). The constants and material properties from EQ (12) (π, $v_{(i)}$, $v_{(r)}$, $E_{(i)}$, $E_{(r)}$) are brought outside the radical and named c_1:

$$c_1 = \frac{1}{2}\cdot\sqrt{\frac{\pi}{\left(\frac{4(1-\upsilon_{(i)}^2)}{E_{(i)}}+\frac{4(1-\upsilon_{(r)}^2)}{E_{(r)}}\right)}} \qquad (13)$$

Then, EQ (12) simplifies to:

$$\sigma_{c(mean)} = c_1\cdot\sqrt{\frac{F_{(i)}\cdot\left(\frac{1}{OR_{(i)}}+\frac{1}{OR_{(r)}}\right)}{L_{(r)}}} \qquad (14)$$

Solving EQ (14) in terms of total force per roller ($F_{(i)}$) results in the following equation:

$$F_{(i)} = \frac{L_{(r)}\cdot\left(\frac{\sigma_{c(mean)}}{c_1}\right)^2}{\frac{1}{OR_{(r)}}+\frac{1}{OR_{(i)}}} \qquad (15)$$

For an outer cam roller clutch, we now have two equations (EQ (6) and EQ (15)) that solve for total force per roller ($F_{(i)}$). We set EQ (6) equal to EQ (15) and obtain the following:

$$\frac{T_{(i)}}{N_{(r)}\cdot OR_{(i)}\cdot\sin\alpha} = \frac{L_{(r)}\cdot\left(\frac{\sigma_{c(mean)}}{c_1}\right)^2}{\frac{1}{OR_{(r)}}+\frac{1}{OR_{(i)}}} \qquad (16)$$

In a realistic design of an outer cam roller clutch, we want to determine the optimal inner race outer radius ($OR_{(i)}$) given material property and loading characteristics. That is, EQ (16) is solved in terms of the inner race outer radius ($OR_{(i)}$). Equation (16) does not have a simple solution for the inner race outer radius ($OR_{(i)}$), however, it is a quadratic equation in the following form:

$$0=\left(OR_{(r)}\cdot N_{(r)}\cdot\sin\alpha\cdot L_{(r)}\cdot\left(\frac{\sigma_{c(mean)}}{c_1}\right)^2\right)\cdot OR_{(i)}^2 + \left(-T_{(i)}\right)\cdot OR_{(i)} + \left(-T_{(i)}\cdot OR_{(r)}\right) \qquad (17)$$

and the quadratic equation that solves for the inner race outer radius ($OR_{(i)}$) is:

$$OR_{(i)} = \frac{T_{(i)} \pm \sqrt{T_{(i)}^2 - 4\cdot\left(OR_{(r)}\cdot N_{(r)}\cdot\sin\alpha\cdot L_{(r)}\cdot\left(\frac{\sigma_{c(mean)}}{c_1}\right)^2\right)\cdot\left(-T_{(i)}\cdot OR_{(r)}\right)}}{2\cdot\left(OR_{(r)}\cdot N_{(r)}\cdot\sin\alpha\cdot L_{(r)}\cdot\left(\frac{\sigma_{c(mean)}}{c_1}\right)^2\right)} \qquad (18)$$

That is, given load, material property, and roller parameters, EQ (18) calculates the outer radius of the inner race for an outer cam roller clutch.

To use EQ (18), we must select an operating contact stress at which the roller OWC will operate. Application history has shown that limiting the mean contact stress to 3.79 GPa (550,000 psi) is reasonable for roller OWC applications. Equation (18) yields two real roots, one negative and one positive. The negative root is meaningless when discussing magnitudes of length, and therefore only the positive root is considered.

HOOP STRESS

Contact forces between the rollers and races create a circumferential (*hoop*) stress in the outer and inner races. The hoop stress on the inner race is compressive, and steel does not typically fatigue due to compressive stresses. However, the outer race experiences tensile hoop stresses, which is a typical mode of fatigue failure for steel. Therefore, we consider hoop stress for the outer race only.

The discrete nature of the contact forces also creates bending stresses. If the number of rollers is very low, then the bending has a significant contribution to the circumferential stress. However, if the number of rollers is relatively large (>8), then the circumferential stresses in the outer race are predominantly hoop stresses. Therefore, in the discussion below, we assume the number of rollers is greater than eight and use the average hoop stress equation.

The governing equation for the mean hoop stress in the outer race ($\sigma_{h\,(mean)}$), as defined by thick cylinder analysis, is a function of the pressure on the interior surface of the outer race ($Q_{(o)}$), and the wall thickness of the outer race ($OR_{(o)}$ and $IR_{(o)}$) [7,8]:

$$\sigma_{h(mean)} = Q_{(o)}\cdot\left(\frac{OR_{(o)}^2+IR_{(o)}^2}{OR_{(o)}^2-IR_{(o)}^2}\right) \qquad (19)$$

The pressure on the inner surface of the outer race ($Q_{(o)}$) of an outer cam roller clutch is a function of the radial force per roller ($F_{(o)} \cdot \cos\alpha$), the number of rollers ($N_{(r)}$), and the total inner surface area of the outer race ($2 \cdot \pi \cdot r \cdot l = 2 \cdot \pi \cdot IR_{(o)} \cdot L_{(o)}$).

$$Q_{(o)} = F_{(o)} \cdot \left(\frac{N_{(r)} \cdot \cos\alpha}{2 \cdot \pi \cdot IR_{(o)} \cdot L_{(o)}} \right) \quad (20)$$

If we assume that the length of the roller is equal to the length of the outer race ($L_{(r)} = L_{(o)}$), and substitute EQs (1), (15), and (20) into EQ (19), then the following equation is derived for the mean hoop stress ($\sigma_{h(mean)}$) in the outer race:

$$\sigma_{h(mean)} = \left(\frac{\sigma_{c(mean)}}{c_1} \right)^2 \cdot \left(\frac{OR_{(r)} \cdot OR_{(i)} \cdot N_{(r)} \cdot \cos\alpha}{\left(OR_{(r)} + OR_{(i)}\right) \cdot 2 \cdot \pi \cdot IR_{(o)}} \right) \cdot \left(\frac{OR_{(o)}^2 + IR_{(o)}^2}{OR_{(o)}^2 - IR_{(o)}^2} \right) \quad (21)$$

However, we want to *assume* the average hoop stress ($\sigma_{h(mean)}$) that the outer race will encounter, and calculate the outer radius accordingly ($OR_{(o)}$). Note that the inner radius of the outer race ($IR_{(o)}$) is determined based upon the geometry of the inner race calculated earlier via EQ (18) and the geometry of the rollers.

Therefore, EQ (21) is put in terms of the outer radius of the outer race, where the pressure on the interior surface of the outer race ($Q_{(o)}$) was previously defined in EQ (20).

$$OR_{(o)} = \sqrt{\frac{Q_{(o)} \cdot IR_{(o)}^2 + \sigma_{h(mean)} \cdot IR_{(o)}^2}{\sigma_{h(mean)} - Q_{(o)}}} \quad (22)$$

ROLLER DYNAMICS

Roller float is defined as a condition wherein the roller loses contact with the rollerway. In certain applications (e.g. in the torque converter) this is a design objective of the OWC engineer. However, in most applications roller float is deleterious to the operation of the clutch, as will be explained later.

There are two conditions that may cause roller float. The first occurs in clutches of outer cam design and happens when the clutch and outer race are spinning. The centrifugal force acting on the roller as a result of this rotary motion "pulls" the roller away from the center of rotation and, hence, away from the rollerway. We refer to this condition as *centrifugal roller float*. In applications where the inner race is grounded, such as in the torque converter, this type of roller float is advantageous in that the roller/inner race contact is lost, thereby reducing wear and spin losses. If, however, the clutch must be "engagement ready" with the outer race spinning, this type of roller float is detrimental because the clutch cannot engage unless there is contact between the rollers and the inner race.

The second type of roller float can occur in clutches of either outer or inner cam design, and is called *eccentricity roller float*. This type of roller float is the result of a rollerway that has a finite eccentricity (or runout) with respect to the cammed race. As one race rotates with respect to the other, the roller will travel "up" and "down" the cam surface while tracking the rollerway. If the over-running speed is great enough and the eccentricity large enough, the race will "throw" the roller off of the rollerway.

During the operation of a roller OWC, either of these conditions can occur. In some cases, the inertia forces that lead to these float conditions can occur simultaneously. Should this happen, the effects are additive. Hence, each condition will be addressed separately in the following sections.

CENTRIFUGAL ROLLER FLOAT - In certain applications, the outer race of a roller OWC freewheels about its axis. If the clutch is an outer cam design, then the rollers of the clutch will revolve with the outer race. As a result of this motion, the roller must accelerate towards the center of rotation to maintain its position relative to the outer race. This is known as *centripetal acceleration* and, by Newton's second law, the force needed to create this acceleration is

$$P_{cen} = m_{(r)} a = m_{(r)} c\, \omega_{(o)}^2 \quad (23)$$

where P_{cen} is the roller centrifugal force, $m_{(r)}$ is the mass of the roller, a is the acceleration of the roller, c is the distance from the center of the clutch to the center of the roller, and $\omega_{(o)}$ is the angular velocity of the outer race. The resulting system of forces is shown in Figure 5.

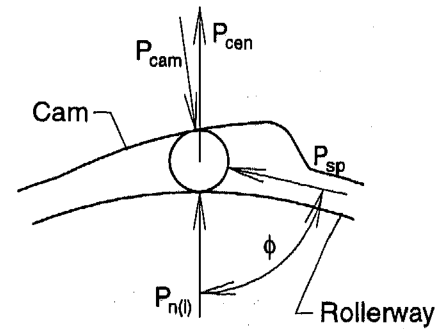


Figure 5: Forces acting on the roller for the centrifugal roller float analysis.

There exists an outer race speed above which the centrifugal roller force P_{cen} overcomes the spring force P_{sp} and the roller loses contact with the inner race. This condition is called *centrifugal roller float*. As $P_{n(I)}$ reaches zero, we

have a condition of equilibrium in which the roller is acted upon by three forces: P_{sp}, P_{cen}, and P_{cam}. Utilizing a force triangle for this condition (Figure 6), the spring force can be derived using the law of sine's

$$P_{sp} = P_{cen}\frac{\sin(2\alpha)}{\sin(\phi-2\alpha)} = m_{(r)} c\,\omega_{(o)}^2 \frac{\sin(2\alpha)}{\sin(\phi-2\alpha)} \qquad (24)$$

where ϕ is the angle between the line of action of the centrifugal force P_{cen} and the spring force P_{sp}. Equation (24) is used to compute the spring force required to allow centrifugal roller float for a given clutch and outer race speed. Alternatively, we may re-arrange EQ (24) to obtain the roller float speed for a given spring load (EQ (25)).

$$\omega_{(o)} = \sqrt{\frac{P_{sp}}{c\,m_{(r)}}\frac{\sin(\phi-2\alpha)}{\sin(2\alpha)}} \qquad (25)$$

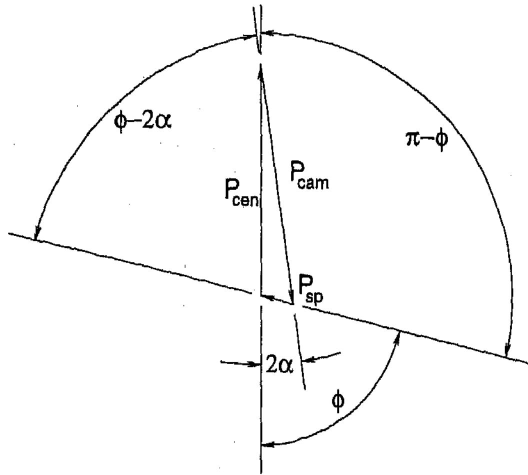


Figure 6: Force triangle for the centrifugal roller float analysis.

Another variation of this analysis (which actually is unrelated to roller float) is to develop a force triangle for the condition where the outer race is stationary, i.e. $\omega_{(o)}=0$. Performing this analysis, we obtain

$$P_{n(i)} = \frac{P_{sp}}{\tan 2\alpha} \qquad (26)$$

Using EQ (26) we obtain the nominal contact force between the roller and the rollerway as a function of the nominal spring force.

ECCENTRICITY ROLLER FLOAT - The function of the energizing spring is to position the roller for proper clutch operation. In the following analysis, proper operation is defined as maintaining contact between the roller and the rollerway during the freewheel mode of operation. In the event that the rollerway has a finite runout with respect to the cammed race, the roller will travel up and down the cam surface with each rotation of the rollerway. The energizing spring will compress and extend with each rotation, keeping the roller and rollerway in contact. This condition is illustrated in Figure 7, wherein the geometry of an outer cam OWC system has been utilized. We will be examining the outer cam OWC in this section; however, the following analysis can also be applied to a clutch having an inner cam configuration.

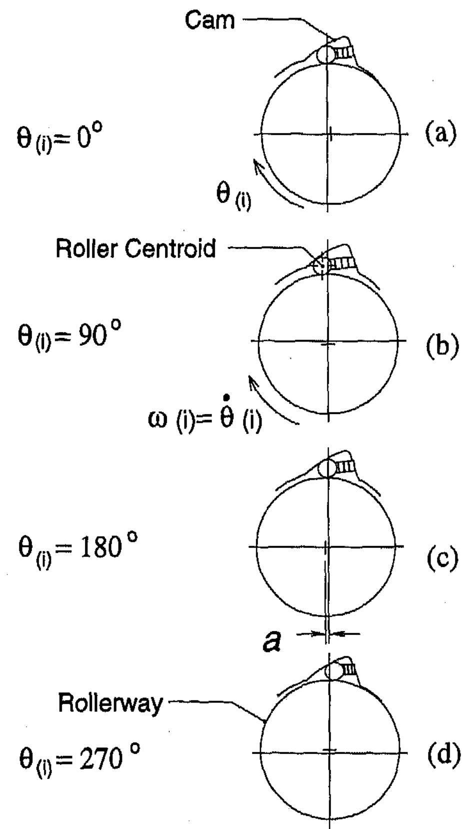


Figure 7: Roller motion due to rollerway eccentricity.

As the rotational speed of the rollerway increases, the translational inertia of the roller starts to play a significant role in the functioning of the clutch system. If the rotational speed is great enough and the eccentricity of the rollerway large enough, the roller inertia force will overcome the energizing spring force and the roller will lose contact with the rollerway. We call this condition *eccentricity roller float.* This condition is illustrated in Figure 7d, where the clutch is shown at an orientation of $\theta_{(i)} = 270°$. Elevated levels of spring compression may result from this condition, which could possibly lead to spring fatigue.

Using the method of kinematic equivalence [9], the dynamics of the roller can be examined using the analysis methods established for the four-bar mechanism [10]. A synopsis of this analysis is presented in this section, and the reader is directed to Reference [11] for a detailed discussion.

In Figure 8 the schematic of the clutch in the 0° orientation is shown. The eccentricity of the system has been exaggerated to help emphasize the definition of some of the parameters listed in the figure. It is our objective to define the position, velocity, and acceleration of point $C_{(r)}$, the centroid of the roller. To facilitate this, a number of derived dimensions are needed, and are listed below without further elaboration.

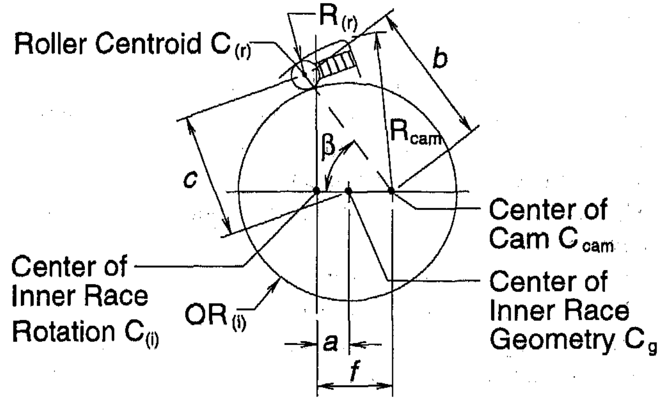


Figure 8: Definition of the *eccentricity* roller float analysis variables.

$$b = R_{cam} - R_{(r)} \qquad (27)$$

$$c = OR_{(i)} + R_{(r)} \qquad (28)$$

$$d = \sqrt{a^2 + f^2 - 2\,a\,f\cos\theta_{(i)}} \qquad (29)$$

$$\delta = \arctan\left(\frac{a\sin\theta_{(i)}}{f - a\cos\theta_{(i)}}\right) \qquad (30)$$

$$\lambda = \arccos\left(\frac{d^2 + b^2 - c^2}{2\,b\,d}\right) \qquad (31)$$

$$\beta = \delta + \lambda \qquad (32)$$

$$\gamma = \arctan\left(\frac{b\sin\beta - a\sin\theta_{(i)}}{f - b\cos\beta - a\cos\theta_{(i)}}\right) \qquad (33)$$

Note that, in the above equations, the parameter a is the eccentric offset of the geometric center C_g of the inner race from the rotational center $C_{(i)}$. The parameter β is the angular coordinate of the roller centroid with respect to the center of curvature of the cam C_{cam}. Taking the derivative of EQ (32) and EQ (33) with respect to time, we obtain the angular velocities $\dot{\beta}$ and $\dot{\gamma}$, which are given as

$$\dot{\beta} = \frac{a\,\sin(\gamma - \theta_i)}{b\,\sin(\gamma + \beta)}\dot{\theta}_{(i)} = k_\beta \dot{\theta}_{(i)} \qquad (34)$$

and

$$\dot{\gamma} = \frac{-a\,\sin(\beta + \theta_i)}{c\,\sin(\gamma + \beta)}\dot{\theta}_{(i)} = k_\gamma \dot{\theta}_{(i)} \qquad (35)$$

respectively. Therein, $\dot{\theta}_{(i)}$ is the angular velocity of the rollerway (which is also listed in this paper as $\omega_{(i)}$), and $\dot{\beta}$ is the angular velocity of the roller about the center of curvature of the cam. Note that in EQ (34) and EQ (35) we have implicitly defined the coefficients k_β and k_γ.

Taking the derivative of EQ (34) with respect to time, the angular acceleration of the roller about C_{cam} is obtained

$$\ddot{\beta} = k_\beta \ddot{\theta}_{(i)} + k'_\beta \dot{\theta}^2_{(i)} \qquad (36)$$

where the coefficient k'_β is defined as

$$k'_\beta = k_\beta\left[\left(k_\gamma - 1\right)\cot\left(\gamma - \theta_{(i)}\right) - \left(k_\gamma + k_\beta\right)\cot(\gamma + \beta)\right] \qquad (37)$$

A study of the relative magnitude of each of the two components on the right hand side of EQ (36) has shown that the term $k_\beta \ddot{\theta}_{(i)}$ is insignificant when compared to $k'_\beta \dot{\theta}^2_{(i)}$, especially at the critical orientation of $\theta_{(i)} = 270°$ [11]. Hence, with little loss of accuracy, the angular acceleration of the roller $\ddot{\beta}$ can be approximated by $k'_\beta \dot{\theta}^2_{(i)}$. The tangential component of the roller inertia force is then given as

$$P_{t(r)} = m_{(r)} b \ddot{\beta} = m_{(r)} b\, k'_\beta \dot{\theta}^2_{(i)} \qquad (38)$$

Equation (38) yields the eccentricity roller float force, $P_{t(r)}$. We may say that $P_{t(r)}$ is the force which we must supply to the roller for it to remain in contact with the inner race roller-way. This force is supplied via the energizing spring.

Another component of the roller inertia force is due to the angular velocity of the roller about the center of curvature of the cam. This force is given as

$$P_{n(r)} = m_{(r)} b \dot{\beta}^2 = m_{(r)} b\, k^2_\beta \dot{\theta}^2_{(i)} \qquad (39)$$

Equation (39) defines the contact force between the roller and the cam surface. The forces acting on the roller are depicted in Figure 9.

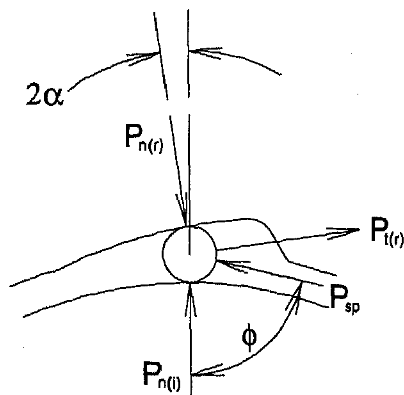


Figure 9: Forces acting on the roller for the eccentricity roller float analysis.

In the event of roller float, $P_{n(i)}$ is equal to zero. As with the previous analysis, we can build a force triangle (see Figure 10) from the remaining loads. Applying the law of sine's to the force triangle, then substituting EQ (38) and re-arranging, we obtain the rollerway speed at which *eccentricity roller float* will occur for a specified spring force

$$\dot{\theta}_{(i)} = \sqrt{\frac{P_{sp}\sin(\phi - 2\alpha)}{m_{(r)} b\, k'_\beta}} = \omega_{(i)} \qquad (40)$$

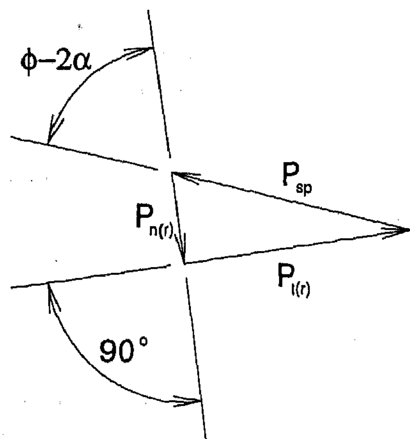


Figure 10: Forces triangle for the eccentricity roller float analysis.

As you may note from Figure 9 and Figure 10, we are neglecting the frictional forces which act between the roller and the spring, outer race, and inner race. It can be shown that by ignoring the frictional forces we compute a more conservative roller float speed.

As an alternative analysis, we can re-write EQ (40) to give us the spring force required to prevent *eccentricity roller float* for a given clutch and inner race overrunning speed. This is given as

$$P_{sp} = \frac{m_{(r)} b k'_\beta \omega_{(i)}^2}{\sin(\phi - 2\alpha)} \qquad (41)$$

wherein we have substituted $\omega_{(i)} = \dot{\theta}_{(i)}$. With the spring force obtained via EQ (41), the OWC engineer can compute the stress level in the selected energizing spring. This, in turn, can be compared to the endurance limit of the selected spring material to evaluate the spring fatigue life for the application.

SUMMARY

This paper described the static (engaged) and dynamic (freewheel) behavior of an outer cam roller one-way clutch. Explicitly derived were equations for:

- The free body forces acting on a roller,
- The limit to the strut angle based upon the operating environment,
- The contact stress on the inner race,
- The hoop stress on the outer race,
- The centrifugal roller float speed, in which the roller loses contact with the inner race due to outer race rotational speed, and
- The eccentricity roller float force, in which the roller loses contact with the inner race due to the lack of concentricity.

Although the equations were derived for outer cam roller clutches, they can easily be derived for inner cam roller clutches.

The objective of the paper was to re-derive the fundamental equations used by engineers and designers when designing roller one-way clutches. The derivation of the equations enable an engineer to better understand the mechanics of a one-way clutch. The equations are derived in terms of dimensionless units, and free the engineer from the implied constraints of using either metric or English units, exclusively.

REFERENCES

[1] Sauzedde, R.E. and Bowie, E.F., "Design of Roller One-Way Clutches in Current Passenger Car Automatic Transmissions", **Design Practices: Passenger Car Automatic Transmissions**, Second Edition, SAE, New York, NY, 1973, pp. 17-26.

[2] Fanella, R.J., "Roller One-Way Clutches", **Design Practices: Passenger Car Automatic Transmissions,** Third Edition, SAE, Warrendale, PA, 1994, pp. 441-467.

[3] Haka, R.J. and Michany, R.P., "Roller Skewing as a Factor in One-Way Clutch Operation", SAE Paper 840576, SAE, Warrendale, PA, 1984.

[4] Young, C.W., "Roark's Formulas for Stress & Strain", Sixth Edition, McGraw-Hill, Inc., New York, NY, 1989, pp. 647-665.

[5] Jones, A.B., "New Departure Engineering Data; Analysis of Stress and Deflections", New Departure Division, General Motors Corporation, 1946.

[6] Kerendian, H., "Development of Roller Clutch Design Computer Program; Design Analysis Information", Hydra-Matic Division, General Motors Corporation, May 6, 1985.

[7] Juvinall, R.C., "Engineering Considerations of Stress, Strain, and Strength", McGraw-Hill, New York, NY, 1967.

[8] Timoshenko, S., "Strength of Materials", Van Nostrand Reinhold, New York, NY, 1958.

[9] Jensen, P.W., "Cam Design and Manufacture", Second Edition, Marcel Dekker Inc., New York, NY, 1987.

[10] Paul, B., "Kinematics and Dynamics of Planar Machinery", First Edition, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[11] Kremer, J.M., "Roller Float as a Consideration in Outer-Cam, Roller One-Way Clutch Design", SAE Paper 950670, SAE, Warrendale, PA, 1995.

[12] Van Vlack, L.H., "Materials for Engineering; Concepts and Applications". Addison Wesley, Reading, MA 1982.

APPENDIX A: RACE DESIGN

This section contains an annotated example of a typical design sequence of a roller clutch. The discussion shows the practical application of the earlier described equations that answers:

Given the following load and materials, is it possible to construct an outer cam roller clutch in the size indicated by the packaging constraints?

A. Inputs

Cam Type:	Outer Cam
Packaging Constraints:	
Minimum Effective Radius ($IR_{(min)}$)	30 mm
Maximum Effective Radius ($OR_{(max)}$)	60 mm
Maximum Length ($L_{(max)}$)	12 mm
Materials:	
Outer Cam	1020 Steel
Inner Race	1020 Steel
Rollers	1020 Steel
Torque:	800 N-m

B. Approach

1. Set the effective inner radius of the inner race ($IR_{(i)}$) equal to the minimum effective radius ($IR_{(min)}$). That is, we are using the packaging constraint for the minimum radius as the smallest possible dimension that the roller clutch may have.

2. Calculate the outer radius of the inner race ($OR_{(i)}$) based upon the equation derived for contact stress (EQ (18)).

3. Determine the effective inner radius of the outer race ($IR_{(o)}$) based upon the outer radius of the inner race ($OR_{(i)}$), roller radius ($OR_{(r)}$), and strut angle (α).

4. Calculate the effective outer radius of the outer race ($OR_{(o)}$) based upon the equation derived for hoop stress (EQ (22)).

5. Compare the effective outer radius of the outer race ($OR_{(o)}$) with the maximum effective radius ($OR_{(max)}$). If the packaging constraint ($OR_{(max)}$) is less than the calculated radius ($OR_{(o)}$), then a roller clutch with the given materials and load requirements cannot be constructed in the given packaging constraints.

C. Calculations

The initial goal of the calculations is to determine the outer radius of the inner race ($OR_{(i)}$) using EQ (18). The calculation of $OR_{(i)}$ is dependent upon c_1 (EQ (13)). The inner race and rollers are both 1020 steel, and therefore the material properties are [12]:

$\upsilon_{(i)}$	0.29
$E_{(i)}$	2.07E11 Pa (30E6 psi)
$\upsilon_{(r)}$	0.29
$E_{(r)}$	2.07E11 Pa (30E6 psi)

and c_1 is calculated as 1.489E5.

The outer radius of the inner race ($OR_{(i)}$) is calculated using c_1 and some additional inputs. Specifically, the inputs that are required are the torque ($T_{(i)}$), default values for the rollers ($N_{(r)}$, $L_{(r)}$, $OR_{(r)}$), the strut angle (α), and the mean contact stress for the inner race ($\sigma_{c(mean)}$). The values are as follows:

$T_{(i)}$	800 N-m (590.05 ft-lbs)	(input)
$N_{(r)}$	12	(default)
$L_{(r)}$	0.012 m (12 mm)	(default)
$OR_{(r)}$	0.004 m (4 mm)	(default)
α	0.0698 rad (4°)	(default)
$\sigma_{c(mean)}$	3.79E9 Pa (550,000 psi)	(empirical limit)

Substituting the above values into EQ (18) results in the outer radius of the inner race ($OR_{(i)}$) equaling 0.03409 m (34.09 mm). Based upon the geometry of the inner race and

rollers, the effective inner radius of the outer race ($IR_{(o)}$) is 0.043023 m (43.023 mm).

The next step is to solve for the effective outer radius of the outer race ($OR_{(o)}$) using EQ (22). However, prior to solving for $OR_{(o)}$, the pressure on the inner surface of the outer cam ($Q_{(o)}$) is required and is calculated using EQ (20). We use the following values:

$N_{(r)}$	12	(default)
$L_{(r)}$	0.012 m (12 mm)	(default)
$OR_{(r)}$	0.004 m (4 mm)	(default)
$\sigma_{c(mean)}$	3.79E9 Pa (550,000 psi)	(empirical limit)
c_1	1.489E5	(EQ (13))
$OR_{(i)}$	0.03409 m (34.09 mm)	(EQ (18))
$IR_{(o)}$	0.04302 m (43.02 mm)	
$L_{(o)}$	0.01200 m (12 mm)	(input)

These values are used to determine the pressure on the inner surface of the outer race ($Q_{(o)}$ = 1.0346E8 N/m^2).

Finally, we assume that the endurance limit for the outer race material is 4.48E8 Pascal (65,000 psi), and set the mean hoop stress of the outer cam ($\sigma_{h\ (mean)}$) to this value. Then we substitute the pressure on the inner surface of the outer cam ($Q_{(o)}$) into EQ (22), and obtain the effective outer radius of the outer race ($OR_{(o)}$ = 0.05443 m = 54.43 mm).

The conclusion from this section is that it is possible to design an outer cam roller clutch that meets the input requirements and has the following dimensions:

$OR_{(r)}$	4.00 mm
$IR_{(i)}$	30.00 mm
$OR_{(i)}$	34.09 mm
$IR_{(o)}$	43.02 mm
$OR_{(o)}$	54.44 mm

and withstands a 3.79E9 Pascal (550,000 psi) mean contact stress on the inner race and a 4.48E8 Pascal (65,000 psi) mean hoop stress on the outer race. Note that the original maximum effective radius ($OR_{(max)}$) defined in INPUTS was 60 mm. Therefore, since the calculated outer radius of the outer race ($OR_{(o)}$ = 54.43 mm) is less than the allowed outer radius of the outer race ($OR_{(max)}$ = 60 mm), then there is additional space that is usable to increase the safety factor for the one-way clutch.

APPENDIX B: DYNAMIC ANALYSIS

In this appendix we will explore the dynamics of the hypothetical clutch design in the previous section. To enable us to do this, we will assume a duty cycle for the clutch, create an energizing spring specification, and finalize the outstanding cam design parameters. In Table 1 the hypothetical operating conditions for our clutch are given, and in Table 2 we have the cam and energizing spring design parameters. Note that the cam data presented in Table 2 was computed given the data developed in Appendix A. These computations, however, are outside the scope of this paper.

Table 1: Duty Cycle for the Hypothetical Roller Clutch

Gear	Race Speed (rpm)			Torque (N-m)
	Inner	Outer	Overrun	
R	-4800	-4800	0	0
1	0	0	0	800
2	0	0	0	250
3	4000	0	4000	0
4	8800	0	8800	0

Table 2: Assumed Cam and Spring Design Parameters

Parameter	Variable	Value
Roller, Rollerway and Cam Data:		
Roller Mass Density	$\rho_{(r)}$	7800 kg/m^3
Critical Rollerway Orientation	$\theta_{(i)}$	270°
Cam Radius	R_{cam}	42.464 mm
Cam Offset	f	5.353 mm
Rollerway Eccentricity	a	0.051 mm
Spring Data:		
Free Height	H_f	13.72 mm
Nominal Working Height	H_w	9.53 mm
Spring Rate	k	0.3 N/mm
Nominal Spring Force	P_{sp_nom}	1.47 N
Spring Orientation	ϕ	85°

From Table 1 we see that the maximum overrunning speed for the clutch occurs in fourth gear, and is 8800 rpm. This occurs with the outer race grounded. These are the conditions for which we will evaluate the clutch for *eccentricity roller float.*

In reverse we see that both races are rotating, but no torque is being transmitted. From this we can deduce that the inner race is driving the outer race, and is working against outer race inertia and friction clutch drag alone. To function properly in this environment our roller clutch must remain engaged while in reverse gear, or else the clutch will disengage due to *centrifugal roller float* forces. This would be followed by an impact re-engagement when the outer race and clutch slow down due to friction clutch drag. Such a sequence of events would lead to degradation and possibly failure of the OWC.

Applying the clutch data given above to the equations provided in the ROLLER DYNAMICS Section, we obtain the computed data given in Table 3. Note that in applying these equations, we must use consistent units. For example, when using the SI system, *meters* (instead of millimeters) should be used throughout. Also, angular coordinates must be entered in *radians* and angular velocities in *radians per second,* regardless of the measurement system employed. These

conversions have been done for the reader in this example problem.

Table 3: Intermediate Computational Results for Roller Dynamic Analysis

Variable	Equation No.	Value
$m_{(r)}$		0.0047 kg
b	27	0.0385 m
c	28	0.0381 m
d	29	0.0054 m
δ	30	-0.0094 rad
λ	31	1.4312 rad
γ	33	-1.5613 rad
$\dot{\gamma}$	35	1.3119 rad/sec
β	32	1.4217 rad
$\dot{\beta}$	34	-0.0829 rad/sec
$\ddot{\beta}$	36	8046 rad/sec^2
k_γ	35	1.424 E-3
k_β	34	-8.996 E-5
k'_β	37	9.474 E-3

In applying the results of Table 3 to the analysis of *eccentricity roller float*, we find that we can either assume that the spring force remains constant during the cycle depicted in Figure 7, or we can compute the change in the spring force (i.e. due to the kinematic roller travel) as the clutch progresses from configuration (c) to configuration (d) in Figure 7. Doing the latter, we define the roller travel as

$$H_\delta = b \cdot \left(\beta|_{\theta=270} - \beta|_{\theta=180} \right) \qquad (42)$$

where H_δ is the change in the spring height and, for example, $\beta|_{\theta=270}$ is the value of β when $\theta_{(I)}$ equals 270°. For our hypothetical clutch, the roller travel (and, hence, the additional spring compression) is computed to be 0.365 mm. This translates to an increase in spring load of 0.13 N. Therefore, the spring force P_{sp} that we use in EQ (40) is 1.60 N. (Note that if we assume that the spring force is constant, we will be inducing a "conservative" error of 8.8% into our computations for *this* example application.) Applying EQ (40) we find that the *eccentricity roller float* speed is 952 radians per second (9090 rpm). From these results we conclude that our clutch will operate with a very small margin of safety in the freewheel mode of operation.

When the transmission is in reverse, we must apply our *centrifugal roller float* analysis. Applying the results of Table 3 to the equations of that section, we note that we must use the nominal spring force in EQ (25). The result of this exercise is a computed centrifugal roller float speed of 239 radians per second (2286 rpm). This is significantly less than the 4800 rpm required by the application. Hence, we must conclude that our hypothetical energizing spring is inadequate in maintaining the clutch in an *engagement ready* state, and that we must go back and redesign the spring-roller system.

970686

A Study of Forces Acting on Rings for Metal Pushing V-Belt Type CVT

Hideaki Yoshida
Honda R&D Co., Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Four forces act in rings for a metal pushing V-belt. These forces are: two kinds of intercepting forces which prevent blocks from going outside of pulleys (one caused by pulley thrust, the other caused by centrifugal force), frictional force acting between the rings and the blocks, and bending force in longitudinal direction.

In the previous paper [1] [2] [3] [5], distribution of three forces, excluding centrifugal force, were presented at low belt speed. We successfully measured all four kinds of forces including centrifugal force continuously at practical operation conditions for layered rings. In this paper, distribution of these four forces on the innermost ring is described at steady states.

INTRODUCTION

A metal pushing V-belt (in short, metal V-belt) which is used in Continuously Variable Transmission (CVT), is composed of steel rings and blocks. Fig.1 shows the structure of the metal V-belt assembly. Various forces act on these rings. To grasp forces acting on rings is very important for argument of a metal V-belt's durability. It has been shown experimentally that the durability of a metal V-belt depends on belt speed. To grasp forces acting on rings, including centrifugal force, at high belt speed is very important for design of transmission. Although there have been some papers on the study of forces acting on rings recently, it has not been shown what the distribution of forces is acting on rings at high belt speed. The purpose of this study is to reveal the effect on the metal V-belt's life by measuring the distributions of the four forces acting on rings at practical operation conditions.

EXPERIMENT

STRUCTURE OF A METAL V-BELT

A metal V-belt developed by Van Doorne's Transmissie b.v. consists of a number of thin, flat rings and metal V-blocks which fill the entire length of the rings. Fig.1 shows the metal V-belt assembly. Each ring is approximately 0.2mm thickness. Thin blocks are strung together by 2 sets of the thin rings positioned in slots at each sides of the blocks.

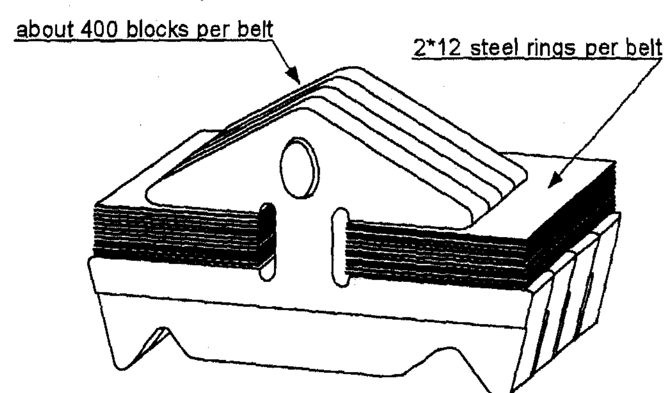


Fig.1 Structure of the metal V-belt assembly

EXPERIMENTAL APPARATUS

Fig.2 shows an assembly drawing of a testing machine for the metal pushing V-belt type CVT. The driving pulley is placed on a slide rail, held by a load cell. The force which acts between the two pulley shafts is measured by the load cell. The driving pulley is rotated by an electric motor. The driven pulley is connected to an eddy-current dynamometer.

Table 1 Test conditions

Speed ratio i=Nin/Nout	2.46, 1.00, 0.61, 0.46
Rotational speed of the driving pulley Nin [rpm]	100, 1000, 2000
Transmitted torque Tin [Nm]	0.0
Thrust of the driven pulley Qdn [N]	17.7, 35.5, 53.2
Oil temperature T_{ATF} [°C]	40

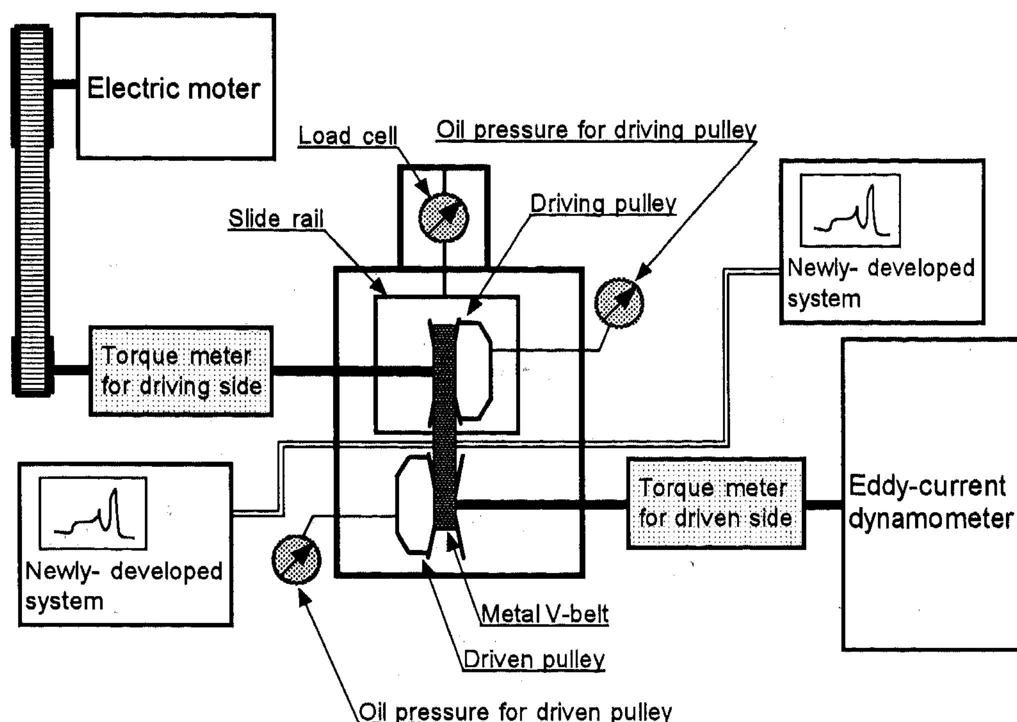


Fig.2 Assembly drawing of a testing machine for the metal pushing V-belt type CVT

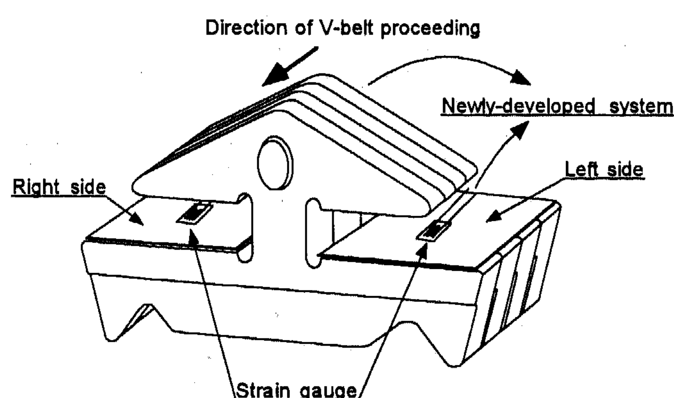


Fig.3 Location of the strain gauges on both innermost rings

EXPERIMENTAL PROCEDURE

Strain gauges are pasted on outer surfaces of both innermost rings. These are the only rings which are set in the metal V-belt (Fig.3). Signals from strain gauges are continuously measured in practical operation conditions by a newly-developed system. Test conditions are given in Table 1 with the definition of each measured item. Speed ratio is defined as rotational speed of the driving pulley divided by that of the driven pulley.

RESULTS AND DISCUSSION

DISTRIBUTION OF RING STRESS

Fig.4,6,7 and 8 show distribution of ring stress for several ring speeds and for various driven pulley thrusts. Abscissa denotes normalized time. Ordinate denotes ring stress.

A distribution of ring stress is reviewed using one of the test conditions as a sample : speed ratio i=0.61, rotational speed of the driving pulley Nin=1000[rpm], driven pulley thrusts Qdn=17.7[N], transmitted torque Tin=0.0[Nm] (Fig.4). In the figure, region a-b and region c-d designate that the strain gauge is in the driving and driven pulley grooves, respectively. Region b-c is the string where the metal V-belt goes from the exit of the driving pulley to the entrance of the driven pulley, and region d-a designates the string where the metal V-belt goes from the exit of the driven pulley to the entrance of the driving pulley. In the figure, ring stress of the left side is moved right slightly, for easier understanding. Definition of right and left sides for rings are shown in the figure 3.

(1) Distribution of ring stress in both pulleys

The ring stress increases suddenly because of the bending stress in region a-b and region c-d. Ring stress in these regions is different for each cycle. Especially, distribution of ring stress in the driven pulley grooves is not equal to that of strain wave figure model (Fig.5(1)) , but changes in a complicated manner.

Saddle surface speed takes different value on each pulley due to the separation between saddle surface and pitch radius. Saddle surface speed is faster than ring speed on the small pitch radius pulley, which will cause unstable behavior of blocks. It seems that the shape of the saddle surface and tilting motion due to the unstable behavior of blocks have detrimental effects on bending stress in the driven pulley grooves.

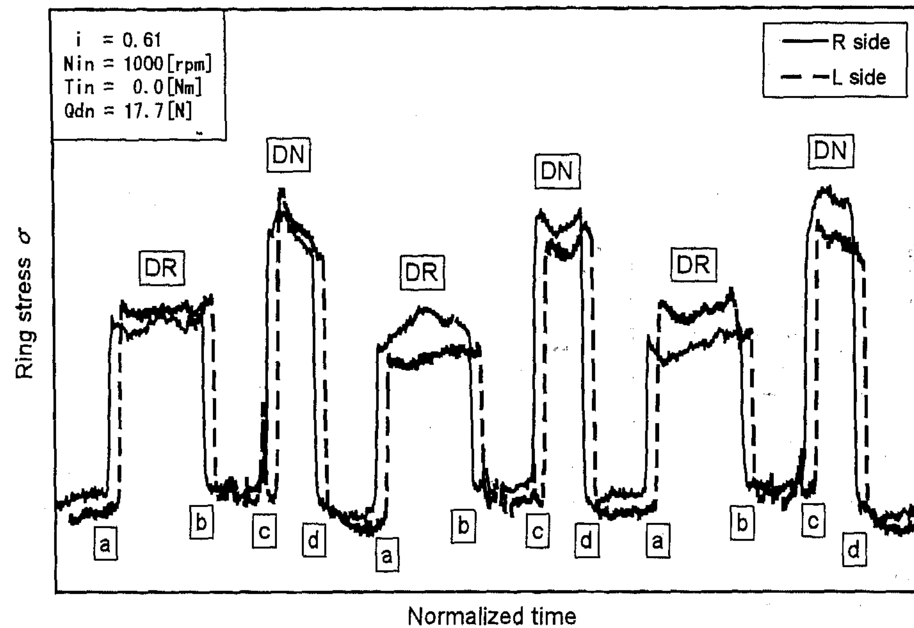


Fig.4 Distribution of ring stress

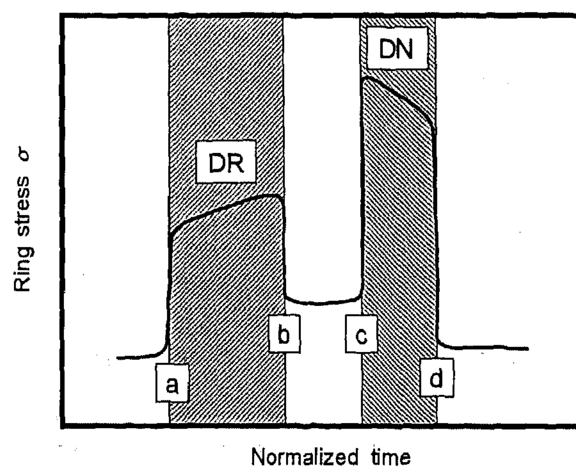


Fig.5(1) Model of strain wave figure

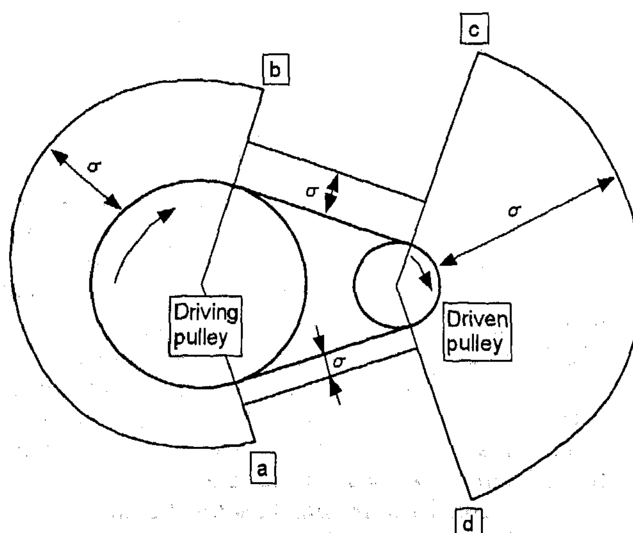


Fig.5(2) Graphic distribution of ring stress

(2) Ring stress at strings

Fig.6 shows distribution of ring stress at several speed ratios. Ring stress in region b-c and region d-a is equal at speed ratio i=1.0. However, at speed ratio i=2.46, ring stress in region d-a is higher than region b-c. And at speed ratio i=0.46, ring stress in the region b-c is higher than the region d-a. This means that the string where the ring stress is higher changes with change of speed ratio. Therefore, ring stress at the string which is the entrance of the small pitch radius pulley is usually higher. Fig.4 and 8 show the same tendency of the ring stress.

In Fujii's paper [5], it is described that difference of ring stress between strings is caused by speed difference of saddle surfaces between two pulleys.

By measuring ring stress at the strings at high belt speed in this study, this tendency which is already revealed at low belt speed is confirmed.

(3) Distribution of ring stress at both right and left sides

Ring stress at region b-c and region d-a is equal on both right and left sides. However, bending stress acting on rings in both pulley grooves is not equal on both sides. Especially, this tendency appears notably in the driven pulley. And this tendency is the same for all conditions. It seems that bending stress changes in a complicated manner due to a little difference in shape of each right and left saddle surface and a little difference in tilting motion of blocks. This is the same reason as considered in the "distribution of ring stress in the both pulleys".

Difference in ring stress acting on both right and left layered rings is a significant factor for life of the metal V-belts. To reveal the reason for this phenomenon is very important for assuming the metal V-belt's life, which is the purpose of this study

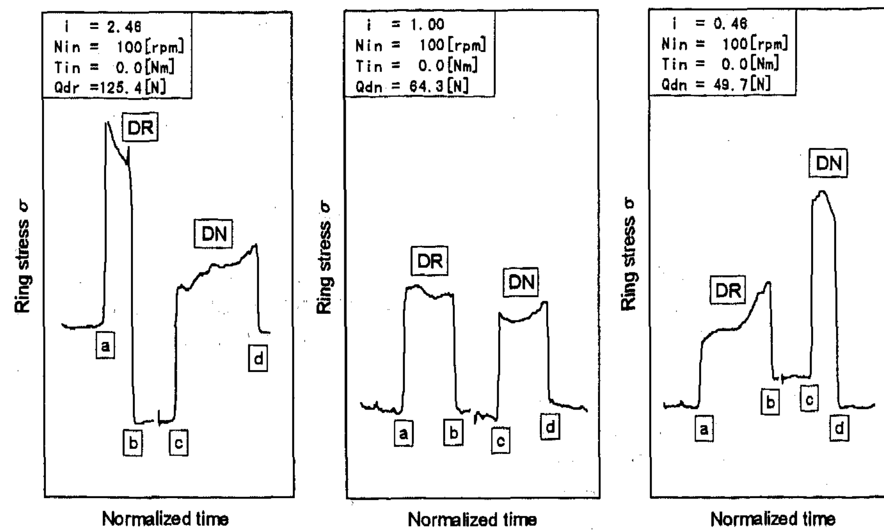


Fig.6 Distribution of ring stress at several speed ratio

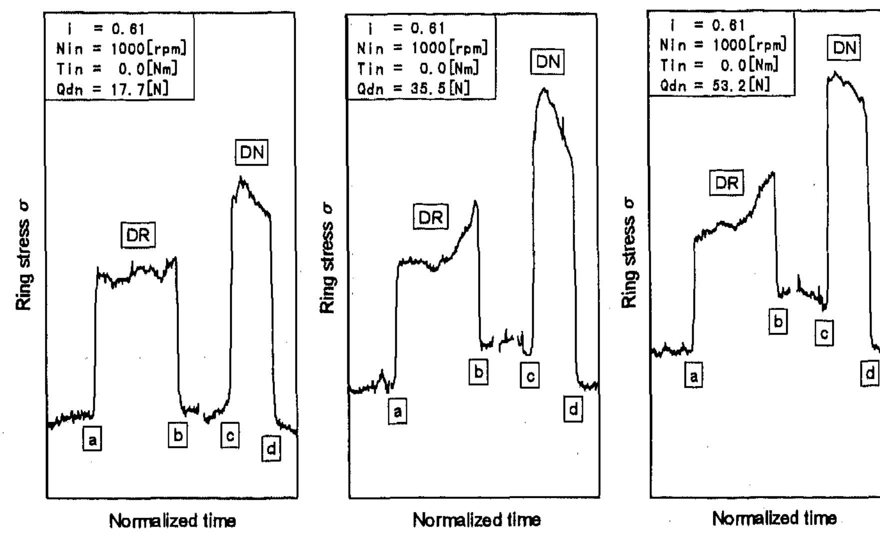


Fig.7 Distribution of ring stress for several driven pulley thrusts

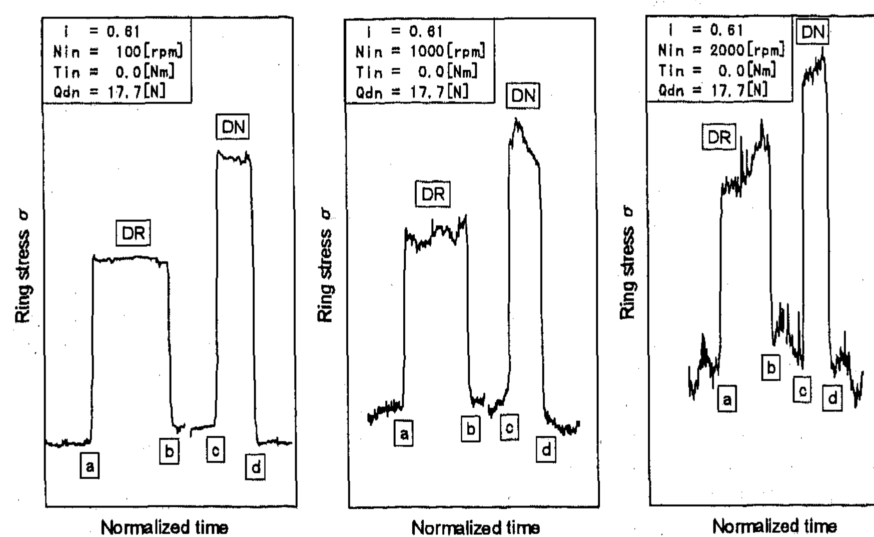


Fig.8 Distribution of ring stress at several rotational speed of the driving pulley

RELATIONSHIP BETWEEN RING STRESS AT THE STRINGS AND THRUST OF THE DRIVEN PULLEY

Fig.7 shows distribution of ring stress for several driven pulley thrusts. Ring stress in both strings is increasing and difference in ring stress between the two strings is growing with increasing driven pulley thrust.

Eular's theory is usually used to analyze the behavior of elastic flat belts. Rings of the metal V-belt do not have the properties of gum elastic. However, Eular's theory may be able to be applied for analysis of this tendency.

Ring speed and saddle speed is different in the small pitch radius pulley. Namely, the innermost rings of both right and left sides are sliding on the saddle surfaces in the small pitch radius pulley. It is assumed that the active arc is spread over the inside surface of the ring which contacts the saddle surfaces in the small pitch radius pulley. The coefficient of friction between the rings and the saddle surfaces is calculated by use of the following Eular's expression.

$$\frac{T_1}{T_2} = e^{\mu \theta}$$

Each character are defined as follows :

T_1 : ring tension for tensile side
T_2 : ring tension for slack side
θ : wrap angle of active arc
μ : coefficient of friction between rings and the saddle surfaces

At several conditions, calculated coefficient of friction is approximately constant : μ=0.09~0.11. This calculated coefficient of friction is equivalent to the coefficient of friction between pulley and blocks. The calculated results are the same at several speed ratios. Therefore, relationship between speed ratio and ring tension ratio (=T_1/T_2 : ring tension for tensile side divided by ring tension for slack side) can be estimated easily. It seems that Eular's theory can be applied for analysis of the ring tension in a metal V-belt.

RELATIONSHIP BETWEEN RING STRESS AT THE STRINGS AND BELT SPEED

Fig.8 shows the distribution of ring stress at several rotational speeds of the driving pulley (Nin). Ring stress at the strings is growing with increasing rotational speed. Ring stress at the strings is constant at rotational speed : Nin=100[rpm]. However, it seems that ring stress at the strings changes in a complicated manner at rotational speed : Nin=1000,2000[rpm]. Especially, this tendency appears in region d-a. In this region d-a, there is no consistent block compression force. Therefore, blocks in the region d-a are free to move. Because of this situation, unstableness of block behavior increases with increasing rotational speed and blocks contact inside surface of the innermost rings. This is the reason why ring stress in region d-a is not constant but changes in a complicated manner.

Fig.9 shows the relationship between ring stress in the strings and rotational speed of the driving pulley. In this figure, ring stress at the strings is defined as an average value of ring stress in region b-c in Fig. 8. It seems that ring stress at the strings is growing on the traces of quadratic curves. Therefore, the relationship between centrifugal force acting on rings and belt speed is expected to be a quadratic curve. The calculated value by use of Kanehara's expression [8] is plotted in Fig.9. The experimental and calculated value are approximately coincident. The coincidence of the experimental value and calculated value confirm the reliability of Kanehara's expression.

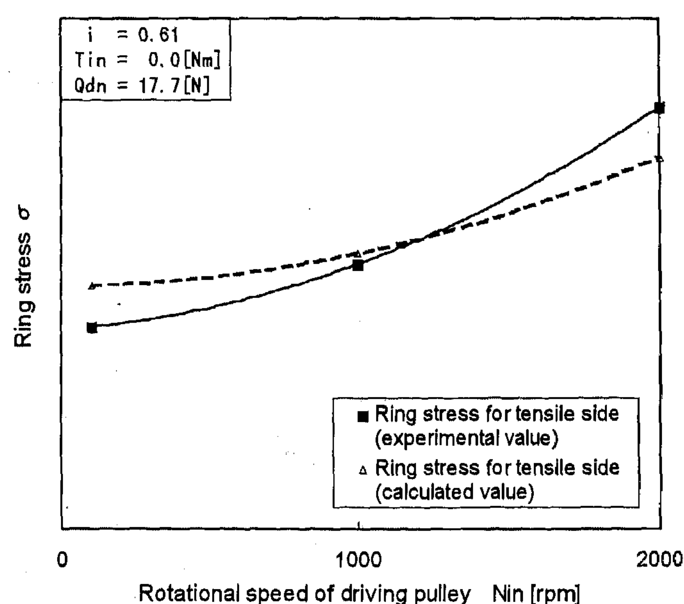


Fig.9 Relationship between ring stress at the strings and rotational speed of driving pulley

CONCLUSION

By measuring all four kinds of forces including centrifugal force continuously at practical operation conditions, the following conclusions have been deduced.

(1) Distribution of ring stress in both pulley grooves is not coincident with a model of strain wave figure which is presented in previous papers, but changes complicatedly.

(2) Ring stress acting on right and left rings is equal at strings, but bending stress acting on both rings is not equal between the two pulleys.

(3) Eular's theory can be applied for the analysis of the ring tension in a metal V-belt with the assumption that the active arc area is spread over the inner surface of the ring which contacts the saddle surfaces in the small pitch radius pulley.

(4)Relationship between centrifugal force acting on rings and belt speed is a quadratic curve.

REFERENCES

(1) K. Katuya et al. ; Analysis of Behavior of CVT Belt, Proc. of JASE 891041 (1989) pp.171-174

(2) S. Ookawa et al ; Stress Measurement of CVT Steel-ring by Strain-gauge, Proc. of JSME 1701 (1992) pp.285-287

(3) H. Asayama et al ; Mechanism of Metal Pushing Belt, Proc. of JSAE 9432804 (1994) pp.121-124

(4) T. Fujii et al ; A Study of a Metal Pushing V-Belt Type CVT - Part 1: Relation Between Transmitted Torque and Pulley Thrust, SAE Tech. Paper Series 930666 (1993)

(5) T. Fujii et al ; A Study of a Metal Pushing V-Belt Type CVT - Part 2: Compression Force Between Metal Blocks and Ring Tension, SAE Tech. Paper Series 930667 (1993)

(6) S. Kanehara et al ; A Study on a Metal Pushing V-Belt Type CVT (Part 3: What Forces Act on Metal Blocks?), SAE Tech. Paper Series 940735 (1994)

(7) S. Kanehara et al ; A Study on a Metal Pushing V-Belt Type CVT (Part 4: Forces Act on Metal Blocks when the Speed Ratio is Changing), SAE Tech. Paper Series 950671 (1995)

(8) S. Kanehara et al ; A Study on a Metal Pushing V-Belt Type CVT Macroscopic Consideration for Coefficient of Friction between Belt and pulley, Proc. of JSAE CVT '96Yokohama Paper Series 9636277 (1996)

970729

Applications and Advances in Autodeposition Coatings

Thomas C. Jones and William E. Fristad
Parker Amchem, Henkel Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Increased restrictions on air and water emissions have led to expanded automotive industry applications for the environmentally-friendly autodeposition process which eliminates chemical conversion coatings, toxic metals and volatile organic compounds (VOC). Underbody suspension components and seating assemblies are currently being painted via autodeposition at many facilities worldwide, several of which have converted existing equipment to save on installation costs.

Daily line operations, control and maintenance of the process will be emphasized. Recent process improvements will also be discussed.

INTRODUCTION

The CLEAN AIR ACT AMENDMENTS (CAAA) of 1990 are now having an impact in the selection of finishing equipment and materials for the appliance industry. The delay in enactment of specific appliance regulations toward the type and quantity of organic emissions from a finishing operation has not diminished the Clean Air Act's impact on the selection of equipment and related processes. Several state environmental agencies have taken the initiative to enact regulations that are more restrictive and more timely than federal agencies. The large investment and relatively long life of finishing equipment mandates that regulations taking effect in 2000 or 2010 must be addressed with the best available technologies today.

Autodeposition is a totally-waterborne painting process which has enabled manufacturers and their toll-coating suppliers to improve the quality of finished products while achieving compliance with increasingly-restrictive environmental codes. The autodeposition coating process offers many advantages:

- Zero VOC content and high coating transfer efficiency performance to meet future air and water quality regulations.
- Less floor space and lower capital expense required than conventional phosphate plus topcoat systems.
- Excellent physical properties such as flexibility, hardness, and solvent resistance.
- Good performance in severe testing environments such as cyclic corrosion and extended salt fog exposure.
- Lower temperature cure (stoving) requirements, enabling entire assemblies to be coated without damage to temperature sensitive components. (Energy inputs are also dramatically reduced).
- Uniform coating thickness control and the ability to coat recessed areas and complex assemblies (i.e., no limitations on "throwing power" or "Faraday cage" effects).
- Safe operation due to absence of flammability

and externally applied current to create the coating reaction.

Ease of operations due to a minimum of chemical control checks and equipment requirements.

Since a zinc (or iron) phosphate pretreatment is **not** required, no toxic heavy metals (e.g., zinc, nickel, manganese) are present in the effluent from this process.

THE AUTODEPOSITION COATING PROCESS

Coating application occurs on metal (e.g., iron/steel) surfaces in an immersion step by the simple dissolution of the cleaned-only substrate by a mildly acidic (pH ~2.5) bath to form positive ions which immediately react with negatively charged latex particles to produce a tightly-adherent coating on the surface. The porosity of the wet film to the metal surface allows the deposition reaction to continue. Succeeding layers of coating provide a thicker barrier to diffusion resulting in increasingly slower deposition rates which effectively prevent the application of unnecessary paint. Because of this mechanism, the coating formation is diffusion controlled (self-limiting). The result is uniform film thickness over the entire part.

After the coating deposition stops, the wet film is water-rinsed to remove any paint puddles and also to render insoluble any unreacted ions. Residual water is next removed and the resins barrier crystalline structure is formed in the final drying step by heat-induced evaporation. While a drying time as low as four minutes (with an infrared oven) has been sufficient for some sheet metal parts, normal forced-air convection oven dwells range from twenty to thirty minutes to allow for substrates of any thickness and to ensure complete water removal from severely recessed areas.

Dry film thickness ranges from 0.5 to 1.0 mil (12 to 25 microns) and is easily controlled by slight changes in paint bath solids and/or the rate of paint tank circulation achieved by variable-speed propeller-type mixers. The length of immersion time in the paint bath directly affects film thickness; however, this time is almost always fixed by the production rate and tank size and is normally not varied in production.

AUTODEPOSITION: EVALUATING COATING UNIFORMITY

Autodeposited coatings are resistant to the "wedge" effect which is an increase in film thickness from the top to the bottom of the vertical dimensions of a work piece and which results from the need for unattached liquid to "flow downhill." This is a typical problem in conventional waterborne immersion paint systems which are both low in viscosity and solids. With the autodeposition process, the coating is chemically reacted to the substrate and any "wedge" effect is minimized.

In a test conducted to demonstrate this, a steel test panel was immersed in a water-reducible alkyd enamel having a viscosity of 18 seconds with a Number 2 Zahn cup. The average film thickness was 1.1 mils.

A second steel test panel was immersed in an autodeposition bath to develop the same film thickness as a baseline. (Note that commercially used autodeposited films are not this high, since 500+ salt spray hours are readily achieved at 0.6 mil.)

On each panel fifty thickness measurements were made using a 5-down and 10-across pattern. The alkyd enamel panel showed thickness variations from 0.4 to 2.4 mils. (The average "wedge" down a vertical surface was 0.5 mil to 1.7 mil.) In contrast, the panel coated by autodeposition showed an overall range of only 0.9 to 1.3 mils, with an average "wedge" variance of $\pm$ 0.2 mils.

AUTODEPOSITION: EVALUATING UNLIMITED THROWING POWER

Although conventional immersion processes have theoretically unlimited throwing power, wet films cannot be rinsed to remove puddled volumes. Furthermore, paint formulations containing volatile organics allow "solvent wash" within restricted substrate areas (where solvent vapors actually dissolve the coating). Autodeposited coatings can be rinsed and dried without removal of coatings from interior sections of parts.

One measure of "throwing power" is the ability of an immersion coating process to produce films inside a standard box:

> Two 4 x 18-inch test panels separated by 3/8-inch strips of a non-conductive plastic spacer are assembled. Both sides of the "box" are contained with adhesive tape leaving the top and bottom areas open, yet allowing liquid access only from the bottom in an immersion process.

The box is coated, water-rinsed, placed in an oven to allow paint cure, and then opened to allow a comparison of coating formation between the inner surface (masked) and the outer (exposed). With autodeposition, both inner and outer surfaces are coated to the level of immersion.

AUTODEPOSITION PERFORMANCE:

Cured film performance properties are listed in Table 1. The most common latex polymer used for deposition is a PVDC (polyvinylidene chloride) copolymer with excellent barrier properties. This resin has been widely used in the food packaging industry for many years to ensure a barrier to moisture and oxygen. The film produced is both hard (for durability) and flexible. The latter property allows post-forming of painted shapes.

INDUSTRIAL APPLICATIONS

Autodeposition is not new. This process highlighted here represents the third generation of polymer technologies. Commercial history covers twenty years. Some commercial applications are listed below.

AUTODEPOSITION APPLICATIONS

- Automotive Jacks
- Fan Blower Housings
- Automotive Lighting
- Leaf Springs
- Seat Tracks/Frames
- File Frames
- Refrigeration Condensers
- Ride Control Equipment
- Suspension Components
- Steering Components
- Brake Boosters
- Drawer Slides
- Outdoor Furniture
- HVAC Equipment

EQUIPMENT CONSIDERATIONS

Since the autodeposition process does not require a zinc or iron phosphate pretreatment, less floor space (and capital equipment costs) are required in comparison with other quality paints which do require a surface conversion coating for optimum performance. This effect is summarized in Table 2. The assumption is made that equivalent substrate cleaning (usually one minute alkaline spray followed by two minutes alkaline immersion) and post-cleaner water rinsing (normally two separate one-minute cascading stages) is required. The autodeposition process gains its time/space advantage in the absence of a chemical pretreatment process and a shorter oven due to lower curing requirements.

Similarly, a lower number of control and maintenance procedures are required (Tables 3 and 4), again primarly due to the absence of stages required by a conversion coating process.

While the specific comparisons are with a waterborne coating, similar comparisons are valid with other quality stoving finishes (e.g., powder, high-solids enamels).

The equipment involved for an autodeposition process is simple with all stages, with the exception of the coating tank, identical to those employed in other finishing processes. For this reason, retrofit of an existing coating system to autodeposition is a direction that should be followed when possible. Both conveyor and hoist designs are used.

.New equipment technology has been introduced which now gives the ability to treat the coating bath solution in-line with an ion exchange system, greatly extending bath life, reducing solid waste effluent, and increasing the transfer efficiency level to that of powder coating and electrodeposition processes.

AUTODEPOSITION APPEARANCE

Standard autodeposited paint films have a black, matte appearance with typical 60 degree gloss readings of 5-15%. The value of the autodeposited coating to the user is usually seen in terms of its overall function in corrosion protection. However, recent developments have made significant improvements in the overall appearance of the coating on parts. The appearance and gloss of the film can be controlled within the autodeposition process by varying the process chemistry. New rinse technologies have been able to increase the gloss level while simultaneously improving the overall surface uniformity. Gloss values in the 15-25% and 25-35% range are attained and the improved overall appearance can meet furniture needs. (Table 5)

ENVIRONMENTAL COMPLIANCE

Federal, state, and local agencies are generating new regulations involving environmental compliance and worker safety. Additionally, there are a number of environmental "volume of compliance" laws affecting the coatings industry today. The most significant laws are the Clean Air Act Amendments of 1990 (CAAA), Superfund Amendments and Reauthorization Act (SARA), Resource Conservation and Recovery Act (RCRA) and Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). These enactments deal with air quality, chemical inventory reporting, hazardous waste management, and waste site remediation, respectively.

A complete review of the current and proposed legislation surrounding compliance issues is far too lengthy to address in this format. The use of solvents in organic coatings manufacture and application is constantly evolving. Years ago, the coatings formulator and applicator had the luxury of focusing only on the polymeric properties of the system to meet performance requirements. Today, we must analyze every ingredient and its environmental impact on the system as a whole. This includes polymers, pigments, extender pigments, biocides, surface active agents, rheological modifiers, **and** solvents.

Autodeposition coating chemicals, because of their aqueous nature, allow the formulation of coating systems which are naturally low in coalescent solvent. There are, in fact, resin species which require only water for manufacture and application. Because of relatively low glass transition temperature (Tg), these polymeric species require no solvents and very low baking temperatures to yield continuous, well-formed, coating films.

One such (low Tg) resin species is polyvinylidene chloride (PVDC). In emulsified and pigmented form, PVDC can be applied while complying with all current and proposed air quality regulations and legislation. The most modern autodeposition process containing PVDC polymeric dispersions has the advantage of containing **no** volatile organic compounds (V.O.C.'s) as determined by the U.S.E.P.A. procedure for certifying quantity of volatile organic compounds emitted by paint, and other coatings (know as **Federal Reference Method 24**, and/or ASTM Method D-3960-92).

VOC DETERMINATION: FEDERAL REFERENCE METHOD 24 AND AUTODEPOSITION COATING (PVDC)

For the determination of VOC in a specific material, the EPA has utilized "Federal Reference Method 24" which combines several test methods published by the American Society for the Testing of Materials (ASTM). The key method employed is ASTM D 3960-92, "Standard Practice for Determining Volatile Organic Compound (VOC) Content of Paints and Related Coatings." This method provides a measure of the VOC content of solventborne and waterborne paints by determining the quantity of material released from a sample under specified bake conditions. The VOC content is calculated as a function of (1) the volume of coating less water and exempt solvents, and (2) the volume of coating solids.

A sample of the most commonly employed autodeposition coating was evaluated according to EPA Method 24. This material is supplied as a concentrated make-up/replenisher which is diluted with water prior to application.

The "as supplied" coating density, $(D_c)s$, was determined to be 9.86 lb/gal using ASTM D 1475 - "Standard Test Methods for Density of Paint, Lacquer, and Related Products." This method uses a weight-per-gallon cup. The weight percent of total volatile in the coating, $(W_v)s$, was determined to be 62.5 by ASTM D 2369 - "Standard Method for Volatile Content of Coatings." The method involves accurately weighing a paint sample before and after drying at 110C for one hour.

The weight percent water, $(W_w)s$, was determined to be 62.5 by ASTM D 3792 - "Standard Test Method for Water Content of Water Reducible Paints by Direct Injection into a Gas Chromatograph." In this procedure, the paint specimen is diluted whereby dimethylformamide, and an internal standard (2-propanol) are added, and an aliquot of the mixture is injected directly into the gas chromatograph.

The water content, in volume percent, $(V_w)s$, can be calculated by the equation:

$$(V_w)s=\frac{(W_w)s(D_c)s}{D_w}$$

Where D_w is the density of water, 8.33 lbs.gal.
In this case,

$$(V_s)s=\frac{(62.5\%)(9.86 lbs/gal)}{8.33 lb/gal}=74.0\%$$

The organic volatile content, $(W_o)s$, i.e., the VOC content expressed as a percent by weight, is determined by the following equation:

$$(Wo_s)=(W_v)s-(W_w)s=62.5-62.5=0$$

The VOC content of the "as supplied"coating, (VOC)s, is calculated and expressed in terms used by State and Federal regulations by:

$$(VOC)s=\frac{(W_o)s(D_c)s}{100\%-(V_w)s}$$

which, in the case of this PVDC based autodeposition coating is:

$$(VOC)s=\frac{(0)9.86 lb/gal)}{100\%-74\%}=0$$

It is apparent that the use of this type of autodeposition coating can eliminate any cost penalties for exhaust surcharges or the purchase/installation of pollution abatement equipment.

CONCLUSION

The autodeposition coating process reduces the complexity and capital required by competitive technologies. It requires no electric current and develops self-limiting organic coatings through a chemical reaction with the metal surface.

Because the process coats wherever it wets, it generates exceptionally uniform protective coatings even on difficult to coat parts...and does so at a fraction of the conventional energy costs.

Autodeposition coatings can be applied on the same production line for parts with many different end-use market applications. No conversion coating is required, the only pretreatment necessary is cleaning. With no volatile organic compounds, no toxic heavy metals, and low temperature curing, these environmental and operational benefits also produce a more worker-friendly environment.

The autodeposition coating process presents an opportunity to satisfy air quality requirements while still applying a high quality barrier coating. In addition, this autodeposition process saves money by generating no toxic heavy metal sludge and utilizing a low temperature (110 degree C) cure, leading to energy savings

Recent advances in the autodeposition process have resulted in better coating appearance. Improvements in both the coating uniformity and higher gloss have been achieved. Gloss values (60 degree angle) in the 5-15%, 15-25%, and 25-35% reflectance range are now possible by modifying the reaction rinse all from the same coating bath. More uniform coloring in the coating has also been demonstrated

REFERENCES

Holyk, N.M., Roberto, O.E., Jones, T.C., "Autodeposition Update", Proceedings Corrosion and Prevention Conference., SAE 1989 Paper 892560

Jones, T.C., "Autodeposition - The Environmental Advantage",

FINISHING '91, SME Technical Paper FC91-371, Society of Manufacturing Engineers, 1991.

Hart, R.G., Roberto, O.E., "The Performance of Autodeposited Coatings for Vehicle Underbody, Components", Proceedings, Automotive Corrosion and Prevention Conference, SAE, 1991, Paper 912291

Hart, R.G., Jones, T.C., "Metal Finishing by Autodeposition of Organic Coatings", Sur-Fin '92, AESF, Proceedings of Sur-Fin '92, Atlanta, GA, June 1992.

Woock, K.L., Jones, T.C., Garavanta, J., "Environmental Air Quality Legislation: The Impact of the Autodeposition Process: Surface Coatings '93 Conference, Chemical Coaters Association, Proceedings of Surface Coatings '93, Grand Rapids, MI, April 1993

Table 1

AUTODEPOSITION

Cured Film Properties

Resin Type	PVDC
Film Thickness (Mils)	0-1.0
Pencil Hardness	5H+
Zero T-Bend	No Loss
Reverse Impact	160 inch-lbs.
Cross Hatch Adhesion	No Failure
Humidity/Tape Adhesion (1000 hours)	No Failure
Water Immersion/Adhesion (240 hours)	No Failure
Gravelometer (GM-9508-P)	> 7
Scribed Salt Spray (Typical)	600 hours
Unscribed Salt Spray	2000 hours
Cyclic Corrosion Exposure	20 to 60* Cycles
Solvent Resistance	Good
Cure Temperature	200 to 230 degrees F
VOC	Zero

* Depends on the specific test and film build.

Table 2

COMPARATIVE FLOOR SPACE
(ASSUME EQUIVALENT CONVEYOR SPEED AND CONFIGURATION)

	Autodeposition	Conventional Quality Waterborne Coating*
Clean	3 minutes	3 minutes
Rinse	2 minutes	2 minutes
Conditioning Rinse	-	1 minute
Zinc Phosphate	-	2 minutes
Rinse	-	1 minute
Chrome Seal Rinse	-	1 minute
Deionized Water Rinse	1 minute	1 minute
Paint	2 minutes	3 minutes
Rinse	2 minutes	2 minutes
Oven	25 minutes	35 minutes
	35 minutes	51 minutes
	(+30% savings)	

*This particular sequence represents a typical waterborne electrodeposition process

Table 3

PROCESS CONTROL PROCEDURES PERFORMED BY LINE ATTENDANT

	AUTODEPOSITION		WATERBORNE (Electrodeposition)	
	Tests Required	Frequency/ Shift	Tests Required	Frequency Shift
1. Cleaning/Phosphate/ Rinses				
- Titrations	4	1	9	2
- Conductivity Tests (Meter)	4	1	6	1
- pH Tests	-	-	2	1
- Post-Coating Rinses Solids Test	-	-	3	1
2. Coating Bath				
- Temperature	Yes	1	Yes	1
- Film Build	Yes	1	Yes	1
- % Solids	Yes	1	Yes	1
- Pigment/Binder Ratio	No		Yes	Weekly
- Pigment Solids	No		Yes	Weekly
- Conductivity				
· Paint	Yes	1	Yes	1
· Anolyte	No		Yes	1
· Phosphate Drippings	No		Yes	1
- pH				
· Coating Bath	No		Yes	1
· Anolyte System	No		Yes	1
- MEQ Titration	No		Yes	1
- Iron Titration	Yes	1	No	
- ORP Level (Meter)	Yes	3	No	
- 101 Level (Meter)	Yes	3	No	
- Ultrafilter Flow Rate	No		Yes	1
- Voltage Level	No		Yes	7 (hourly)
- Amperage Level	No		Yes	7 (hourly)
- Oven Temperature	Yes	1	Yes	1
Total Process Control Tests Per Shift	20		54	

Table 4

MAINTENANCE

		AUTODEPOSITION	WATERBORNE (Electrodeposition)
1.	Yearly Phosphate System Boilout	No	Yes
2.	# of Stages Requiring Pump Maintenance	2	11
3.	Ultrafilter Maintenance - Membrane Cleaning - Membrane Replacement - Pumps & Seals - Rebuild (every 5 years)	N/A	Yes
4.	Paint Tank Bag Filter and Oil Filter Inspection & Maintenance	No	Yes
5.	Paint Tank Transfers/Year	1	2
6.	Oven Cleaning	No	Yes
7.	Stages Requiring Nozzle Maintenance	2	10
8.	Anolyte System Flushing and Fungus Removal	No	Yes
9.	Electrode Replacement	No	Yes
10.	Rack Stripping	Every 2 years on average	Twice a month

Table 5

COMPARATIVE GLOSS VALUES

Autodeposition Coating	60^{O} Gloss Value
Standard Rinse #1	6
Rinse #2	15
Rinse #3	32

970732

Effects of Coating Weight and Pretreatment on the Painted Corrosion Performance of Coated Steel Sheet

Theresa C. Simpson and Jay D. Hoffman
Bethlehem Steel Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A program was initiated to determine corrosion performance of coated sheet products. The program examined the performance of galvannealed, hot-dip galvanized and electroplated zinc coated steel sheet in a variety of cyclic laboratory corrosion tests. Both perforation and cosmetic corrosion susceptibility were evaluated. Materials were pretreated using immersion and spray phosphate systems and were also prepared without a phosphate pretreatment. Results of these studies will be presented in this paper.

INTRODUCTION

The purpose of this study was to determine the effects of coating weight and pretreatment type on the painted corrosion performance of galvannealed, hot-dip galvanized and electroplated steel sheet. Materials with coating weights ranging from 45g/m^2 - 150g/m^2 were tested. Materials were pretreated using immersion and spray phosphating systems and were also prepared in the cleaned-only (no phosphate) condition. Cosmetic corrosion resistance of these materials was evaluated using the GM9540P-Method B test, the Nissan CCT-IV test, the Volvo test, and SAE J2334[1-4]. These will be referred to as the 9540P, CCT4, Volvo and J2334 tests, respectively, for the balance of this document. Perforation corrosion resistance was evaluated using the 9540P test. Phosphatability was also evaluated for each material type prior to corrosion testing.

Results of these studies indicate some dependence of cosmetic and perforation corrosion resistance on coating weight of these materials. A stronger dependence on pretreatment type was observed. The studies also indicate some differences in cosmetic corrosion that are cyclic test dependent. Results of these studies will be presented in this paper.

EXPERIMENTAL

Materials used in these studies included galvanneal and hot dip galvanized produced via both flux and hot line processes, and electroplated zinc. A list of the material types and chemistry of each coating is given in Table 1.

The pretreatment processes used in these studies were based on standard automotive practices. Materials were cleaned with a non-etching alkaline cleaner prior to phosphate application and painting. Tri-cationic spray and immersion phosphates were applied after cleaning. A high build cathodic e-coat and a base-coat/clear coat were then applied. Samples without phosphate were alkaline cleaned prior to e-coat and topcoat application. Phosphate processing and painting was completed by ACT Laboratories, Hillsdale, MI.

Cyclic testing was completed using automatic cabinets for the 9540P and the CCT4 tests. The J2334 test was completed manually.

Table 1: Materials Used in Test Program

Material Code	Material Type	Average Coating Weight	Wt % Fe (Galvanneal Only)	Tests Completed on Material C= CCT4, G=9540P, V=Volvo, J=J2334
L1, 828089	Flux Line Galvannealed	56 g/m^2	10.7	G,J
L3, 832257	Flux Line Galvannealed	70 g/m^2	9.8	G,C,J
L4, 842382	Flux Line Galvannealed	60 g/m^2	11.1	G,C,J
B1, 549223	Hot Line Galvannealed	60 g/m^2	10.8	G,C,J,V
B2, H21614	Hot Line Galvannealed	45 g/m^2	9.3	G,C,J,V
B4, H15925	Hot Line Galvannealed	50 g/m^2	9.3	G
B5, H68356	Hot Line Galvannealed	91 g/m^2	10.3	G,C,J,V
G1, 846252	Flux Line Hot Dip Galvanized	150 g/m^2	-	G,C,J
G2, 644939	Hot Line Hot Dip Galvanized	70 g/m^2	-	G,C,J,V
G3, H70958	Hot Line Hot Dip Galvanized	59 g/m^2	-	G,C,J,V
G4, H70980	Hot Line Hot Dip Galvanized	84 g/m^2	-	G,C,J
E6, 698174	Electroplated Zinc	68 g/m^2	-	G,C,J,V
E7, 501915	Electroplated Zinc	80 g/m^2	-	G,C,J,V

Table 2: Phosphate Characterization of Materials

Material Code	Phosphate Coating Weight (g/m^2)		Phosphate Crystal Size (μm)	
	Immersion	Spray	Immersion	Spray
L1, 828089	3.78	3.79	5-15	15-25
L3, 832257	3.96	3.75	10-20	10-20
L4, 842382	3.91	4.11	10-20	20
B1, 549223	3.81	4.10	10-15	10-15
B2, H21614	4.46	5.06	5-15	10-20
B4, H15925	3.97	4.61	10-20	10-30
B5, H68356	3.41	3.16	5-15	10-30
G1, 846252	2.89	3.61	5-10	5-15
G2, 644939	3.10	3.66	5-10	10-15
G3, H70958	3.24	3.11	5-10	10-20
G4, H70980	3.05	2.94	3-10	5-15
E6, 698174	3.03	3.13	3-10	10-15
E7, 501915	3.54	3.87	3-10	5-10

RESULTS AND DISCUSSION

PHOSPHATABILITY - Phosphate quality, crystal size, and uniformity was evaluated for each material type prior to corrosion testing. Materials were phosphated with tri-cationic phosphates, applied by both spray and immersion methods. Average phosphate coating weights (2 samples each) and crystal size for each material type are listed in Table 2.

COSMETIC CORROSION PERFORMANCE TEST SEVERITY EVALUATION - A number of cyclic tests were evaluated as part of this study. Since nominally identical materials were exposed in each test, it was possible to make some judgments about the relative severity of these tests. Figure 1 compares average scribe creep data for all samples as a function of pretreatment type in three cyclic tests: 9540P, CCT4, and J2334. Data are plotted for 40 cycles of 9540P, 50 cycles of CCT4, and 80 cycles of J2334. Based on this comparison the severities of these tests are similar, although CCT4 may be slightly more severe and J2334 slightly less severe than 9540P. The standard deviations of the three tests are also similar.

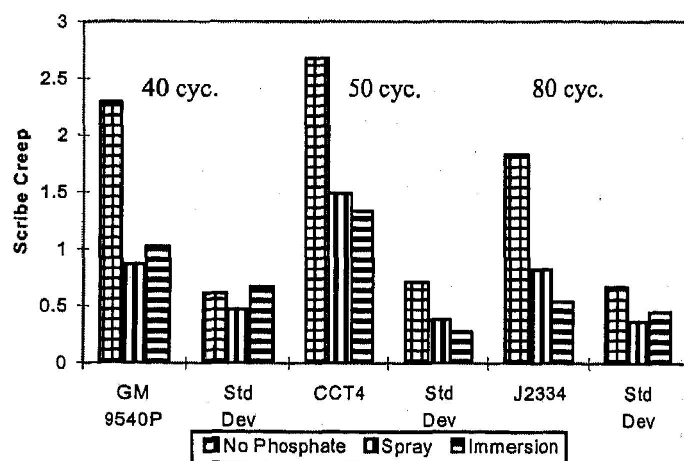


FIGURE 1-Comparison of severity of cyclic tests for various pretreatment conditions.

TEST SELECTIVITY EVALUATION, COMPARISON OF CYCLIC TEST DATA - Selectivity of each test was also evaluated. Figure 2 is a plot comparing control materials tested in the three tests. The plot shows scribe creep results for triplicate (duplicate for CCT4) samples of a cold rolled steel material (CRS) and a hot dip galvanized material (HDG) after test. Note that in both 40 cycles of the 9540P test and 50 cycles of the CCT4 test there is a relatively small difference in the scribe creep performance of these materials. In 80 cycles of J2334, however, the CRS has significantly poorer scribe creep performance than the HDG. Comparing these results to data from actual field exposures for zinc coated sheet and cold rolled steel sheet shows that the values in J2334 are much more field relevant[5]. Standard deviations for each material are also shown. The standard deviation for the CCT4 test are slightly less than for the other two tests. It should be noted that the times (# of cycles) selected for these tests were based on standard amounts of time typically used for each test. It is possible that the selectivity of the 9540P and CCT4 tests would improve if their times were extended.

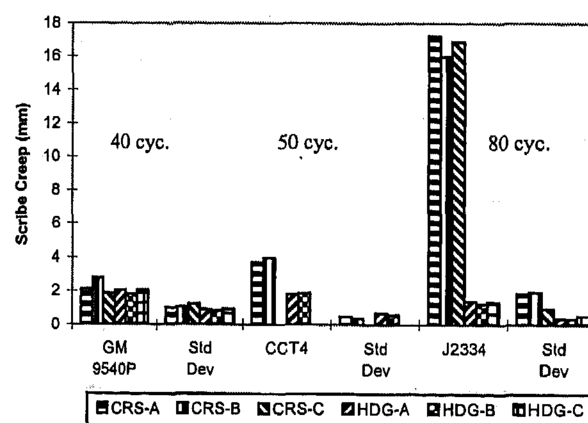


FIGURE 2-Comparison of selectivity of cyclic corrosion tests for CRS and HDG.

EFFECTS OF COATING WEIGHT ON COSMETIC CORROSION PERFORMANCE - The primary purpose of this study was to determine if coating weight and pretreatment affect the painted corrosion performance of these materials. Materials were tested in a variety of laboratory and one accelerated field test. Results are grouped according to pretreatment type. Figures 3-5 show an example of scribe creep data as a function of coating weight for materials tested in J2334. Coating weights given are zinc coating weights to enable direct comparison of galvannealed to hot dip galvanized and electroplated zinc. Note that no real trend in scribe creep as a function of coating weight occurs. Similar data were obtained for all of the cyclic tests. Figures 6-8 show the relationship

between red rust susceptibility and coating weight after 6 months of exposure in the Volvo test. Note that there is some evidence of a dependence on red rust susceptibility with coating weight. These data are early data in this test and it will be interesting to see if this dependence continues with time. No evidence for a relationship between rust susceptibility and coating weight was observed in the other tests.

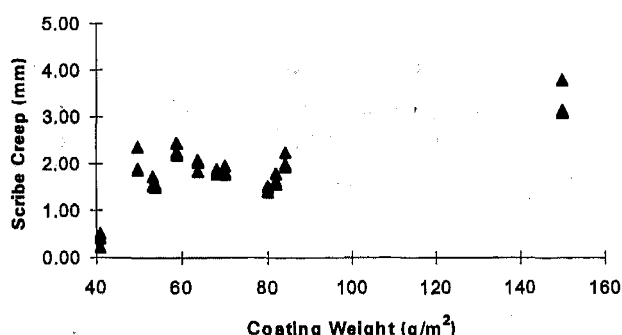


FIGURE 3-Scribe creep as a function of zinc coating weight, J2334, 80 cycles, No Phosphate Condition.

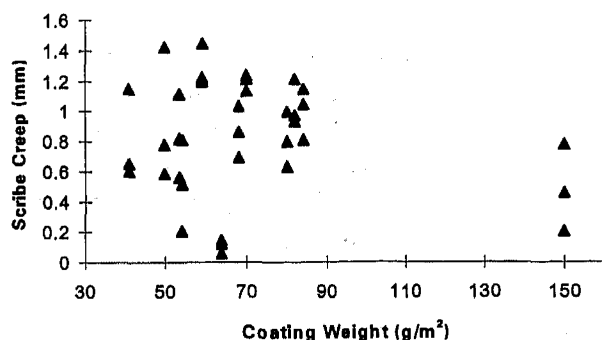


FIGURE 4-Scribe creep as a function of zinc coating weight, J2334, 80 cycles, Spray Phosphate Condition.

EFFECTS OF MATERIAL TYPE- It was also possible to compare the effects of material type on the cosmetic corrosion performance. In order to make this comparison, galvannealed, hot dip galvanized and electroplated zinc materials were selected that had coating weights of approximately 60 g/m^2 (electroplated zinc had a coating weight of ~68 g/m^2). Figure 9 compares average scribe creep for these materials. Two different galvannealed materials with similar coating weights are shown in this comparison. Data are averages of scribe creep performance in three cyclic tests - 9540P, CCT4 and J2334. Triplicate samples of each material were exposed in the 9540P test and the J2334 test; duplicates were exposed in the CCT4 test. Figure 10 compares the same materials using scribe creep information from only the J2334 test. In general, the trend in scribe creep that was shown in Figures 3-5 - no phosphate>spray phosphate>immersion phosphate - holds for these materials. All materials have good scribe creep performance and the galvannealed materials have equivalent to marginally better performance on average than the hot dip galvanized or electroplated zinc materials.

The conclusion from these studies is that cosmetic corrosion resistance is affected by the pretreatment type, and the severity of attack will differ based on the type of cyclic test performed. Cosmetic corrosion resistance has a small if any dependence on coating weight for these materials over this range of coating weights. Performance of these coated sheet products is similar; in some cases galvannealed materials have slightly better scribe creep resistance than hot dip galvanized or electroplated zinc.

PERFORATION CORROSION PERFORMANCE - Perforation corrosion resistance was also evaluated on these materials using an 80 cycle 9540P test. Pit depth was measured at the conclusion of the test. Two sample configurations were used in these studies: SAE ACAP perforation-type coupons consisting of a test specimen:cover plate sandwich with a Teflon spacer separating them and a circular uncoated region on the test specimen [6] and spot welded couples of each material to cold-rolled-steel. All perforation specimens were phosphated with a tri-cationic immersion phosphate and received a high build cathodic electrocoat.

Figures 11-13 are plots of average maximum pit depth as a function of coating weight.

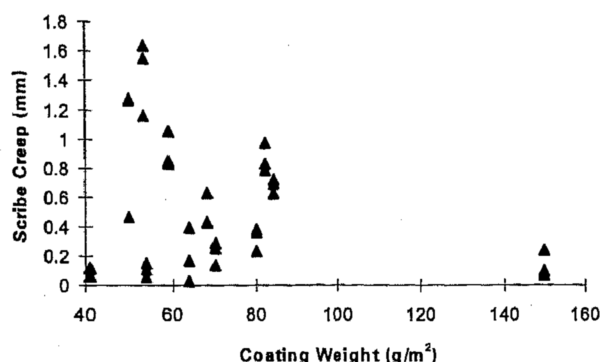


FIGURE 5-Scribe creep as a function of zinc coating weight, J2334, 80 cycles, Immersion Phosphate Condition

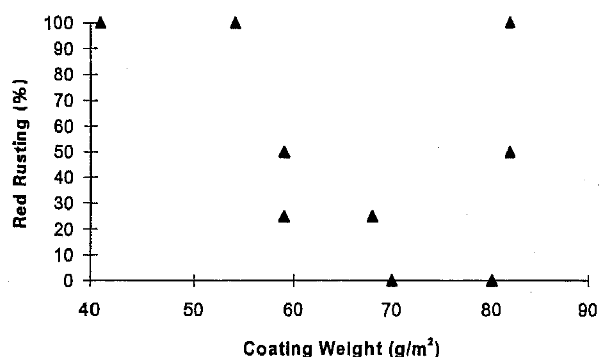


FIGURE 6-Red Rust susceptibility as a function of zinc coating weight, Volvo Test, 6 month removals, visual observation, No Phosphate Condition.

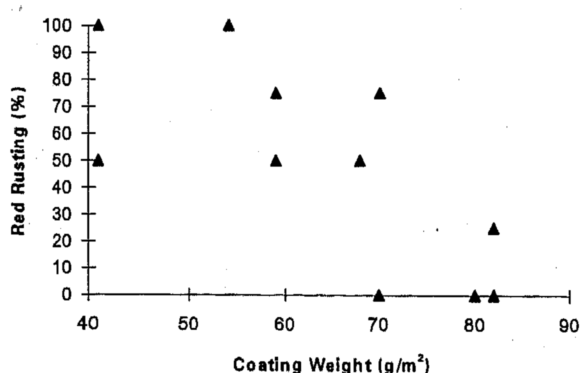


FIGURE 7-Red Rust susceptibility as a function of zinc coating weight, Volvo Test, 6 month removals, visual observations, Spray Phosphate Condition.

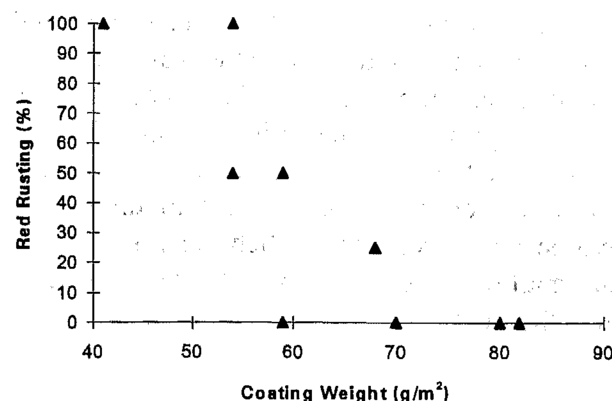


FIGURE 8-Red Rust susceptibility as a function of zinc coating weight, Volvo Test, 6 month removals, visual observations, Immersion Phosphate Condition.

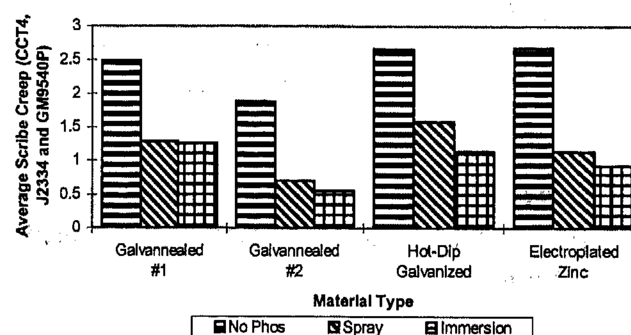


FIGURE 9-Comparison of Scribe Creep Performance as a function of material type and pretreament type. Data plotted are averages of three cyclic test results (length of test as in Figs. 1-2). All materials had ~60 g/m^2 coating weights (electroplated zinc, 68 g/m^2)

Average maximum pit depth for the spot welded couples is the average of the five deepest pits across the pitted surface. For the SAE coupons it is the average of the deepest pit in each of the four quadrants of the uncoated area of the coupon. Figures 11 and 12 are for spot welded couples; Figure 13 is for SAE perforation-type coupons. Regression lines are shown on all plots. Figure 12 includes a cold-rolled steel control material in the regression analyses; Figures 11 and 13 are for zinc coated products only. A trend in pitting susceptibility as a function of coating weight does exist, but it is only a trend and is far from a linear correlation. It is possible that this trend would have

been more pronounced if the materials were exposed for longer periods of time.

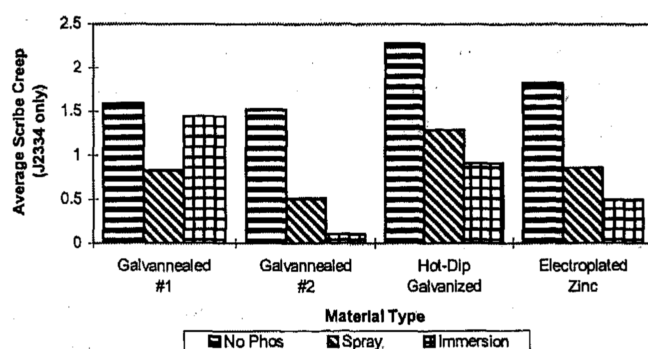


FIGURE 10-Comparison of Scribe Creep Performance as a function of material type and pretreatment type. Data plotted are averages of triplicate samples in the J2334 test (80 cycles). All materials had ~60 g/m² coating weights (electroplated zinc, 68 g/m²)

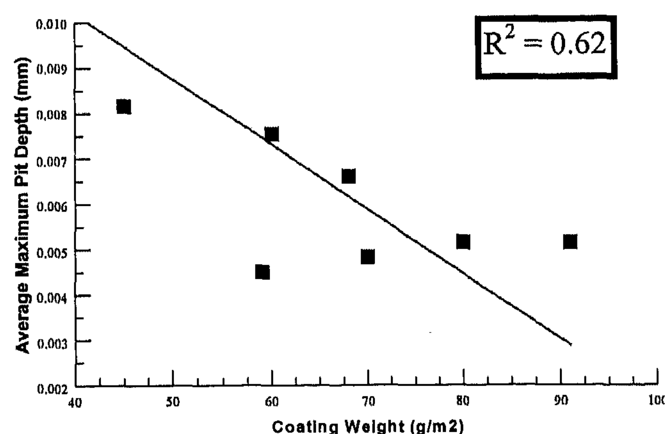


FIGURE 11-Pitting susceptibility as a function of zinc coating weight, coated products only, 80 cycles 9540P, spot welded couples of each coated product to cold rolled steel.

The conclusion from these studies is that pitting susceptibility is somewhat related to zinc coating weight and that within the range of coating weights examined it does not differ dramatically for these materials.

CONCLUSIONS

Overall, these studies showed that there appears to be a weak influence of coating weight on the cosmetic and perforation corrosion resistance of zinc coated products in the zinc coating weight range of 40-150 g/m². It should be noted that for the perforation corrosion studies the time of exposure may be one reason that a consistent effect of coating weight was not observed. Further studies need to address this by taking exposures to longer times. The early results of the Volvo test show some evidence of increased red rusting resistance with increased coating weight. Results further showed that phosphate pretreatment does influence the cosmetic corrosion resistance. The best corrosion resistance was seen with immersion phosphates.

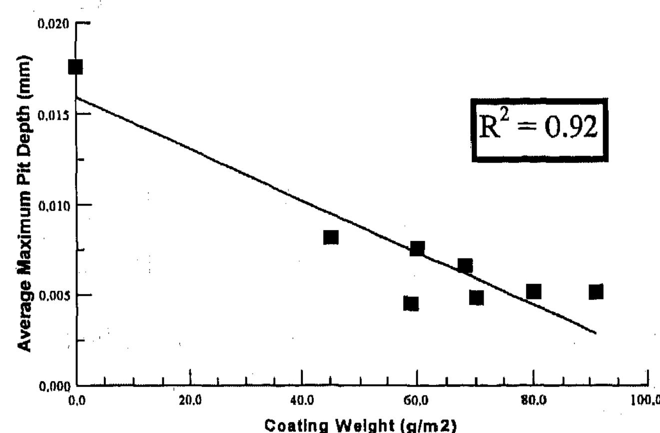


FIGURE 12-Pitting susceptibility as a function of zinc coating weight, coated products and cold rolled steel controls, 80 cycles 9540P , spot welded couples of each coated product to cold rolled steel.

Data from these studies further showed that there are significant differences in both the severity and selectivity of cyclic corrosion tests. The J2334 test appears to have the most field relevant selectivity for zinc coated vs cold rolled steel products.

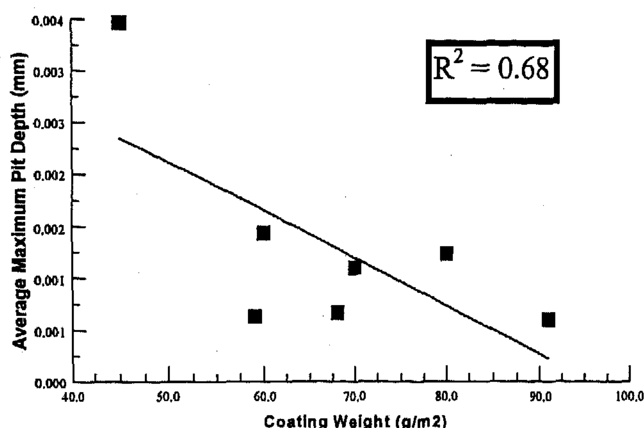


FIGURE 13-Pitting susceptibility as a function of zinc coating weight, coated products only, 80 cycles 9540P , SAE ACAP-type perforation coupons.

REFERENCES

1) General Motors Engineering Standard, "Accelerated Corrosion Test - GM9540P", 1992, Available General Motors Corporation, Detroit, MI.

2) Composite Corrosion Test, CCT-IV Method, Available Nissan Motor Manufacturing Corporation.

3) H. E. Townsend, "Development of an Improved Laboratory Corrosion Test By the Automotive and Steel Industries", Proceedings of the 4th Annual ESD Advanced Coatings Conference, ESD, The Engineering Society, Ann Arbor, MI, 1994.

4) F. Lutze and R. J. Shaffer, "An Assessment of Nine Accelerated Atmospheric Corrosion Sites on the Cosmetic Corrosion Performance of the AISI Materials - Interim Report", in Proceedings of the 5th Automotive Corrosion and Prevention Symposium, P-250, Paper # 912278, Available SAE, Warrendale, PA, 1991.

5) M. R. Ostermiller and H. E. Townsend, "On-Vehicle Evaluations of Coated and Cold Rolled Steel Sheet", Proceedings of the 6th Automotive Corrosion and Prevention Conference, Paper # 932335, Available SAE, Warrendale, PA 1993.

6) T. E. Dorsett and D. D. Davidson, "SAE ACAP Division III Report - Design of a Standard Test Coupon for Perforation Corrosion", Proceedings of the 6th Automotive Corrosion and Prevention Conference, Paper # 932366, Available SAE, Warrendale, PA 1993.

ACKNOWLEDGMENTS

Linford Hahn is acknowledged for SEM evaluations of phosphated coatings and Aparna Kumar for chemical analyses of these materials. In addition, Herb Townsend is acknowledged for technical discussions throughout these studies.

970733

Automotive Corrosion Protection Practices in Europe

Horst Gehmecker
Chemetall GmbH

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A survey of the corrosion protection practices in Europe is presented. The materials in use for the construction of the carbody and the chemical pretreatment processes applied are described in detail. Trends in electrocoating are also discussed briefly. The use of precoated/prephosphated steel and aluminum is continuously increasing. In addition trials are made with magnesium and plastic hang-on parts. In pretreatment, multi-metal and environmentally friendly processes free of nickel, nitrite and chrome are increasingly used. Some twelve body lines process aluminum hang-on parts and all-aluminum carbodies, respectively. Two body lines operate with a nickel and nitrite free phosphating process and chrome free post-treatment since early 1996. Future trends are presented for conclusion.

INTRODUCTION

Automotive corrosion protection practices in Europe include the application of pre-coated steel, corrosion protection by design and construction, seam sealers, trication phosphating, cataphoretic electrocoating, underbody coating and cavity waxes. These protection practices seem to be similar to the ones employed in North America in the first run. A closer analysis reveals interesting differences, some of them being rather distinct.

The corrosion protection practices in Europe will be discussed in detail. Particular emphasis will be put on the materials used for the car body construction, chemical pretreatment and electrocoating. The protection measures associated with design and construction, sealers and waxes are beyond the scope of this paper.

CAR BODY CONSTRUCTION MATERIALS

For car body construction, the sheet materials depicted in Table 1 are used by the European vehicle manufacturers. Some of these materials are still being investigated in pre-production trials and are not yet employed in routine production.

The ranking included in Table 1 indicates the extent to which the material is used in the European vehicle industry excluding transplants. The amount of mild steel is decreasing while the amounts of pre-coated steel and aluminum are increasing. Several car models are made 100 % of pre-coated steel sheet. In addition to the production of an all-aluminum luxury car (approximately 70 units/day), several medium volume cars are being manufactured with aluminum hang-on parts. There do exist ambitious plans by several car manufacturers to use aluminum for hang-on parts, aluminum intensive vehicles and further all-aluminum cars. By 1998, the amount of aluminum to be processed in paint shops may be 35 - 50 % in several plants.

Several manufacturers have investigated the application of cast magnesium for skin parts. Problems encountered so far are the surface quality of the material having a large impact on the finish of the painted car.

The use of plastics is still discussed controversially in Europe. By the beginning of 1998 a high volume subcompact car will be introduced which will be composed of a metal frame and plastic hang-on parts.

Pre-primed pre-coated steel has been employed to some extent over many years. There are indications that the use of pre-primed electrogalvanized steel will be increased in the near future. A new compact car which will be introduced in 1997 will have the underbody coating replaced by the use of pre-primed electrogalvanized sheet in combination with a shield made of plastic.

PRE-PHOSPHATING OF COATED STEELS - Pre-phosphating of electrogalvanized steel was introduced some years ago in order to improve mainly the forming properties of the material. Initially, two cationic systems have been in use based on Zn-Mn-phosphate and Zn-Ni-phosphate, respectively. Quality constraints led to the development of a trication phosphating system based on Zn-Ni-Mn-phosphate from 1991 onwards. Details of the process and the coatings precipitated are depicted in Table 2. The coatings comply with the quality requirements of most of the automobile manufacturers.

In addition to two-sided electrogalvanized steel, one-sided electrogalvanized, Galvanneal and to some extent HDG are being pre-phosphated. In the case of one-sided EG particular emphasis must be taken to avoid excessive reaction with the steel side, leading to undesired sludge generation. Trication pre-phosphating of EG is being performed in seven lines in Europe. The material is purchased without passivating post-treatment by the OEM industry.

Several high volume cars are currently being made from pre-coated pre-phosphated steel, with a strong upwards trend for the near future. All car bodies made from pre-phosphated sheet are conducted through a chemical pretreatment line comprising degreasing, trication phosphating and post-treatment in order to treat sanded areas as well as areas where

the coating had been damaged, e. g. by forming operations etc.

With state-of-the-art pre-phosphating processes, coating weights on Galvanneal are difficult to control. This causes problems in sheet welding, especially of double layers. This led to the development of no-rinse trication phosphating systems which can be applied by spray/squeeze or roll-on methods. This latest development allows control of the coating weight in a sufficiently low and narrow range. At present, no-rinse pre-phosphated material is being investigated or evaluated at an industrial scale.

IN-LINE PRE-TREATMENT OF ALUMINIUM SHEET - Most of the aluminum sheet material employed to date is purchased by the car manufacturer in a bare surface state and either used as such or given a deoxidising/ passivating treatment prior to stamping and joining. One manufacturer uses chromated, pre-painted (Al pigmented primer, approximately 7 μm) material.

With the increasing use of aluminum, improved properties regarding stamping, welding, adhesive bonding and transport/storage are demanded by the body shops. This initiated extensive collaborative investigations, with various approaches to apply an inorganic or organic coating on the strip (see Table 3). At present, dry lubricants are being investigated or evaluated at an industrial scale.

CHEMICAL PRE-TREATMENT OF MULTI-METAL CAR BODIES

Due to strict environmental and working safety regulations, the European vehicle manufacturers are on the leading edge with respect to the use of environmentally-conscious phosphating products. The technology presently used in Europe is shown schematically in Figure 1. The various pre-treatment stages will be discussed in more detail, especially with respect to environmentally-conscious products. Depending on regulations in Europe, the priorities may be different for individual countries.

DEGREASING - Thorough degreasing is an important precondition to achieve a high quality conversion coating. The alkaline cleaners used are formulated with biodegradable and APEO (= Alkylphenolethoxylate) free surfactants. Powder cleaners and liquid cleaners are used in a ratio of approximately two to one. A considerable amount of cleaners are formulated as two component products in order to cope with the different permeability of builder and surfactants when passed through ultrafiltration devices.

ACTIVATION - Grain refining (activation) with colloidal titanium phosphate polymers is state-of-the-art world-wide. Although it is generally accepted that the grain refiners give better performance in a separate stage than in a cleaner stage, many spray and also spray/dip lines are not equipped with separate activation stages.

Most activating products still in use have very limited bath-lifetimes necessitating frequent bath make-up and replenishment. Some five years ago grain refiners with a greatly extended bath-lifetime were introduced. Currently some twenty car body lines and numerous parts lines are operated with these materials. Through their use a fair amount of process water can be saved.

Characteristic data of the liquid grain refiners are given in Table 4. An interesting feature is that they are supplied in liquid form and can be dosed directly by means of metering pumps, eliminating manual additions of powder.

NITRITE-FREE PHOSPHATING - In order to achieve fine-crystalline zinc phosphate coatings in a reasonably short time, oxidising reagents must be added to accelerate the phosphating reaction.

Two basic types of nitrite-free processes have been introduced in recent years. The accelerators used in these processes are hydrogen peroxide and hydroxylamine compounds, respectively. The formation of toxic fumes (NOx) in the plant and the contamination of effluents with NO_2^- are avoided by both of them. Peroxide-accelerated processes may be more sound due to the lack of accelerator breakdown products which could enrich in the phosphating bath and hence contribute to the rinsing water contamination. Hydroxylamine-accelerated processes may be easier to operate due to the integrated accelerator but form breakdown products such as ammonium ions.

Peroxide-accelerated processes tend to be more advantageous in spray application whereas hydroxylamine-accelerated processes give better results in immersion applications. Characteristics of interest are summarized in Table 5.

NICKEL-FREE PHOSPHATING - Trication phosphating processes contain heavy metal ions like zinc, nickel and manganese. The concentrations of heavy metals in the waste water are limited in many countries by state and also local regulations. In Germany, for example, the maximum permitted waste water concentration of nickel is 0,2 to 0,5 ppm. It is common knowledge that such low limits are difficult to reach even when applying special waste water treatment. These problems may be solved by installing a selective ion exchanger or by employing a nickel-free process. Nickel-free processes, as referenced in this paper, are also cobalt-free.

In recent years a nickel-free process has been developed using traces of Cu ions as a nickel replacement. For steel and aluminum treatment the process may be accelerated by sodium nitrite or hydroxylamine. Hydrogen peroxide seems to be less appropriate. If pre-coated steel should also be treated, hydroxylamine is most appropriate as an accelerator. This is due to the ability to control the nitrate concentration which turned out to be a critical factor with respect to the formation of white spots.

This process has been successfully employed in two vertically indexed immersion lines for car

bodies since early 1996. Materials processed are CRS, EG two-sided, EG one-sided, Al and Al pre-primed. Data of interest are compiled in Table 6.

CHROME-FREE PASSIVATION - Passivating post-treatments are still being used by most of the vehicle manufacturers throughout North America and Europe. Japanese and Korean manufacturers normally do not apply post-treatments.

Most frequently, products containing hexavalent/trivalent chromium are still being employed. Since hexavalent chromium compounds are highly toxic and carcinogenic, the maximum permitted concentrations are very low (i.e. 0,1 mg/m^3 in air and 0,1 mg/l in effluents in Europe).

Chrome-free alternatives on the basis of flourozirconates are generally accepted by the European automotive manufactureres to give equivalent paint adherence and creep resistance as chromium based products. Approximately twelve car body lines currently employ these products. The majority of these lines are located in Germany.

Unlike chrome based products, it is possible to establish either a small or large rinse cascade when using the chrome-free post treatment. In both cases considerable amounts of water can be saved (see Figure 2). Since the chrome-free process does not provide substantial protection to bare phosphated steel panels it is recommended to electrocoat the bodies wet-on-wet after chemical conversion treatment.

More recently, a trend towards the elimination of post treatments has also been observed in Europe. At present, too little data are available to discuss the long term impact of this change.

PHOSPHATING OF ALUMINIUM - Trication phosphating with controlled addition of fluoride is the current state-of-the-art pre-treatment of multi-metal car bodies with aluminum hang-on parts or all-aluminium car bodies in Europe. Approximately ten lines are currently employing this technique, with spray and dip application balancing each other. All processes do contain nickel, and the majority are accelerated by sodium nitrite. Some of the practical experience collected over the recent years is depicted in Table 7.

Spray phosphating of approximately 80 % aluminum together with steel and pre-coated steel has been operated successfully for many years. In immersion applications some preconditions have to be met in order to achieve satisfactory coating results. These are, among others, the use of intensive bath agitation and the minimization of bath contaminants. To date, less than 20 % of aluminum are being processed by immersion application. In Europe there do exist very ambitious plans to increase the amount of aluminium to app. 35 % for certain models, in some cases to 50...100 %.

CATHODIC ELECTROCOATING

With respect to electrocoating in Europe, lead-free cathodic electrocoats have been introduced in several car body lines most recently. Some of the materials in use are also tin-free. The solvent content has been decreased to under approximately two percent in most applications.

A further trend is the conversion of the electrocoat tanks to the so-called generation five materials which are said to give higher throw power at lower coating thickness. Also so-called low-density materials are being used to some extent. Most of these materials still contain lead.

In several lines the effluents from the electrocoating section have been decreased to zero due to the establishment of rinsing cascades and the employment of a final rinse with ultrafiltrate.

Low bake materials are not a topic in Europe because of the intensive use of bake-hardening steel or aluminum sheet. Bake-hardening materials need oven temperatures exceeding 180°C in order to obtain their final mechanical strengths.

SUMMARY

Trends in Europe with respect to corrosion protection practices are the increasing use of coated steel, pre-phosphated steel and aluminum. Several projects are under way to use these materials with inorganic or organic functional layers. The use of light materials like magnesium and plastic are intensively investigated. Due to environmental issues, new pretreatment products are used including cleaners with biodegradable and APEO-free surfactants, liquid activators with longer bath-lifetime, nitrite-free and/or nickel-free zinc phosphating agents and chrome-free post-treatments. Lead- and tin-free cathodic electrocoats have been introduced in several car body lines. A further trend in electrocoating is the conversion of the state-of-the -art materials to so-called generation five and low-density products.

TABLE 1: Car body construction materials employed in Europe

Cold-rolled steel	- soft, unalloyed - high strength,e.g. dual-phase, triple-phase - bake-hardening	3 2 3
Hot-dip galvanized steel	- Zn + 0.1- 0.2 % Al - Zn + ca. 10 % Fe (Galvanneal) - Zn + ca. 5 % Al + 1 % Rare Earths(Galfan) - Zn + ca. 55 % Al + 1.5 % Si (e.g.Galvalume)	3 2 0 0
Electrogalvanized steel	- Zn - Zn + ca. 10 % Ni - Zn + ca. 16 % Fe - Zn + ca. 16 % Fe / Zn + ca. 83 %Fe	3 3 0 0
Hot-dip/Electrogalvanized steel	- Zn + ca. 16 % Fe / Zn + ca. 83 % Fe	0
Pre-coated steel, pre-primed	- Zn or ZnNi (el.) + zinc rich paint (3 - 7 μm) - ZnNi (el.) + org. coating (~1 μm)	2 1
Aluminium	- AlCuMg (AA 2xxx) - AlMg, AlMgMn (AA 5xxx) - AlMgSi (AA 6xxx)	1 1 2
Aluminium, coated with primer	- AlMgSi (AA 6xxx) + org. coating paint (~7 μm)	1
Magnesium	- AZ 91 HP	0
Plastic	- thermoplastics	1

0 = not used/investigated   1 = limited application   2 = medium amounts   3 = extensive use

TABLE 2: Trication pre-phosphating process for EG strip

Treatment times:	5 - 15 sec
Treatment temperature:	60 - 62 °C (140 - 143 °F)
Coating weight:	1.6 ± 0.4 g/m^2 (two-sided EG) 1.3 ± 0.4 g/m^2 (one-sided EG)
Coating composition:	38 - 40 % Zn 30 - 33 % P_2O_5 ~ 1.0 % Ni 4 - 8 % Mn
Lines in Europe:	7

TABLE 3: Aluminum for car body construction

Surface state	**Status**
bare	in use
bare, deoxidized + passivated	in use
chromated, rinse or dry-in-place *	in use
chromated + conductive paint (~ 7 μm) *	in use
chromated + conductive paint (~ 2 μm) *	R&D
pre-treatment chrome-free, rinse or dry-in-place *	R&D
pre-treated chrome-free, with conductive paints *	R&D
anodized *	R&D
zinc phosphated *	R&D
thin Zn layer (plasma, electrolytic, zincate) *	R&D
thin Zn layer (plasma, electrolytic, zincate) + Znph *	R&D
dry lubricant *	R&D
pre-treatment chrome-free + dry lubricant *	in use
	R&D

* These pre-treatments are applied in-line onto the strip

TABLE 4: Properties of liquid activators

Supply product	• liquid • milky white • shelf life-time 6 months • Storage temperature > -7 °C (20 °F) • automatically dosable
Activation bath	• stable up to 50 °C (120 °F) • bath life-time ~ 8 weeks • constant activation • completely dispersible • only compensation of carry-over losses • consumption 50 - 200 mg/ft^2

TABLE 5: Characteristics of nitrite-free phosphating processes

Peroxide-accelerated process	• no breakdown products • more appropriate in spray application • minimum time 1,5 min for steel • separate neutralisation necessary • frequent accelerator determination • closed-loop possible
Hydroxylamine-accelerated process	• more appropriate in immersion • breakdown product are ammonium ions • accelerator contained in replenisher • rare accelerator determination • process slower than nitrite-accelerated • operates on FeII side

TABLE 6: Nickel-free phosphating process - line data

Type of line	- 2 indexed vertical immersion lines
Throughput	- approximately 700 units/day per line
Material processed	- CRS, EG, Al, Al pre-primed
Number of stages	- 11
Nickel replacement	- traces of Cu
Accelerator	- hydroxylamine
Treatment time	- 180 secs
Fell limitation	- with pressurized air
Coating weight on CRS	- 2.5 g/m^2
Crystal size on CRS	- 4 - 10 microns
Post-treatment	- fluorozirconate
Performance with CED	- similar to ZnNiMnph process

TABLE 7: Practical experience with aluminum phosphating

- Applicable by spray and immersion
- Spray process easier to operate than dip process
- Percentage of aluminum being processed to date
  - spray: < 80 %
  - dip: < 20 %
- Treatment of 35 - 50 % Al by dip expected for 1998
- Preconditions in dip application
  - high bath agitation
  - good activating pre-rinse
  - minimize bath contamination with phosphating poisons
- fluoride control with ion-sensitive electrode

FIGURE 1: Pre-treatment technology used in European OEM industry

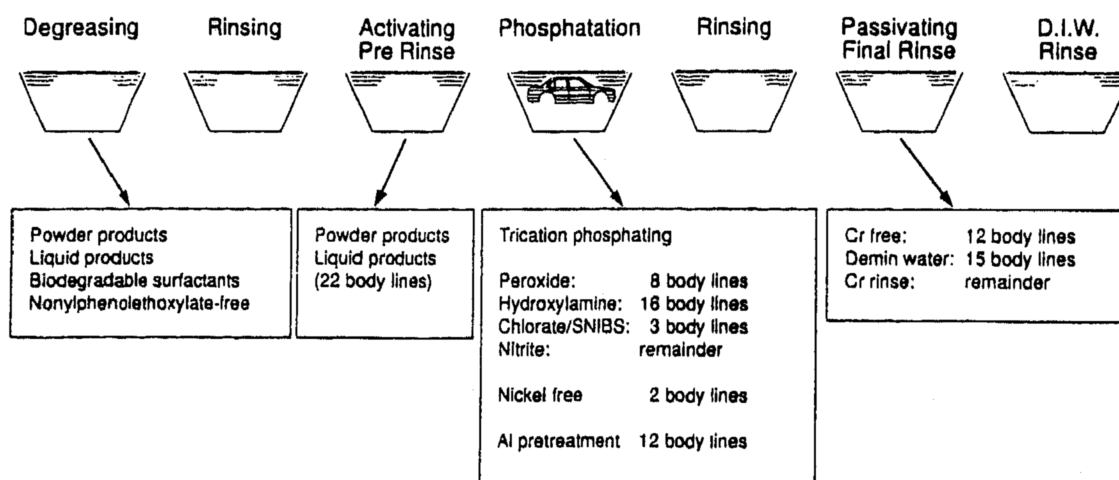
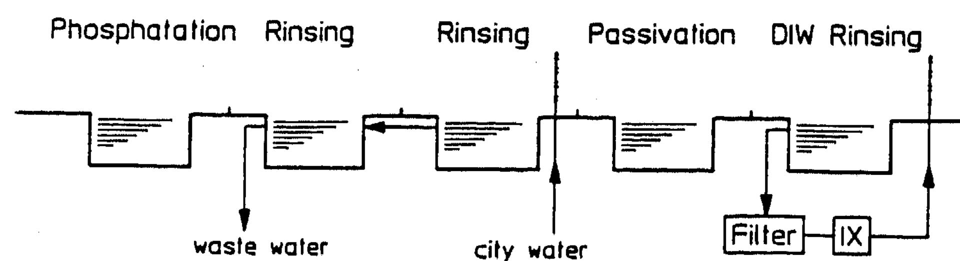


FIGURE 2: Recommended cascades for Cr-free post-treatment

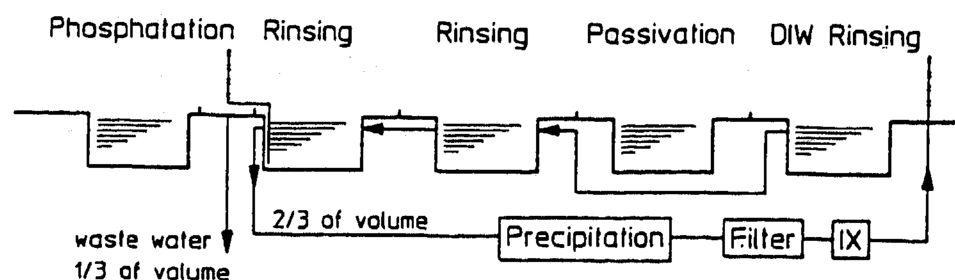
Rinse Water Recycling after Phosphating

Passivation with Bonder-Nachspüler D 6800 (chrome-free)

Small cycle



Large cycle



970734

Round-Robin Evaluation of a New Standard Laboratory Test for Cosmetic Corrosion

Herbert E. Townsend
Bethlehem Steel Corp.

Duncan C. McCune
Quality/Statistics

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A new, standard laboratory test (SAE J 2334) for evaluation of the cosmetic corrosion resistance of autobody steel sheet has been developed through the joint efforts of SAE's Automotive Corrosion Prevention Committee (SAE/ACAP) and the Auto/Steel Partnership's (A/SP) Corrosion Task Force. Results from this test give an excellent correlation with those of on-vehicle tests conducted in Montreal, Quebec, and St. John's, Newfoundland. To establish the precision of the new test, nine laboratories participated in a round-robin evaluation. Each laboratory tested ten standard, painted automotive steel sheet products. The results of this work are analyzed in terms of within-lab repeatability and lab-to-lab reproducibility, and they are an important and final milestone in the development and specification of the new test.

INTRODUCTION

In the early 1980's, the automotive and steel industries recognized that an improved laboratory test for evaluation of the cosmetic corrosion resistance of coated steel sheet products was sorely needed. The widely used salt-spray test was notoriously unreliable. Although existing cyclic tests were better than salt spray, they did not completely agree with real-world performance. Moreover, the large number of different types of cyclic tests made it difficult for auto manufacturers and their suppliers to agree on performance criteria.

To address this need, the Automotive Applications Committee of the American Iron and Steel Institute established the Corrosion Task Force in 1984 with the mission of developing an improved, standard laboratory test. This Group joined forces with the SAE/ACAP Cosmetic Corrosion Subcommittee and the two groups proceeded with the test development. The Auto/Steel Partnership added its support to the groups in 1995.

The 12-year efforts of these groups have culminated in the establishment of a new corrosion test which has been tentatively designated as SAE J 2334. Development of this test involved a number of key steps. As documented in a series of papers published during the course of this development [1-10], these steps included:

o A reservoir of standard test materials was established. These included cold rolled and coated sheet products representing a broad range of performance. The materials were phosphated and painted to give ten material variables. This reservoir made possible the evaluation of a broad range of tests with a common set of materials, thus giving a well-defined and consistent frame of reference for comparing test results.

o To establish the real-world performance of the reference materials, on-vehicle tests were conducted for a 5-year period on pickup trucks operated in Montreal, Quebec, and St. John's, Newfoundland. The results of these tests served as the real-world standard against which the results of all other tests were compared.

o License-plate test racks containing a subset of the standard materials were distributed to committee members. In 1992, fifty racks were prepared for mounting on the front license-plate brackets of cars driven in many locations throughout the USA, including Florida, Pennsylvania, and Michigan. Results from these tests will ensure that the real-world standard of results established in Montreal and St. John's is truly representative of a wide range of environments.

o The standard materials were tested in a wide variety of test environments by the committee members and other cooperating laboratories worldwide . These included 19 laboratory tests, three proving ground tests, and static exposure with twice-weekly application of salt solution at seven outdoor

sites.

o Methods were developed for comparing the results of the various tests to those of the real-world standard. These comparisons clearly showed that none of the existing test methods evaluated was completely satisfactory, see Table 1. In particular, all of the evaluated tests underestimated the performance of zinc and zinc-alloy coated sheet relative to cold rolled.

Table 1: Rating of various tests.

Test	Rating*
J 2334	97
Best existing cyclic test	80
Best outdoor test	56
Best proving ground test	52
Salt spray test	12

* Rating is based on the the R^2 + C criterion given in Reference 1, normalized to a scale of 0 (worst) to 100 (perfect correlation).

o To overcome the shortcomings of existing tests, we conducted a statistically designed experiment (Plackett-Burmann L8 matrix)[1,2]. The effects of seven variables in a typical cyclic test were evaluated. The results of these tests indicated that two of the variables are critical to obtaining a good correlation with real-world behavior. First, it was discovered that the salt solution used in most tests (typically 5%) was too concentrated. Much better results are obtained with lower concentrations, such as 0.5 to 0.85%. Second, it was found that best results are obtained when the relative humidity in the dry stage is controlled at 50%.

o Based on the above findings, the new J 2334 test cycle (see Figure 1) was designed and evaluated. As shown in Table 1, the new test shows an excellent correlation with the on-vehicle tests conducted at St. John's and Montreal, and it is a major improvement over any of the existing tests evaluated.

o To determine whether the mechanisms of corrosion occurring in the new test environment are similar to those taking place in the on-vehicle tests, Lehigh University conducted a detailed study of the test specimens. This study showed an excellent match of the corrosion products and morphologies of attack for J 2334 and the real-world[11].

o Finally, to determine the precision of the new test, a round-robin, or interlaboratory, test program was conducted in which nine laboratories conducted the J 2334 test on the standard set of ten materials. This paper describes the results of the round-robin tests.

EXPERIMENTAL PROCEDURES

TEST MATERIALS - A variety of automotive steel sheet products were sheared into 100- x 300-mm panels, and coated with a full automotive paint system, see Table 2. These materials, including bare, cold rolled as well as metallic-coated sheet, were selected to give a wide range in performance, and to be representative of the types of products in use for the construction of automobiles. Specifically, they included steel sheet with: two coating weights each of electroplated zinc, electroplated zinc-nickel alloy, and hot-dip zinc-iron alloy, plus one coating weight of hot-dip zinc. Also included are three uncoated cold rolled steels, two with different phosphate pretreatments, and one with no phosphate. These are the same ten material combinations that were used in on-vehicle and laboratory tests throughout the development of J 2334.

Also included in the tests were duplicate, preweighed 25- x 50-mm coupons of bare cold rolled steel intended to provide quantitative mass-loss data on the corrosivity of the test conditions.

TEST SPECIMENS - Painted test panels were cut in half to form 100- x 150-mm coupons. Metallic adhesive tape was applied to the single cut edge to protect it from corrosion. A diagonal scribe was made with a carbide lathe tool to scratch through the paint to expose bare, uncoated steel. All of the test panels were prepared and scribed at ACT laboratories. It was intended that the test panels would be oriented at an angle of 15 degrees to the vertical.

TEST ENVIRONMENT - The J 2334 test cycle is shown schematically in Figure 1. Based on the results of the statistically designed experiment, the key factors which make this test correlate successfully with on-vehicle tests are the use of a dilute salt solution during the immersion stage, and a controlled 50% relative humidity during the dry stage. Test coupons were exposed to this environment for a target duration of 80 cycles. Total test time to complete 80 cycles including weekends is 110 days, plus any holidays. Based on the 10.4 acceleration factor found for this test [1], the 80 cycles corresponds to roughly 3.4 years in an severe environment equivalent to an average of those at St. John's and Montreal.

The test cycle shown in Figure 1 is shown as it would run manually, that is, by manually moving the test specimens from a dry-stage chamber to a humid-stage chamber, then to an immersion tank, and finally back to the dry-stage chamber on a daily basis. The manual cycle can also be run by use of a single chamber which is switched back and forth between dry and humid conditions. Three of the participating laboratories ran the manual test.

An alternate method is to run the cycle as a fully automated, continuous operation by use of a chamber which can be programmed to create all three stages. When this is the case, the dry-stage hold on weekends and holidays is omitted. This reduces the test time for 80 cycles to 80 days. Four laboratories in the round-robin series operated the test in this fashion. Another two laboratories ran an automated test that included the weekend hold.

ROUND-ROBIN PARTICIPANTS - Nine laboratories have thus far participated in the round-robin study, see Table 3. Additional tests may be completed at a later date. All participants received identical sets of triplicate painted and scribed panels, as well as four pairs of bare, cold rolled steel mass-loss coupons from ACT. The tests were intended to be conducted as nearly as possible according to the the cycle shown in Figure 1. Conditions were monitored and recorded throughout the tests. Significant deviations from the procedure reported by participants are noted in Table 3.

EVALUATION - After 80 cycles, the coupons were removed from test and returned to ACT for final evaluation. Scribe creep, the delamination of paint next to the scribe, was evaluated by scraping and taping to remove any loose paint. Total width of paint delamination was measured at 10, 1-cm intervals along the scribe and averaged. This was done for each of the triplicate coupons.

RESULTS

SCRIBE CREEP - Results of the round-robin tests are shown on Figures 2 to 11 for each of the ten standard materials. All laboratories obtained results that generally give similar rankings for the performance of the ten standard materials, as follows (in order of decreasing performance):

thick coated > intermediate-thickness coated > thin coated > phosphated cold rolled > unphosphated cold rolled

This ranking is consistent with that of the on-vehicle tests.

MASS-LOSS COUPON RESULTS are shown in Figure 12. In general, mass-loss results fall in an order generally similar to that of the phosphated cold rolled steels, cf. Figures 9 and 11. An exception is noted in the case of Laboratory 9 which showed relatively low mass-loss values. This was found to be the result of improper mounting of the coupons -- spacers which should have been used to separate the coupons from the rack were omitted -- this caused one side of the coupon to be shielded from the environment by the rack.

PRECISION - The precision of a test method can be considered in terms of two components: (1) repeatability, the ability to replicate a result within a single laboratory; and (2) reproducibility, the ability to replicate a result from one laboratory to another.

Following 80 cycles of testing, and removal of loose paint, the ten measurements of scribe creep on each test panel were averaged to give a single panel value, x. Following the procedures and nomenclature outlined in References 12 and 13, the three values for each panel of a particular material tested in a given laboratory were used to calculate a cell average, x', and a cell standard deviation, s.

Repeatability - The repeatability standard deviation for the material, s_r, is defined as the the root-mean-square average of the cell standard deviations for all laboratories, that is,

$$s_r = \left(\sum s^2/p\right)^{1/2}$$

where p is this number of laboratories.

The repeatability interval, r, is defined as the 95% confidence interval for the difference between the scribe-creep values between individual panels of a particular material tested in the same laboratory, and is given by,

$$r = 2.8\ s_r$$

Values of r for each of the materials are plotted in Figure 13 as a function of the material's average scribe creep (that is, as a function of x" as defined below). Note that average scribe creep varies from roughly 2 mm for coated steels to 24 mm for unphosphated cold rolled steel.

A least squares line is drawn through the data on Figure 13. It shows that, on average, the repeatability interval is about 25% of the creep value.

Reproducibility - For the reproducibility of a

particular material, we calculated x" which is an average of the cell averages (x') from all laboratories, and a corresponding standard deviation, $s_{x''}$.

The reproducibility standard deviation, s_R, is defined as the larger of either $s_{x''}$, or the term,

$$\left((s_{x'})^2 + (s_r)^2(n-1)/n\right)^{1/2}$$

Finally, the reproducibility interval, R, is defined as the 95% confidence interval for the difference between the scribe-creep values for individual panels of a particular material as measured in different laboratories It is calculated as follows.

$$R = 2.8\ s_R$$

Values of R for the ten materials are plotted in Figure 14, along with a least-squares line. These data indicate that, on average, the reproducibility interval is about 144% of the measured value.

We note that the lab-to-lab reproducibility intervals, R, are significantly greater than the within-lab repeatability intervals, r. The relatively large reproducibility intervals probably result from deviations to the standard test procedures by the participating laboratories. To examine this hypothesis, we normalized the results in terms of the ratios of scribe creep of cold rolled steel to that of electrogalvanized sheet, as shown in Figure 15. This ratio is of particular significance because one of the salient strengths of the J 2334 test is that it yields a large value comparable to that of on-vehicle tests. Similar conclusions are also reached from looking at the absolute values. Also shown on Figure 15 are the deviations from the normal test procedures as given in Table 3.

Effect of specimen orientation - Laboratory 3, the only laboratory with a vertical specimen orientation gave the highest ratio plotted on Figure 15. Conversely, Laboratory 4, the only laboratory with a 30-degree orientation produced a low ratio relative to the laboratories without deviation (Laboratories 2 and 5). This result concurs with the those of on-vehicle tests conducted in Montreal and St. John's where it was found that this ratio was higher for vertical panels than for horizontal panels[9]. Both results indicate that the specimen orientation is an important factor to control for improving the reproducibility of this test.

Effect of dry-stage relative humidity - Laboratories 6 and 8 gave the lowest value of the ratio plotted in Figure 15. Previous work in the development of this test demonstrated conclusively that 50% relative humidity in the dry stage is one of the two most critical factors in achieving adequately high creep rates for cold rolled steel, thus giving a good correlation with the results of on-vehicle tests[1,2]. Moreover, the Lehigh University mechanism study showed that the 50% dry-stage humidity gives the best match of corrosion products and morphology[11]. The relatively low creep rates for cold rolled steel found by Laboratories 6 and 8 in this study provide further confirmation of the need to control dry-stage humidity at 50%.

Effect of salt application method - J 2334 was initially developed as a manual test. As shown in Figure 1, application of the salt solution in the manual test is easily and reproducibly accomplished by a simple 15-minute immersion. For all but one of the automated tests, other methods of salt application were employed. As shown in Figure 15, the alternate methods generally are associated with a lower ratio. The results suggest that the effect of application method is to decrease ratio as follows:

immersion > heavy spray > fog > light spray.

If so, use of immersion by all laboratories should improve the reproducibility of the test. Alternatively, it would be necessary to define and control the conditions of spray or fog in a way that can be easily replicated, and can be shown to be equivalent to immersion. Given the number of variables needed to define spray and fog conditions, such as spray volume, nozzle configuration, spray velocity, angle of impingement, droplet size, and collection rate, specification of reproducible spray and fog parameters may be difficult.

It is also noted that the spacing of test panels becomes more important when salt is applied by spray or fog because of the need to avoid sheltering. As an example, we note that Laboratory 9 provided a 1.2-cm spacing between panels. Since the panels were inclined at a 15-degree angle from the vertical, there could easily have been sheltering from the fog by adjacent, overhanging panels.

Other considerations - Other, yet unidentified, factors may be contributing to the reproducibility interval. Although there were no apparent effects of 80 vs. 110 days of testing to reach 80 cycles, any possible effect of this variable could be simply avoided by defining the total test time in an automated test.

CONCLUSIONS

As a final step in the development of SAE J 2334, the new laboratory test for determining the relative corrosion resistance of coated automotive steel sheet products, a round-robin study was conducted to

determine the precision of the test. Nine laboratories tested ten material combinations. Results of these tests lead to the following conclusions:

o Participants obtained a similar ranking of materials that is generally in agreement with those of on-vehicle tests.

o In terms of a percentage of the creep measured for a particular material, the within-lab repeatability interval was about 25%, and the between-lab reproducibility interval was 144%.

o It should be possible to improve the reproducibility by specifying and adhering to precisely defined conditions, particularly for conditions of specimen orientation, dry-stage humidity, and method of salt application.

ACKNOWLEDGEMENTS

The authors are grateful for the many contributions by members of the SAE/ACAP Cosmetic Corrosion Subcommittee, and the Auto/Steel Partnership Corrosion Task Force during the development of J 2334. Financial support for this work was provided by the Automotive Applications Committee of the American Iron and Steel Institute, and by the Auto/Steel Partnership. Most of all, we are pleased to acknowledge the participants in the round-robin study: Frank Lutze, ACT Laboratories, Inc.; Paul Ziman, Alcoa; George Donchez, Bethlehem Steel Corporation; Pankaj Shah, Ford Motor Company; Mike Ostermiller and Larry Thompson, General Motors; Steve Grossman, Q-Panel Lab Products; Mike Simko, USS Corporation; Kevin Smith, Atotech, USA, Inc.; and Cindy Meade, Singleton Corporation.

REFERENCES

1. Townsend, H. E., "Development of an Improved Laboratory Corrosion Test by the Automotive and Steel Industries," in Advanced Coatings Technology, Proceedings of the fourth Annual ESD Advanced Coatings Conference, The Engineering Society, Ann Arbor, Mich., 1994, pp. 29-49.

2. Roudabush, L. A., Townsend, H. E., and McCune, D. C., "Update on the Development of an Improved Cosmetic Corrosion Test by the Automotive and Steel Industries," Automotive Corrosion and Prevention Conference Proceedings, P-268, Society of Automotive Engineers, Warrendale, Pa., 1993, pp 53-63.

3. Townsend, H. E., "Accelerated Corrosion Testing: A Cooperative Effort by the Automotive and Steel Industries," Proceedings of the Symposium on Corrosion-Resistant Automotive Sheet Steels, ASM Materials Congress, ASM International, Metals Park, Ohio, 1988, pp 55-67.

4. Townsend, H. E., "Status of a Cooperative Effort by the Automotive and Steel Industries to Develop a Standard Accelerated Corrosion Test," Automotive Corrosion and Prevention Conference Proceedings, P-228, Society of Automotive Engineers, Warrendale, Pa., 1989, pp 133-145.

5. Townsend, H. E., Granata, R. D., McCune, D. C., Schumacher, W. A., and Neville, R. J., "Progress by the Automotive and Steel Industries Toward an Improved Laboratory Cosmetic Corrosion Test," Automotive Corrosion and Prevention Conference Proceedings, P-250, Society of Automotive Engineers, Warrendale, Pa., 1991, pp 73-97.

6. Stephens, M. L., "SAE ACAP Division 3 Project: Evaluation of Corrosion Test Methods," Automotive Corrosion and Prevention Conference Proceedings, P-228, Society of Automotive Engineers, Warrendale, Pa., 1989, pp 157-164.

7. Lutze, F. W., and Shaffer, R. J., "Accelerated Atmospheric Corrosion Testing of AISI Panels," Automotive Corrosion and Prevention Conference Proceedings, P-250, Society of Automotive Engineers, Warrendale, Pa., 1991, pp 115-127.

8. Petschel, M.,"Statistical Evaluation of Accelerated Corrosion Tests and Correlation with Two-Year On-Vehicle Tests," Automotive Corrosion and Prevention Conference Proceedings, P-250, Society of Automotive Engineers, Warrendale, Pa., 1991, pp 179-203.

9. Davidson, D. D. and Schumacher, W. A.,"An Evaluation and Analysis of Commonly Used Accelerated Cosmetic Corrosion Tests Using Direct Comparisons with Actual Field Exposure," Automotive Corrosion and Prevention Conference Proceedings, P-250, Society of Automotive Engineers, Warrendale, Pa., 1991 pp 205-219.

10. Ostermiller, M. R., and Townsend, H. E., "On-Vehicle Cosmetic Corrosion Testing of Coated and Cold-Rolled Steel Sheet," Automotive Corrosion and Prevention Conference Proceedings, P-268, Society of Automotive Engineers, Warrendale, Pa., 1993, pp 65-83.

11. Granata, R. D. and Moussavi-Madani, M., "Characterization of Corrosion Products and Corrosion Mechanisms on Automotive Coated Steels Subjected to Field and Laboratory Exposure Tests," Lehigh University Report to the AISI Corrosion Task Force, January 10, 1996; also presented at this

symposium.

12. ASTM E 691-92, "Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method," Annual Book of Standards, American Society for Testing and Materials, West Conshohocken, Pa.

13. ASTM E 177-90a, "Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods," Annual Book of Standards, American Society for Testing and Materials, West Conshohocken, Pa.

Table 2: Standard steel sheet test materials.

Coating Metal	Coating Process	Coating Mass, g/m^2	Phosphate Type
Zinc	Electroplated	30	B
Zinc	Electroplated	70	B
Zinc-iron alloy	Hot Dip	5 to 9	B
Zinc-iron alloy	Hot Dip	67	B
Zinc-nickel alloy	Electroplated	18	B
Zinc-nickel alloy	Electroplated	44	B
Zinc	Hot Dip	108	A
Uncoated	-	-	B
Uncoated	-	-	None
Uncoated	-	-	A

All materials were spray phosphated (except as noted), primed with cationic electrocoat, and sprayed with white base coat and clearcoat layers

Table 3: Laboratories and test conditions .

Organization	Type of Cycle	Time, days	Dry-Stage % RH	Salt Application Method	Panel Orientation(a)	Panel Spacing, cm
Atotech	Automatic	80	50	Heavy Spray	15	2.5
ACT Laboratories, Inc.	Manual	110	50	Immersion	15	2.5
Ford Motor Company	Automatic	110	50	Immersion	0	2.5
USX Corporation	Automatic	80	50	Heavy Spray	30	2.5
Bethlehem Steel Corporation	Manual	110	50	Immersion	15	2.5
Q-Panel Lab Products	Automatic	80	8	Fog	15	2.5
Alcoa	Automatic	110	50	Immersion	0	2.5
General Motors	Manual	110	<20	Light Spray	15	2.5
Singleton	Automatic	80	50	Fog	15	1.3

a - Degrees inclination of panel relative to the vertical.

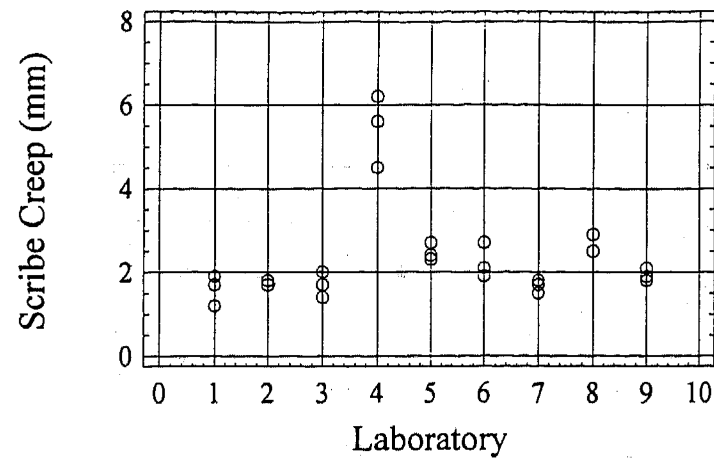


Figure 2: Scribe creep for 30 g/m^2 electroplated zinc.

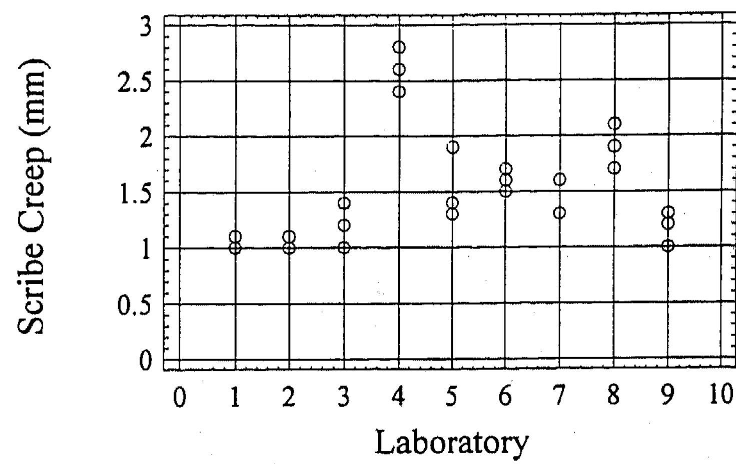


Figure 3: Scribe creep for 70 g/m^2 electroplated zinc.

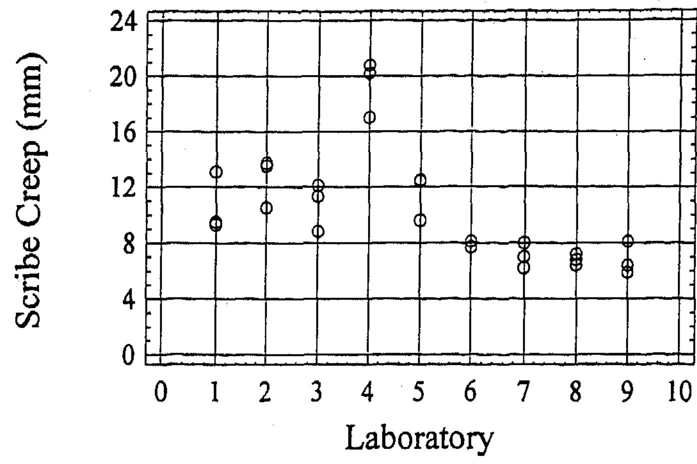


Figure 4: Scribe creep for 5-9 g/m^2 hot-dip zinc-iron.

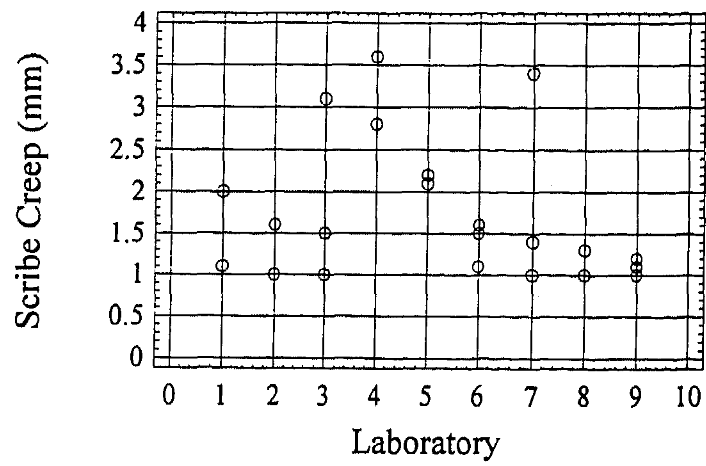


Figure 5: Scribe creep for 67 g/m^2 hot-dip zinc-iron.

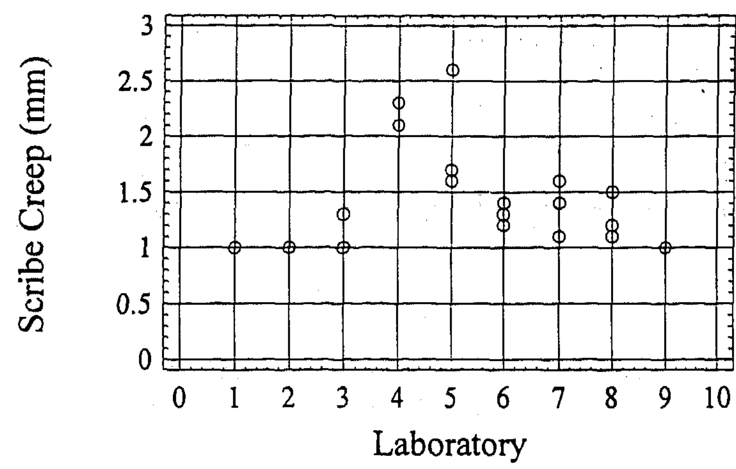


Figure 8: Scribe creep for 108 g/m^2 hot-dip zinc.

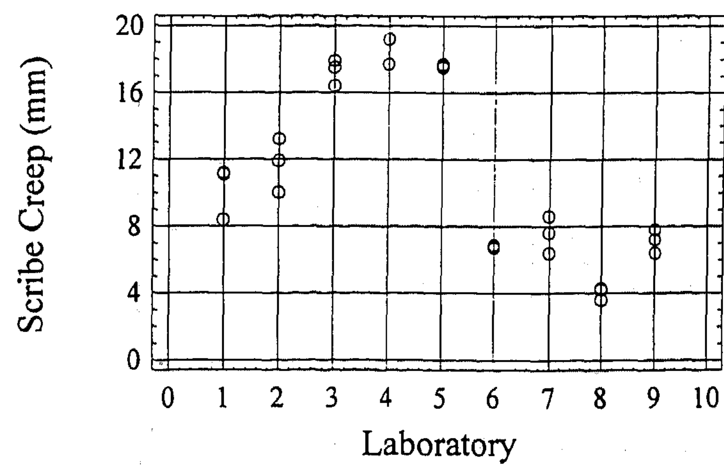


Figure 9: Scribe creep for cold rolled steel with phosphate B.

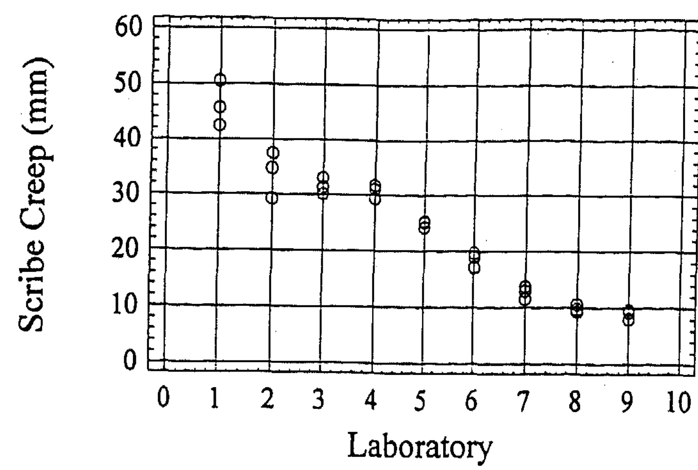


Figure 10: Scribe creep for cold rolled steel with no phosphate.

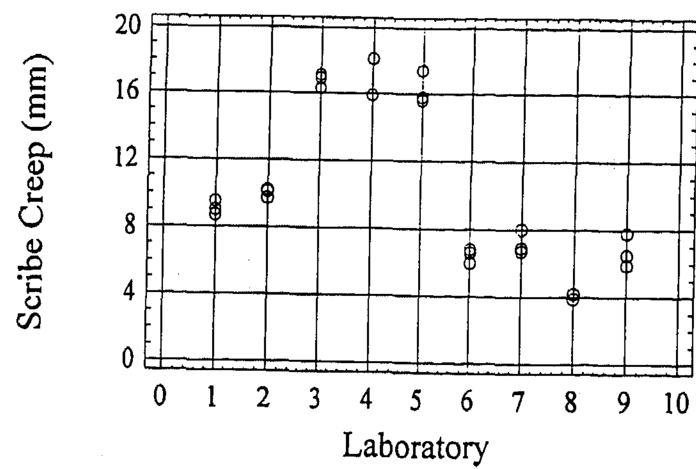


Figure 11: Scribe creep for cold rolled steel with phosphate A.

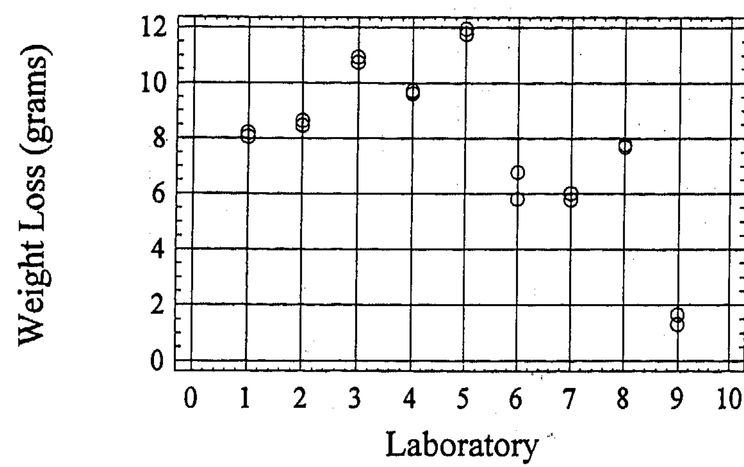


Figure 12: Mass loss of cold rolled steel coupons.

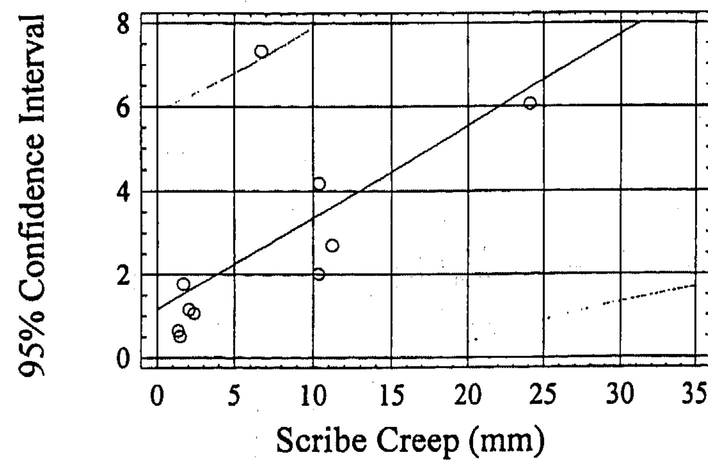


Figure 13: Repeatability (r) intervals for scribe creep .

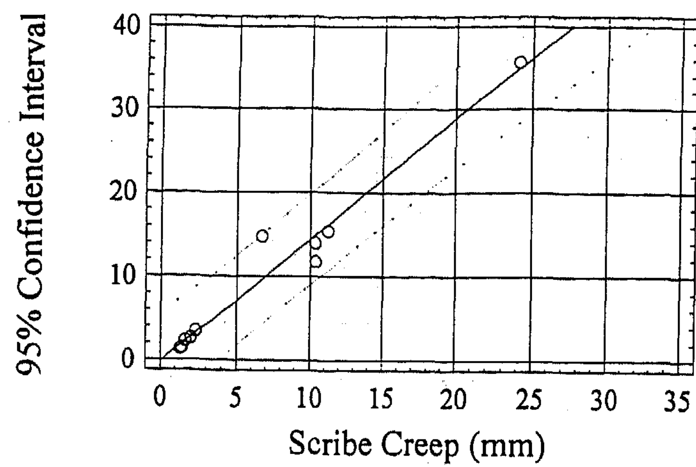


Figure 14: Reproducibility (R) intervals for scribe creep .

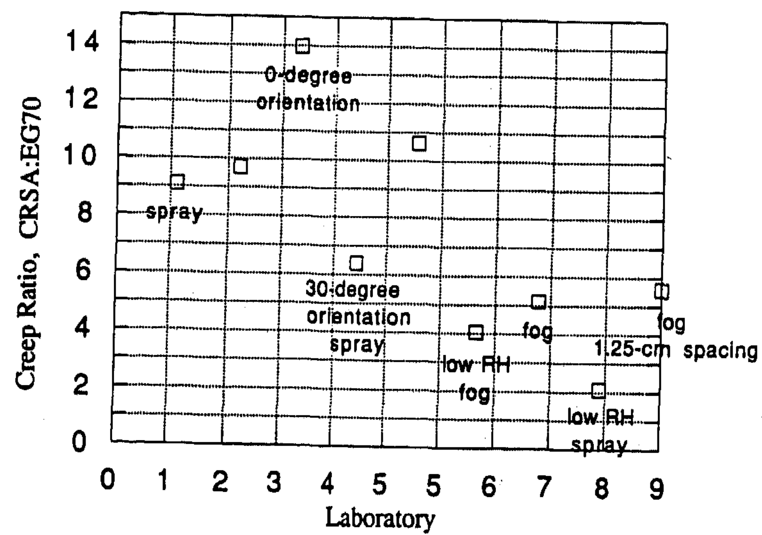


Figure 15: Effects of deviations from standard test conditions on the ratio of scribe creep of cold rolled steel (phosphate A) to 70 g/m2 electrogalvanized.

970762

Methods for Analyzing the Value of Automobiles

J. A. Donndelinger and H. E. Cook
University of Illinois at Urbana-Champaign

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Engineers, marketing specialists, and planners responsible for making the numerous cost/benefit trade-offs that arise in the development of a new product need to understand how the attributes of the product generate value to the customer. A recently proposed methodology for benchmarking product value is tested here using five family automobiles competing in the middle product segment. Reasonable agreement is found between value differences estimated from an analysis of demands and prices and those estimated from attribute differences. This finding supports the use of the methodology in forecasting the demands and profits of future products based upon the improvements expected in their attributes.

INTRODUCTION

Persons developing the design and manufacturing processes for a new product need to understand how the attributes of the product affect its value to the customer. For example, marketing specialists and product planners, working at the system level, make important value assessments for new features at the start of new product programs. As the details of the product design and its manufacturing processes unfold, engineers at the subsystem and component levels make numerous cost versus perceived value trade-off decisions. Because the customer is the ultimate judge of the wisdom of these decisions, the challenge shared by all stakeholders in the development process is to reliably project what the customer's reaction will be to an attribute change and how that reaction will affect the profitability of the product.

For example, suppose an engineer has discovered a design modification requiring a variable cost increase of $100 but no added investment which, if implemented, would improve the fuel economy for a new vehicle under development from its currently planned level of 21 mpg to 23 mpg. With the price of fuel at $1.20 per gallon, the economic value to the customer discounted over seven years is $396 using a 6% discount rate. Based upon value versus cost, it may appear that the decision at this point should be to proceed. However, customers may view the value of fuel economy in a subjective manner rather than in strict economic terms making it advisable to survey how customers would respond to the fuel economy improvement before authorizing the design change and increase in variable cost. Another engineer may have discovered that a reduction of interior noise at 70 mph from 70 to 67 dB could be achieved for a cost increase of $100. The decision of whether or not to proceed rests on understanding how the value to the customer is improved by reducing noise at cruising speeds.

Because the frequency of such decisions during the development of a new product is high, manufacturing enterprises can benefit by having a methodology for systematically forecasting how the value of its products will change as a function of their attributes. To be effective in making trade-off decisions against costs, the methodology must have wide support throughout the organization including marketing, finance, engineering, purchasing, manufacturing, and assembly. Also the methodology should quantify value in a cardinal as opposed to an ordinal sense, should offer insight, and should be straightforward to apply if it is to be widely used throughout the product development process from concept to completion.

The use today of any one of the several available product decision support tools is generally limited to one organizational unit or another. For example, multidimensional scaling [1] and conjoint analysis [2] are important customer survey tools currently used by market research specialists to gain an understanding of the utility of product attributes. However, these tools have received little application by engineers who use tools such as Taguchi's robust design methodology [3], value engineering methods [4], and Quality Function Deployment (QFD) [5] in making product trade-off decisions.

The restricted use of these tools is not surprising as multidimensional scaling and conjoint analysis were developed to examine major product questions at the more deliberate system design stage when the conceptual plans for the product are being established; whereas, the value assessment issues faced by engineers often appear at the subsystem and component design levels with the program launched and the program timing clock running. Taguchi, for example, in developing his theory of quality [3] has focused primarily on parameter and tolerance design issues at the component level.

The field of value engineering has developed as a successful operational discipline for component design, particularly in Japan, by defining value as the ratio of product function divided by cost [4]. However, this approach is sorely limited when needing to make trade-offs between different attributes at the system level because their values are defined in different units.

Because the applications of QFD, Taguchi methods, and value engineering require that value assessments be made, it should be possible to integrate these tools and couple them to the tools of marketing research if a common method for value assessment can be found. The purpose of this paper is to examine an approach to value benchmarking which we believe can be used by all stakeholders in the product realization process. It should be of interest to system level planners because it allows all trade-off decisions at the subsystem and component levels to be made in a manner fully consistent with the guidelines set at the system level. The lack of a common basis for value assessment within and between disciplines, in our view, has been a major impediment in having a product management tool that is broadly used.

Two other criteria must also be satisfied. One is that the tool be sufficiently transparent so that it can be easily understood and used by a wide range of specialists. The other is that it have sufficient rigor so that the results can be embraced with a reasonable level of confidence. In keeping with these criteria, the model used here for product demand [6] was developed following the guidelines of Little's decision calculus which is that a model should be easy to understand, not easily give bad answers, and treat the major areas of interest [7]. In what follows, four complementary techniques:

- Demand and Price Analysis,
- Customer Surveys,
- Value Curves, and
- Economic Assessment,

are integrated to form a value benchmarking methodology for products as summarized in Figure 1.

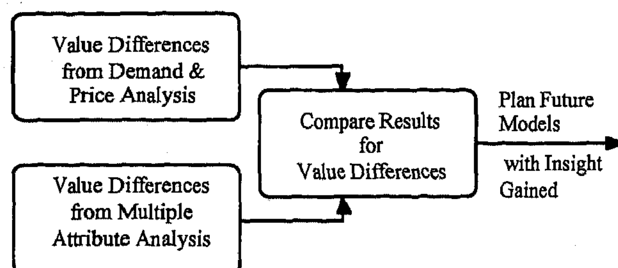


Figure 1: The value benchmarking process.

Vehicles in the mid-sized family automobile segments are evaluated here to provide a partial test of the methodology. The process begins by determining the values and value differences of the products currently competing in the market segment of interest by an analysis of market demand and prices (DP method). The results are then compared to value differences computed from the multiple attribute differences between the products. Three complementary techniques are used to assess how value is influenced by changes in product attributes: a direct value survey (DV method) for highly subjective attributes, value curves (VC method) for performance attributes which are not highly subjective, and economic value calculations (EV method) for attributes which lend themselves to straightforward economic analysis. The survey method can also be used as well for determining the value of selected performance and economic attributes.

There are important advantages to using several methods to evaluate how value is generated by changes in product attributes. Value assessment by one method, for example, can be crossed checked by one of the other methods. Also one method may be easier to use for one attribute versus another in terms of time and resources required. Although variants of each of the three methods for value assessment have been used by others, the effort here of bringing several value assessment methods to bear on the same problem and of comparing differences in product values computed from attributes with those computed from the analysis of demand and price is unique.

If the comparison between the two approaches for computing the value differences are in reasonable agreement, then insight and confidence has been gained for forecasting the value of future products based upon the changes planned in their attributes. Although the products studied here are automobiles, the methodology is general and should be applicable to other products including software and services.

TOTAL PRODUCT VALUE

LINEAR DEMAND MODEL -- The values or utilities of products currently competing within a market segment can be estimated by using a model for their annual demands or market share in terms of their prices and values. Simon [8] has evaluated several models for the demand D_i of product i as a function of its price and the prices of competing products. The simplest model studied by him:

$$D_i = a - bP_i + c\overline{P}_{j\neq i} \quad (1)$$

seemed as good as the more complex models where a, b, and c are constants and $\overline{P}_{j\neq i}$ is the average price of the products competing against i. The above expression can obtained be from a Taylor's expansion of demand as a function of price and keeping only the linear terms. When considering the variables of price <u>and</u> value, Cook and Kolli [6] have shown that demand is given by:

$$D_i = K\left\{V_i - P_i - \frac{1}{N}\sum_{j\neq i}(V_j - P_j)\right\} \quad (2)$$

based upon a Taylor's expansion in the vicinity of a so-called "cartel" reference point, Figure 2, where each competitor has the same reference demand D_R, price, P_R, and value, V_R.

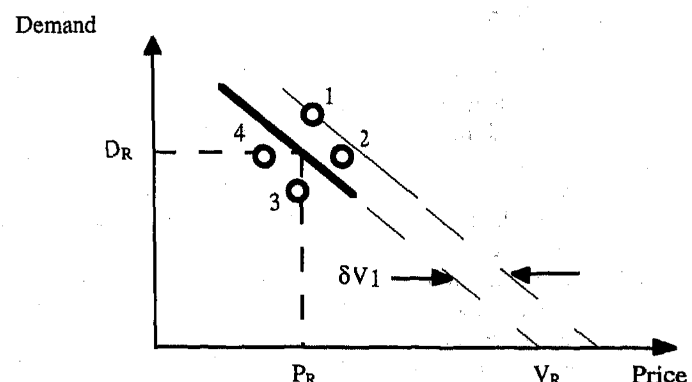


Figure 2: Movement from the cartel reference state when each product is differentiated in value and not offered at the same price.

The interpretation of the intercept shown in Figure 2 as the value of the reference state is fitting as the difference between it and the reference price, P_R, is a measure related to consumer surplus. The demand curve need not be linear over its entire range for this definition of value to be appropriate as linearity is assumed only in a region about the reference point in writing Eq. 2. All of the prices and values in Eq. 2 are treated here as variables.

The value increase from the cartel reference point of product 1 is noted as δV_1 in Figure 2. The slopes drawn are for the case when the prices of the four products are increased or decreased by the same amount and are equal to K/N The parameter $-K$ is the partial derivative of demand, D_i, with price, P_i, of a single competitor at the reference point, N is the number of competitors, and V_i is the value of product i. If one of the producers chooses to change price alone, however, then its slope of demand versus price is steeper and equal to K.

The cartel point is chosen to be the point defined by $\overline{P}$ and $\overline{D}$, the average price and demand, respectively, for the N competitors as this minimizes the error in using the linear approximation. When value is taken as a constant, the coefficient c, on comparing Eq. 1 with Eq. 2, is seen to be equal to $b[N-1]/N$. When prices and all other attributes but one, defined as g, remain unchanged, Eq. 2 leads to an unambiguous, operational definition for the change in value of product i associated with a small change δg:

$$V_i(g+\delta g)-V_i(g)=\frac{D_i(g+\delta g)-D_i(g)}{K}$$

A partial check of the consistency of the demand model given by Eq. 2 can be made by noting that the ratio of the price elasticity E_1, determined from the change in demand of a product when only its price is changed, to the price elasticity, E_2, determined from the demand slope with price shown in Figure 2 when all of the competitors change price by the same amount but hold value constant, should be equal to the number of competitors N [6]. For the automobile market, $E_1 \approx 6$ and $E_2 \approx 1$ [9] which yields, according to the above model, approximately six for the number of vehicles in the market segment that a buyer on average considers with some degree of interest when making a purchase decision which is not unreasonable. It is useful to note that although demand in Eq. 2 is written in terms of two variables and Eq. 1 is written in terms of only one variable, there is only one expansion coefficient that needs to be evaluated in Eq. 2 versus three in Eq. 1. This is a direct result of the high symmetry in the coefficients for an expansion about the cartel point and the fact that the total demand for the N members of the cartel should equal the demand for a monopolist selling the same product.

The quantity E_2 is equal, in theory, to the price elasticity for a monopoly at the reference state and can thus also be defined in terms of the value and price of the product in the reference state:

$$E_2=\frac{P_R}{V_R-P_R}\approx\frac{\overline{P}}{\overline{V}-\overline{P}} \quad (3)$$

where $\overline{V}$ is the average value of the N products being considered. Because E_2 is independent of N and $E_1 = NE_2$ is not, E_2 is the more fundamental elasticity.

For products currently in production, the N simultaneous equations represented by Eq. 2 can be solved for the values for each of the competing products in terms of their demands and prices yielding:

$$V_i=\frac{N[D_i+D_T]}{K[N+1]}+P_i \quad (4)$$

where D_T is the total demand for the N products. Value determined in this fashion is independent of how the producer sets price and represents *perceived value* [10] because the customers have purchased but not yet used the product. The coefficient, K, can be estimated from the ratio of demand and price in the reference state and the price elasticity E_1:

$$K=\frac{D_R}{P_R}E_1\approx\frac{\overline{D}}{\overline{P}}E_1 \quad (5)$$

The difference in value of product i relative to product k in the segment is given by:

$$V_i-V_k=\frac{\overline{P}}{[N+1]E_2}\left[\frac{D_i-D_k}{\overline{D}}\right]+P_i-P_k \quad (6)$$

The above expression can be used to determine value differences between two products from their differences in demand and price (DP analysis) when the demand and price differences are not large. This expression can also be used to determine the value difference between an alternative, i=2, and a baseline, k=1, from paired comparisons in a survey by rewriting it in the form for N=2:

$$4f-2=\left[\frac{3E_2}{P_1}\right]\{V_2-V_1-[P_2-P_1]\} \quad (7)$$

where $f=D_2/[D_1+D_2]$ is the fraction of persons in the survey that choose the alternative over the baseline when the alternative was priced at P_2 versus the baseline priced at a constant reference amount P_1 which is used instead of an average price $\overline{P}$. A linear plot of $4f-2$ versus P_2 goes to zero at the point f=0 ($D_1 = D_2$) and the value of $P_2 = P_N$ (the neutral price) at this point allows the value difference between the alternative and baseline to be computed from the relation $V_2-V_1=P_N-P_1$. The price elasticity E_2 can be solved for from the slope of the line.

LOGIT MODEL -- The logit model [11, 12] for the ratio of any two competing demands is given by:

$$\frac{D_i}{D_k} = \exp^{(U_i - U_k)} \tag{8}$$

where U_i and U_k in Eq. 8 are the respective utilities of competitors i and k. We will use the logit model to provide an assessment of the limits of applicability of the linear model and to explore the relative strengths and limitations of the two models for DP analysis as well as for the analysis of results from DV surveys. To compare the models, the parameters in linear model given by Eqs. 6 and 7 need to be related to the utilities in Eq. 8. On expanding the natural logarithm of Eq. 8 about the same D_T / N cartel demand point described above and noting that $Ln(D_i / D_k)$ is approximately equal to $[D_i - D_k] / \overline{D}$, we find that:

$$\begin{aligned} Ln(D_i / D_k) &= U_i - U_k \\ &= \frac{[N+1]E_2}{\overline{P}} \{V_i - V_k - [P_i - P_k]\} \end{aligned} \tag{9}$$

On rearranging Eq. 9, we obtain the logit model expression for the value difference between products i and k determined from the differences in demand and price (DP analysis):

$$V_i - V_k = \frac{N\overline{P}}{[N+1]E_1} Ln\left(\frac{D_i}{D_k}\right) + P_i - P_k \tag{10}$$

which should be compared to Eq. 6 for the linear model. When paired comparisons are analyzed in a DV survey to determine their value differences, the expression for the logit model equivalent to Eq. 7 is given by:

$$Ln\left(\frac{f}{1-f}\right) = \left[\frac{3E_2}{P_1}\right]\{V_2 - V_1 - [P_2 - P_1]\} \tag{11}$$

DETERMINING VALUE DIFFERENCES FROM ATTRIBUTE DIFFERENCES

Customer needs for automobiles are generally stated in terms of performance, appearance, comfort, reliability, durability, safety and operating costs. Several transformations may be required in going from a statement of needs by a customer to the quantifiable product attributes, which we term *system attributes*, that best connect the product to those needs. A customer may state, for example, that a quiet car is important. However, the physical scale for this attribute, sound pressure intensity, does not correlate as well to what persons perceive as noise as the dB psychometric scale which is based on the logarithm of the physical scale.

A customer may also give importance to quickness which has several system level measures including the acceleration time t_a from 0 to 60 mph and from 40 to 70 mph when passing on a two lane road. For this situation, it is necessary to weight the importance of each subset of the attribute when making an overall assessment of the value of acceleration. Moreover, a transformation to a psychometric scale given by $Log(1/t_a)$ may also be in order for this attribute as the driver's experience is likely dominated by the acceleration force and not time itself [13].

The value change associated with the change in a subsystem attribute should be evaluated from its impact on the system level attributes. For example, the value change due to an increase in horsepower (an attribute of the engine subsystem) should be determined through the sum of the changes in value it generates in system level attributes such as acceleration, top speed, and fuel economy. The weight of a car is a good example of a system level property that influences many system level attributes but is not in itself a system level attribute (by the definition used here) because customers do not experience the vehicle's weight directly. By contrast, the weight of a laptop computer is a system level attribute because customers carry them in their hand.

CUSTOMER SURVEYS -- The first extensive use simulated market conditions for assessing real market demand was made by Pessemier [14]. The DV method [15] used here blends elements of contingent valuation (see for example, [16] and [17]), conjoint analysis [2], and prospect theory [18] to estimate the value difference between a baseline product and an alternative.

There are two major differences between the conventional conjoint approach and the DV method. In conjoint analysis, price is a product attribute like any other whose utility must be determined. In the DV method, however, price is used to arrive at a direct estimate of the difference in value between a baseline product and an alternative based upon stated willingness to pay (WTP) or accept (WTA), Eqs. 7 and 11, respectively, for the linear and logit models. In this fashion, an absolute or cardinal estimate of the changes in the value of a product resulting from changes in its attributes can be obtained from customer surveys in a direct and straightforward manner. The procedure also incorporates a constant baseline in assessing value, the importance of which has been established from prospect theory.

In their development of consumer choice allocation experiments, Louviere and Woodworth [11] have described weaknesses in the theoretical underpinnings of classical approaches to conjoint analysis in forecasting consumer choices. The DV method is more nearly connected to their formalism than conjoint experiments. The major differences are not fundamental but expedient. One-at-a-time experimental designs (paired choice sets) are generally used for DV customer surveys because they are simpler cognitively for the respondent to work through than the statistically more efficient orthogonal arrays. Also a range of prices is used for the alternative versus the fixed price baseline product. From this, a neutral price is determined where the respondents in aggregate are indifferent between the baseline and the alternative, Eqs. 7 and 11. The value difference between the two is then computed directly as the neutral price less the baseline price.

The overall sampling error for the one-at-a-time design can be made equivalent to a more efficient design by increasing sample size. This is a small price to pay for the ease provided to the respondent in working through the survey. The effect of possible interactions between specific main effects can also be evaluated using the one-at-a-time design by simply bundling the factors of interest together as one might plan to market them (not as the combinatorial explosion of all possibilities). Although the DV method can

be used to obtain the value of most any product attribute, it should prove to be particularly useful in assessing the value of highly subjective product attributes such as the value of a product's appearance or the value of a brand name.

VALUE CURVES (VC) -- The VC method for attribute value assessment is based upon the premise that the value of performance attributes can be represented by continuous value curves if the performance measure is itself continuous. A similar concept is, of course, found in classical utility theory, except that the value curves here are not determined from a lottery but from the use of simple extrapolation functions that go through special attribute/value points.

Each performance attribute g is assumed to have three well-defined specification points, the first being the so-called ideal point g_I where value for the attribute is at its highest level (or a point beyond which further attribute improvements should give no noticeable benefit), the second being the baseline point g_0, and the third being the critical point g_C where the product becomes valueless independent of the level of other attributes [19].

For example, if an automobile were satisfactory in all other respects but the turning radius was so large it could not negotiate a right angle turn at a normal intersection, it would have no value. Or if its interior noise level was 120 dB, the threshold of pain, and could not be changed, it would be valueless even though otherwise perfect. The value curve for an individual will depend, in general, upon lifestyle (spirited versus conservative), income, education level, age, sex, etc. of the individual. The value curve used for a particular consumer segment such as buyers of mid-sized family sedans represents an average of their value relationships.

A representative value curve having these properties is shown schematically in Figure 3. Importantly, value curves defined in this manner have been directly linked [19] to Taguchi's loss of inferior quality, the curve in Figure 3 being known as a smaller is better (SIB) condition. The ideal location points used in the unfolding methods of multidimensional scaling described by DeSarbo and Rao [20] and DeSarbo and Manrai [21] are similar, at least conceptually, to the ideal specification point used in constructing value curves of the type shown in Figure 3.

Another feature of the VC method is that it offers a relatively rapid method for estimating value because the curves can often can be generated without a market survey and thereby used to quickly identify where the major opportunities for value improvement lie. If the baseline truly represents the expectations of the persons in the market segment for the attribute, results from prospect theory [18] suggest that value curves should have a discontinuous change in slope at the baseline point with the loss curve, as represented by the dotted curve in Figure 3 beyond g_0, falling off twice as fast as the gain. The baseline should truly represents the expectations of the customer if the value curve is assumed to have a break in slope. Of considerable importance is the fact that the curves can be expressed in absolute units of value by setting the value at the baseline point equal to the value determined for the current product in the marketplace from Eq. 4.

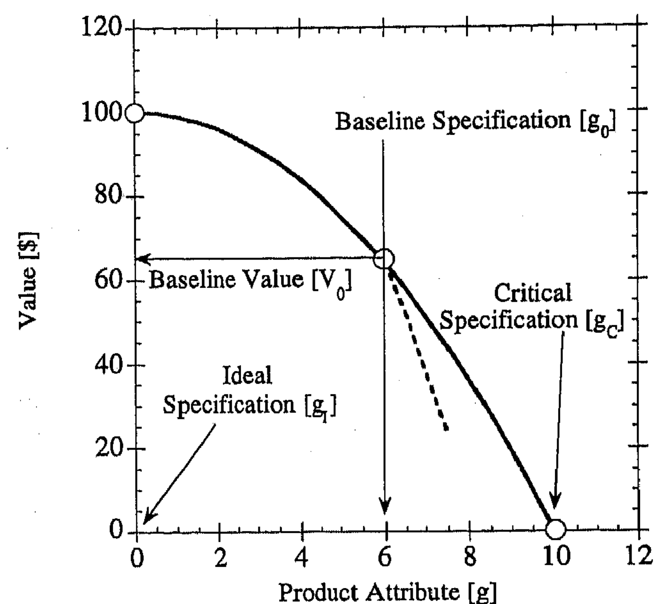


Figure 3: A three point value curve for a smaller is better condition showing a prospect loss break.

The procedure for estimating how a continuous performance attribute is expected to affect product value is to first make estimates of the critical and ideal specifications for the attribute and then substitute them along with the baseline specification into the expression given by [6]:

$$v(g) \equiv \frac{V(g)}{V_0} = \left\{ \frac{[g_C - g_I]^2 - [g - g_I]^2}{[g_C - g_I]^2 - [g_0 - g_I]^2} \right\}^{\gamma} \tag{12}$$

The exponent γ weights the attribute and for making a quick, back of the envelope calculation, it can be taken as equal to the time fraction that the attribute is influential during the use of the product.

For example, interior noise is affected by several environments such as city traffic, idle, full throttle acceleration, and cruising at 70 mph. Noise can be different for each condition with the exponent for each being approximately equal to the time fraction for the condition. The family of curves shown in Figure 4 are based upon one-half of a parabola between 40 and 110 dB having four different weighting exponents. The choice of 40 as ideal (meaning here a point at which no additional benefit accrues by reducing noise) and 110 dB is based upon a comparison by McConville [22] of the value curve with results of Leatherwood, Dempsey, and Clevenson [23]. The choice of 0.05, 0.2, 0.5 for the weighting exponents in Figure 4 are based upon our estimates of the weighting for noise during maximum acceleration in first gear, during idle, and during cruising at 70 mph, respectively.

The relationship of value versus interior noise in Figure 4 for $\gamma = 0.5$ can be used to analyze the hypothetical trade-off discussed a the outset between the reduction of noise at cruising from 70 to 67 dB for a cost increase of $100. From

Figure 4, we find that value is forecasted to improve by 2% for this 3 dB noise reduction. As we will show subsequently, the baseline value for a mid-sized family sedan computed from Eq. 3 is approximately twice its price or $33,000. This results in a forecasted increase in value of approximately $600 for the 3 dB reduction which is $500 more than the hypothetical variable cost increase of $100 stated earlier as needed to make the noise reduction.

The best estimate of the weighting coefficient is obtained by having customers evaluate the difference in value $V(g)-V_0$ between the baseline product and one or more products with the attribute levels different from baseline using the DV method just described. For the noise problem, the different attribute levels could be noise reductions of 5 and 10 dB, for example. This generates additional points on the value curve shown in Figure 4 and allows an estimate of the weighting exponent to be made using a least squares fit to all the points. This approach has been used to determine the weighting exponent of 0.16 for acceleration time from 0 to 60 mph [13]. This process also limits the impact of errors made in the determination of the baseline value V_0 as the DV results calibrate the overall value curve model when the exponent is determined in this manner.

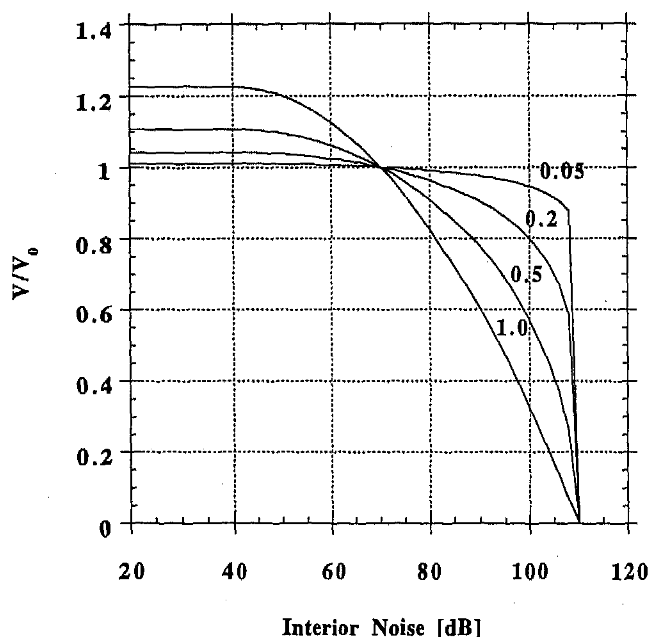


Figure 4: A family of value curves interior noise for four different exponential weighting coefficients.

When multiple product system level attributes g_1, g_2, g_2, ...are considered, the total value of the product is assumed to be determined by the heuristic expression [15]:

$$V_i = V_0 v(g_1)v(g_2)v(g_3)...v(g_j)+\Delta V_{Opt} \quad (13)$$

The term ΔV_{Opt} includes the value of salvageable options as well as the value of other attributes which are assumed to simply add to the overall value of product i in that they do not have a well-defined critical specification. The form of the above equation was chosen as the multiattribute value expression because it has the properties 1) that the value of the product goes to zero, apart form salvageable options, when any of the attributes reaches its critical specification and 2) that a linear relationship for value results with the attribute changes weighted by γ when the attribute changes from baseline are small.

ECONOMIC VALUE (EV) -- The EV method for assessing the value of product attributes directly is based economic considerations. For example, the cost savings resulting from an increase in fuel economy or reduction in maintenance costs of a vehicle should increase product value equal to the savings computed, suitably discounted over time. However, if fuel economy or maintenance costs fall below the expectations of the customer, prospect theory indicates that the value losses could be roughly two to three times their computed economic losses [18].

PERCEIVED VALUE DIFFERENCES OF 1993 MODEL YEAR CARS IN THE MIDDLE SEGMENT

PRIOR INVESTIGATIONS -- Automotive markets offer a rich data base for testing the influence of product attributes on demand. Of direct interest to the problem being considered here is the analysis by Agarwal and Ratchford [24] based upon the pioneering theoretical developments of Lancaster [25] and Rosen [26]. They investigated five vehicle attributes: engine displacement, handling, ride, time to pass, luggage volume, and rear leg room and determined the price elasticities for each attribute using linear regression. A 1% increase in displacement was found, for example, to yield a 0.035% increase in price. More recently, Kwoka [27] has used Katz's [28] price-quality model for market segmentation to analyze certain Chrysler models in terms of vehicle length which was used as a proxy for overall quality.

McCarthy and Tay [29] performed a regression analysis to consider how the utility of attributes as determined from a logit model differed between U.S. buyers who purchased European, Japanese, and domestic vehicles. Horsky and Nelson [30] considered the problem of positioning a new automotive brand at an optimal point in terms of performance, dependability, comfort, prestige, style, and price. Each of the attributes, except for price, were expressed on a perceptual rating scale that ranged from 4.0 to 7.0 for these attributes. Utility differences were determined using the logit model. The results were expressed in terms of a model linear in the attributes and the coefficients were determined using regression methods.

The regression analyses spanned a range of segments to gain a sufficient number of competing vehicles to obtain the coefficients for the attributes considered. There is strong correlation, however, between certain vehicle attributes when viewed across market segments [27] and this should influence the integrity of the regression coefficients if they are to be used to evaluate specific, individual product trade-off decisions within a given segment that do not involve those correlations. Moreover, vehicle features or option content may not be available in smaller, economy vehicles but offered as standard equipment in larger more luxurious vehicles thereby making increased vehicle roominess alone appear more valuable when evaluated across segments than when evaluated within a segment. Because of these limitations, it is necessary to utilize more direct methods such as conjoint analysis, contingent valuation, multidimensional scaling, etc. for estimating how the myriad array of individual attributes affect value.

The DV, VC, and EV methods, are utilized here separately and in combination as already discussed, to estimate how vehicle value is directly influenced by attributes. Also included in the calculations will be the differences in value resulting from different option content which can be sizable even within a segment. The results for the combined attributes are then checked against the overall value differences determined from DP analysis, Figure 1. Alhough a comparison of this kind between the values of competing products as determined from their attribute differences and from their differences in price and demand has not been made before, this type of comparison is important in arriving at a reliable, quantitative insight into the origins of the value differences between complex products and in understanding how those attributes can be best changed to improve customer satisfaction in a cost effective manner.

MIDDLE PRODUCT SEGMENT --- Five vehicles from the 1993 middle product segment of family vehicles which includes Accord, Camry, Century, Ciera, Cutlass, Dynasty, Grand Prix, Lumina, Regal, Passat, Premier, Sable, and Taurus were selected for study. The five had been available for several years without a total redesign thus buyers should have been reasonably informed about them. Total sales for the five selected comprised 70 pct of the total sales (retail plus fleet) for this segment. All vehicle models -- 2 door and 4 door sedans and station wagons -- were included in the value analysis but customers surveyed were restricted to buyers of the 4 door sedans which was the largest group in terms of sales.

Although the ratio of $E_1 / E_2 = 6$ for the two elasticities discussed earlier indicates that customers may roughly consider up to six vehicles when making a purchase decision, limiting the number to five for the analysis reduced the size required for the market survey. In this regard, the error introduced when using Eq. 9 to compute the difference in value between two vehicles for N equal to five versus seven, which is likely an upper limit, is less than 5 pct based upon differences in the ratio of $N/[N+1]$.

The names of the five vehicles are disguised using the letters A through E to avoid any presumption of being judgmental on the merits of one brand versus another because the value metric, although of importance, is insufficient in itself for making a purchase decision as the customer must weigh value against price. Vehicles having a computed perceived value less than the average value of the vehicles in a segment can be very popular in terms of demand and customer satisfaction if their prices are right. Also one individual may judge a given vehicle quite differently than another, but the values computed represent a quantity that must be considered as an average over the market segment. Moreover, the results are for the 1993 model year and they do not represent the values for current vehicles even though the brand names are the same.

DEMAND AND PRICE ANALYSIS RESULTS -- The resulting values determined from Eq. 4 are shown in column 4 of Table 1. The value differences in column 5 were computed by subtracting $33,556 (the average value for the five vehicles) from the values listed in column 4. The figure of $33,556 can be thought of as the value of a baseline vehicle whose dimensions and performance attributes are a composite average of the five. The retail annual sales and retail (median) transaction prices used in the computations are proprietary to one of the sponsors. With permission, we used the figures (but can not disclose them) in computing the value results shown in Table 1. The value differences computed from the logit model, Eq. 9, with D_k constant and equal to 0.2 are shown in the last column. Only for vehicle E, which had the lowest demand, is a significant difference seen between the DP analyses for the two models in Table 1.

Table 1: Values and value differences computed from demand and price analysis.

Model	Annual Number of Retail Sales	Transaction Price	Linear Model Value	Linear Model Diff.	Logit Model Diff.
A	*	*	$33,516	($40)	($63)
B	*	*	$36,458	$2,902	$2,844
C	*	*	$34,598	$1,043	$767
D	*	*	$30,187	($3,369)	($3,483)
E	*	*	$33,020	($536)	($1,516)
\verages=	200,440	$18,303	$33,556		

Because the annual sales figures used for demand were equal to the so-called "retail sales" computed as total sales less fleet sales, buyers in this segment have likely purchased the vehicles for their own personal or family use. Two of the five vehicles chosen (C and D) were available in both two door and four door models and four brands (A, B, C, and E) were also available as station wagons.

The negative demand slope of $K = 65.7$ annual units per dollar for the segment was computed using Eq. 5 for $E_1 = 6$. The reference demand and price were set equal to the average retail sales and transaction prices of the five vehicles. The reference quantities should be near the desired but fictitious cartel point about which the expansion in Eq. 2 is made. The differences in value shown in Table 1 are sizable in terms of how they can affect market share. A $2,000 value difference, on multiplication by K, equates to a change of roughly 120,000 units in annual sales *if no price adjustment were made*. For deviations in K of +/- 10%, the value differences in column 5 of Table 1 were found to deviate by -$181 to $153.

DIRECT ASSESSMENTS OF ATTRIBUTE VALUE -- A total of forty-one product attributes were evaluated for each vehicle, Table 2, to generate the attribute value differences for comparison to the DP results listed in Table 1. Of these, twenty-six represented different levels in available options. Two important attributes not considered were the differences in appearance of the vehicles and vehicle ride quality as they were too costly to properly investigate with the resources available. These differences are not likely of sufficient size in themselves to explain the large value differences shown in Table 1 but their uncertainty, of course, contributes to the error in the value differences determined from the comparison of attributes.

The baseline specifications for all the attributes considered, Table 2 column 1, except for major options and name value, were for a composite vehicle formed from the average of the five attributes for vehicles A, B, C, D, and E listed in Table 3. Total value in Table 3 is broken into six categories: Fuel Economy, Vehicle Range, Option value, Name Value, Value Curves, and Station Wagon Value. The value assessment for each category is described in the Appendix. The results shown in Table 3 were computed from the linear model results for name value and WTP for fuel

economy. The value differences were also computed using the logit model results for these two attributes and the differences were found to be small (see Table A.2). The value coefficients in lines 5 through 15 were multiplied together to arrive at the results in line 16a and then multiplied by $V_0 = \$33,556$ (the average value for the segment located at the bottom of column 4 of Table 1) to compute their combined effect on value, Eq. 13, shown in line 16c.

The total value for each vehicle computed from the attributes considered here was calculated using Eq. 12 as follows: The sum of the values in lines 1a through 4 were added to figures in line 16c computed from the value coefficients using Eqs. 10 and 11 to generate the subtotals in line 18. The station wagon values from Table A.5 were then added to arrive at the total attribute value differences relative to the composite vehicle reported on line 19. Because it seems unlikely that customers would be able to detect many of the small differences in the value coefficients in lines 5 through 15, we also considered what error was introduced by considering a threshold hurdle of 2% before using a coefficient. Only two coefficients (shown within borders) satisfied this constraint. The resulting product of the coefficients under this constraint shown in line 16b are seen to differ by no more than 1% ($336) from the product of coefficients without the threshold constraint.

The value differences in line 19 are compared in Figure 5 to the value differences obtained from the linear and logit DP analyses listed in Table 1. The straight line indicates where the points should fall if the attribute and DP approaches to determining the value differences were in exact agreement. The error bars shown on each side of the linear model points are for two standard deviations computed from the error between the linear model differences in Table 1 and the corresponding attribute differences in line 19 of Table 3. (Note, the degrees of freedom were four (not three) for this computation as the slope of the line is a fixed at unity and thus is not a statistical quantity computed from the experimental data.)

Except for vehicle E, the logit and linear models yield similar results. The minor shifts in the attribute value differences (y direction) in Figure 5 reflect the minor differences in the two model values shown in Table A.2. However, the fact that the linear model places the point for vehicle E closer to the line is not evidence that the linear model is better. We suspect that the larger deviation of E from the line for the logit model may result of the slope K in Eq. 5 being smaller for this vehicle. Although the advertising budgets are not available, there is anecdotal evidence that vehicle E was subordinate to its sister platform, vehicle A, in receiving sales promotion thereby likely reducing K for vehicle. (Less advertising is assumed to reduce demand but value is unchanged to a first approximation; thus, the slope K is reduced.) If a lower K were used in the calculations expressly for vehicle E, its computed DP value would be greater than the amount shown in Figure 5 for both the linear and logit models. Because the linear model is less sensitive than the logit model to changes in K on value, the DP results computed from it are closer to the theoretical line.

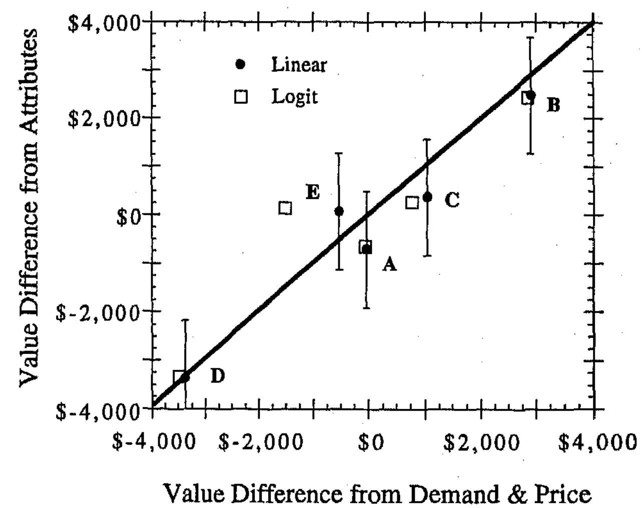


Figure 5: Comparison of vehicle value differences.

DISCUSSION

Apart from vehicle E as already discussed, reasonable agreement is seen in Figure 5 between the value differences computed from attributes and those computed from the analysis of demands and prices. Vehicle B has the highest value of the five. Based upon the results listed in Table 3, this arises from its relatively high name value, its 3 to 7 pct favorable noise level, and its high option content value. Vehicle D has the lowest value as a result of its low option content value, largely from not having an air bag, and its lower name value. The absence of a driver's air bag generated a loss, see Table A.3, of $1,000 for this vehicle relative to the other four which had it as standard equipment. The size of the loss is not unreasonable as it is likely that customers expected vehicles in this segment to offer a driver's air bag which results in vehicle D suffering a loss equal to 2 to 3 times the value of adding an air bag to the vehicle considering the findings of prospect theory [18].

Again, note that lower value does not necessarily imply lower demand and/or lower customer satisfaction as both should be a function of value received minus price paid and not value alone. Vehicle D had the lowest price of the five vehicles. Although vehicle C had the highest name value and highest value for fuel economy, its overall value was reduced below that of vehicle B because of its reduced interior room, higher interior noise level, and lower value of its option content.

Several of the attribute differences computed through the use of value curves, lines 5 through 15 in Tables 2 and 3 were likely too small to have generated a noticeable value difference between the vehicles studied here. These attributes, nevertheless, could be improved upon to differentiate the value of future vehicles.

Table 2: Attributes considered for their impact on value. SIB and LIB refer to smaller is better and larger is better, respectively.

Line	Attributes	Type	Attribute Units	Baseline	Critical	Ideal	Exponent
	Fuel Economy						
1	**Overall mpg**	Economic Survey	mpg	23.69	N/A	N/A	N/A
	Fuel Tank						
2	**Volume Fuel Tank**	Economic	gal	16.92	N/A	N/A	N/A
	Option Value						
3	**Major Options**	Economic	N/A	Vehicle A	N/A	N/A	N/A
	Name Value						
4	**Stated Willingness to Pay**	Survey	N/A	Vehicle A	N/A	N/A	N/A
	Normalized Value Curves						
5	**Interior Room**	Economic & Survey		N/A	N/A	N/A	N/A
6	**Turning Circle**	SIB	meters	12.2	20.54	1.83	0.125
7	**Overall Length**	SIB	millimeters	4853	7620.00	762.00	0.125
8	**Overall Width**	SIB	millimeters	1782.6	2743.20	762.00	0.15
9	**Door-top to ground**	LIB	millimeters	1278.1	1127.76	1498.60	0.03
10	**Luggage Capacity**	LIB	cubic meters	0.46	0.00	0.85	0.08
11	**Idle Noise**	SIB	dBA	42.8	110.00	40.00	0.20
12	**Max 1st Gear Noise**	SIB	dBA	71.3	110.00	40.00	0.05
13	**70 mph Noise**	SIB	dBA	66.8	110.00	40.00	0.50
14	**Time for 0-60 mph***	SIB	seconds	10.8	40.00	2.00	0.16
15	**Skidpad Lateral Acceleration**	LIB	G's	0.72	0.25	1.00	0.20

*Times are transformed to LOG(1/time) which becomes a LIB psychometric force variable.

Table 3: Values computed for attributes (name value and fuel economy WTP from linear model.

Line	Attributes	Vehicle A		Vehicle B		Vehicle C		Vehicle D		Vehicle E	
	Fuel Economy	Attr.	Value	Attr.	Value	Attr.	Value	Attr.	Value	Attr.	Value
1a	Value from EPA Fuel Econ.	23.58	($24)	22.91	($173)	26.13	$473	22.26	($326)	23.58	($24)
1b	Value from Survey WTP		($40)		($278)		$866		($509)		($40)
	Average =		($32)		($225)		$670		($417)		($32)
	Range										
2	Volume Fuel Tank	16.00	($18)	18.50	$11	17.00	$38	17.10	($23)	16.00	($18)
	Option Value										
3	Major Options		$89		$581		($350)		($1,194)		$874
	Name Value										
	Vehicle A Reference		$0		$1,612		$1,938		($1,057)		$0
4	Composite Reference		($499)		$1,113		$1,439		($1,556)		($499)
	Value Coefficients										
5	Interior Room		1.000		1.007		0.962		1.008		1.000
6	Turning Circle	12.19	1.000	11.89	1.003	12.50	0.997	12.19	1.000	12.19	1.000
7	Overall Length	4877	0.999	4770	1.003	4704	1.005	5037	0.994	4877	0.999
8	Overall Width	1808	0.997	1770	1.001	1704	1.008	1821	0.996	1808	0.997
9	Door-top to ground	1252	0.995	1275	0.999	1331	1.007	1281	1.000	1252	0.995
10	Luggage Capacity	0.51	1.005	0.42	0.996	0.40	0.993	0.45	0.999	0.51	1.005
11	Idle Noise	44.00	1.000	41.00	1.000	42.00	1.000	Info. Not Available		44.00	1.000
12	Max 1st Gear Noise	69.00	1.002	69.00	1.002	78.00	0.994	Info. Not Available		69.00	1.002
13	70 mph Noise	68.00	0.992	63.00	1.022	68.00	0.992	Info. Not Available		68.00	0.992
14	Time for 0-60 mph	11.10	0.998	11.00	0.999	10.10	1.006	10.90	1.000	11.10	0.998
15	Skidpad Lateral Acceleration	0.73	1.001	0.72	0.999	0.71	0.997	0.73	1.001	0.73	1.001
16a	Product of Normalized Values		**0.989**		**1.032**		**0.960**		**0.998**		**0.989**
16b	"Threshold" Product		1.000		1.022		0.962		1.000		1.000
16c	Conversion to Value Difference		($385)		$1,070		($1,356)		($57)		($385)
17	Sub-total Value Difference		($844)		$2,550		$441		($3,247)		($59)
18	S/W Average Added Value		$139		($70)		($75)		($124)		$131
19	Total Attribute Value Difference		($705)		$2,480		$367		($3,371)		$71

Where possible, a comparison has been made between the attribute values found here and those reported elsewhere. A value of $250 for a 10% improvement in acceleration performance was used by Greene [31] and Greene and Liu [32] in their study of fuel economy policy. This figure was computed from Sperling, Hungerford, and Kurani's [33] preference survey results of motorists in New York and California in which 50% stated that they would pay $0.10 per gallon extra in fuel costs for a 10% improvement in power. For comparison, the marginal value for a reduction in the 0 to 60 mph acceleration time from 10 to 9 sec computed from the value curve parameters on line 14 of Table 2 is $270. The two are surprisingly close considering that the value curve parameters on line 14 of Table 2 were determined from the stated WTP from persons who actually drove vehicles, which were the same except for their acceleration performances, through a test route [13] and that the result used by Greene was obtained from responses to a written preference survey.

Train and Lohrer [34] reported marginal values for luggage space of 118 to 365 $\$/\mathrm{ft}^3$ (in 1978 \$). However, because this range was greater than the range of 26 to 156 $\$/\mathrm{ft}^3$ for interior volume found by Greene and Liu [31], Greene [32] assumed that Train and Lohrer's results were unreasonably high and chose to use a marginal value of 25 $\$/\mathrm{ft}^3$ instead. We can not support Greene's reason for discounting of Train and Lohrer's finding because the marginal values of interior room and luggage capacity should be the same when the competition between the two spaces is optimally balanced. Since it is in the interest of automobile manufacturers to make design trade-offs properly, the two marginal values should be similar rather than apart.

The calculation here for the marginal value of luggage space of 164 $\$/\mathrm{ft}^3$ (in 1993 \$) taken from its value curve determined from the parameters in Tables 2 and 3 is within Train and Lohrer's [34] range albeit at the low end when inflation is taken into account. Moreover, the assessment here of $404 for one inch of added front leg room to a nominal mid-sized vehicle, which is based upon Simek and Cook's [35] WTP evaluation, translates into approximately two cubic feet of interior room or 202 $\$/\mathrm{ft}^3$ and is comparable, as should be expected from the argument given above, to the marginal value of luggage space obtained here.

Value curves are more robust than marginal values for attributes because they provide a means for estimating diminishing returns as the attribute is improved as well as the size of the prospect loss if the attribute is reduced below what customers have become to expect, Fig 3. Human factor responses are used to estimate the critical and ideal specifications for the parabola and WTP results are used to determine the exponent γ which weights the parabola in Eq. 12.

CONCLUSIONS AND RECOMMENDATIONS

The combined use of the linear and logit aggregate choice models was found to be more beneficial to the overall analysis than the use of either model by itself. The linear model is more insightful because it is directly formulated in terms of demand rather than demand ratio and in terms of dollars for value and price rather than utiles. The model also shows the direct connection between the two elasticities E_1 and E_2 through the relationship $E_1 = NE_2$. It also makes fewer assumptions regarding buyer behavior and readily yields an estimate of total value, V_0, when the price elasticities are known. Knowledge of V_0 is necessary to compute value changes based upon the VC approach.

The logit model is useful for extending the limits of the linear model on demand and price when computing the value differences between existing products. When the two models are combined, Eqs. 10 and 11, the insight of the linear model is preserved and the added range of the logit model is gained. The added range of the logit model is also useful for the analysis of results from DV surveys, although the range of prices used here was not sufficient to detect a significant difference between the linear and logit model results.

The findings of prospect theory, in particular the treatment of gains versus losses and the importance of baselines, which have been developed in laboratory studies [18], should have major implications in real markets as shown by the results of the survey listed in Table A.2. The ownership experience of customers gives them different expectations or baselines.

For example, owners of vehicles having high reliability should view vehicles of lower reliability as having a prospect loss for this attribute. This loss, estimated from Table A.2 at roughly $700 per increased repair during five years of ownership (=$7,025/[2x5]), presents a significant hurdle to an automobile manufacturer desiring to win over customers currently driving a competing vehicle of higher reliability. In other words, adding $300 in value with additional features such as leather seats and an improved stereo system in attempting to win over customers to offset a $300 loss from expectations computed strictly from economic considerations will fail because it represents a true prospect loss of more than $600.

Because the term can be so large, understanding of the sources of the name value attributed to one manufacturer versus another is important to improving sales in competitive markets. If it is largely related to historical product reliability or the treatment of customers at a dealership, these can be corrected over time through well-designed and carefully implemented quality and customer service improvement programs. If it is image, for example, gained over years through superior craftsmanship or a racing heritage, it will be more difficult for other manufacturers to overcome.

ACKNOWLEDGMENTS

The authors are indebted to Greg McConville, John Runnion, and Mark Simek who developed the value curves for several of the product attributes considered here. We are also indebted to Shandon Alderson, Hussein Ali, Eric Monroe, Greg McConville, Becky Silver, and Tony Woods for a critical review of an earlier manuscript. Very helpful discussions with Devanathan Sudarshan are also gratefully acknowledged. Finally, we are deeply grateful to Caterpillar and Ford Motor Company for support of this research through unrestricted grants.

REFERENCES

1. Shocker, A. D. and Srinivasan, V. (1974), "A Consumer-Based Methodology for the Identification of New Product Ideas," *Management Science*, **20**, 921-937.

2. Green, P.E. and Ward, Y. (1975), "New Way to Measure Consumers' Judgments," *Harvard Business Review*, (July-August),107-117.
3. Taguchi, G. (1993), *Taguchi on Robust Technology Development*, The American Society of Mechanical Engineers, N.Y.
4. Fowler, T. C. (1990), *Value Analysis in Design*, Van Nostrand Reinhold, New York.
5. Akio, Y., (1990), ed., *Quality Function Deployment*, Productivity Press, Cambridge, MA.
6. Cook, H.E. and Kolli, R.P. (1994), "Using Value Benchmarking to Plan and Price New Products and Processes," *Manufacturing Review*, 7 (2), 134-147; Kolli, R.P. and Cook, H.E. (1994), "Strategic Quality Deployment," *Manufacturing Review*, 7, (2) 148-163.
7. Little, J.D.C. (1970), "The Concept of a Decision Calculus," *Management Science*, **16**, B466-465.
8. Simon, H. (1989), *Price Management*, Elsevier Science Publishers, Amsterdam, 21.
9. McCurdy, R. (1993), Ford Motor Company, *Personal Communication*, January.
10. Monroe, K. B. (1990), *Pricing Making Profitable Decisions*, 2nd edition, McGraw-Hill, New York, 38.
11. Louviere, J. L. and Woodworth, G. (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, **20**, 350-357.
12. Ben-Akiva, M. and Lerman, S. R. (1985), *Discrete Choice Analysis: Theory and Applications to Travel Demand*, The MIT Press, Cambridge, MA.
13. McConville, G. and Cook, H. E. (1996), "Estimating the Value Trade-off Between Automobile Acceleration Performance and Fuel Economy," *SAE Technical Paper 960004*, Society of Automotive Engineers, Warrendale, PA.
14. Pessemier, E. A. (1959), "A New Way to Determine Buying Decisions," *Journal of Marketing*, **24**, 41-46.
15. Cook, H.E. (1992), "New Avenues to Total Quality Management," *Manufacturing Review*, **5** (4), 284-292.
16. Randall, A., Ives, B. C. , and Eastman, C. (1974), "Bidding Games for Valuation of Aesthetic Environmental Improvements," *Journal of Environmental Economics and Management*, **1**, 132-149.
17. Carson, R. T. (1991), "Constructed Markets," in J. Braden and C. Kolstad, eds., *Measuring the Demand for Environmental Quality*, Elsevier, Amsterdam, The Netherlands.
18. Tversky, A. and Kahneman, D. (1981), "The Framing of Decisions and Psychology of Choice," *Science* , **211**, 453-458.
19. Cook, H.E. and DeVor, R.E. (1991), "On Competitive Manufacturing Enterprises I: The S-Model and the Theory of Quality," *Manufacturing Review* **4** (2), 96-105.
20. DeSarbo, W. and Rao, V. R. (1986), "A Constrained Unfolding Methodology for Product Positioning," *Marketing Science*, **5** (1), 1-19.
21. DeSarbo, W. and Manrai, A. K., (1992), "A New Multidimensional Scaling Methodology for the Analysis of Asymmetric Proximity Data in Marketing Research," *Marketing Science*, **11** (1), 1-20.
22. McConville, G. (1996), *Developing Value Relationships for Automotive Attributes*, M.S. Thesis, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.
23. Leatherwood, J., Dempsey, T., and Clevenson, S. (1980), "A Design Tool for Estimating Passenger Ride Comfort Within Complex Ride Environments," *Human Factors*, **22**, 291-312.
24. Agarwal, M. K. and Ratchford, B. T. (1980), "Estimating Demand Functions for Product Characteristics: The Case of Automobiles," *Journal of Consumer Research*, 7, 249-262.
25. Lancaster, K. (1971), *Consumer Demand: A New Approach*, Columbia University Press, New York.
26. Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, **82**, 34-55.
27. Kwoka, Jr., J. E. (1992), "Market Segmentation by Price-quality Schedules: Some Evidence from Automobiles," *Journal of Business*, **65** (4), 615-628.
28. Katz, M. L. (1984), "Issues In Strategic Pricing," *Journal of Business*, **57** (1), S149-S166.
29. McCarthy, P.S. and Tay, R. (1989), "Consumer Valuation of New Car Attributes: An Econometric Analysis of The Demand for Domestic and Japanese/Western European Imports," *Transportation Research A*, **23A**, 367-375.
30. Horsky, D and Nelson, P. (1992), "New Brand Positioning and Pricing in an Oligopolistic Market," *Marketing Science*, **11**, 160-165.
31. Greene, D.L. (1994), "Alternative Fuels and Vehicles Choice Model," *ORNL/TM-12738*, Oak Ridge National Laboratory, Oak Ridge, TN.
32. Greene, D.L. and Liu, J.T. (1988), "Automotive Fuel Economy Improvements and Consumers' Surplus," *Transportation Research A*, **22A**, 203-218.
33. Sperling, D. and Kurani, K.S. (1987), "Refueling and the Vehicle Purchase Decision: The Diesel Car Case," *SAE Technical Paper 870644*, Society of Automotive Engineers, Warrendale, PA.
34. Train, K. and Lohrer, M. (1982), "Vehicle Ownership and Usage: An Integrated System of Disaggregate Demand Models," report prepared for the California Energy Commission by Cambridge Systematics, Inc., Berkeley, California, March 31, 1982.
35. Simek, M.E. and Cook, H. E. (1996), "A Methodology for Estimating the Value of Interior Room in Automobiles," *SAE Technical Paper 960002*, Society of Automotive Engineers, Warrendale, PA.
36. *Consumer Reports* (1993a), **58** (4), 226-229.
37. *Automotive News 1993 Market Data Book* (1993), Crain Communications, Inc., Detroit, (May), 55.
38. *Automotive News 1994 Market Data Book* (1994), Crain Communications, Inc., Detroit, (May), 74-94.
39. *Consumer Reports* (1993b), 258-263.
40. Simek, M. (1994), *Human Factors Value Modeling Applied to Vehicle Analysis and Development*, M.S. Thesis, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.
41. *Edmund's 1993 New Car Prices* (1993), Edmund Publications Corp., Concord, MA, (June).
42. *Consumer Reports* (1992), **57** (4), 232-233.

43. *MVMA Specifications*, 1993.
44. Runnion J. (1994), *Value Analysis of a Marketable Electric Vehicle Using the S-Model*, Masters Thesis, University of Illinois.
45. Woodson, W.E., Tillman, B., and Tillman, P. L. (1992), *Human Factors Design Handbook*, 2nd edition, McGraw-Hill, Inc. N.Y., 606.
46. *Road and Track* (1992), Hachette Magazines, Inc. New York, (December), 96-97.
47. *Consumer Reports (1991)*, Consumers Union, Yonkers, NY, (June), 395.

APPENDIX: ASSESSMENTS OF ATTRIBUTE VALUE

DIRECT VALUE (CONSUMER SURVEY) RESULTS --The value of the name of the manufacturer, repair frequency, and fuel economy were estimated using the DV method based upon a survey mailed to a random sample of 4000 persons who purchased one of the four vehicles (A, B, C, and D) during the first quarter of 1993. A total 858 quality responses were returned for an overall return rate of 21.5 pct. Although this response rate was low (no monetary inducements or follow-ups were used to encourage people to respond), it is adequate for the purpose of providing a first test of the proposed value benchmarking methodology. The baseline, brand A, was kept constant because findings from prospect theory have shown that it is important make comparisons to a single baseline. The number of respondents that had purchased a given brand was in proportion to the retail sales rate for that brand.

Four surveys were used, the differences being the order of the questions and the prices selected for the alternatives. The order of the questions were varied to minimize possible anchoring effects. For each attribute considered, a range of four prices were evaluated but only three of the four prices were considered in a given survey, the restricted price range allowing the length of the survey to be limited to four pages. This reduced the effective average sample size to 643 (=0.75x858). The results from the survey are summarized in Table A.1.

NAME VALUE OF MANUFACTURER -- The values of the names of the manufacturers of vehicles B, C, and D relative to the name of the manufacturer of vehicle A as determined form the survey are shown graphically in Figure A.1. The resulting values are averages over the WTP or WTA responses (greater or less than the baseline price of $16,000, respectively) given by buyers of vehicles A, B, C, and D. The plot is based upon the logit model, Eq. 11. The equivalent plot for the linear model, not shown, used the function $4f - 2$ for the y-axis, Eq. 7. The slope of the plots should equal $3E_2 / \$16,000$.

Vehicle E was not surveyed because it was manufactured by the same company as vehicle A and had the same platform but was marketed as a separate brand through a different division with certain appearance cue changes.

Table A.1: DV survey results.

Attribute Change	Alternate Price	Total Responses	Number Chosing Alternate	% Choosing Alternate
Vehicle A to B	14000	648	486	75.00
	15000	658	444	67.48
	16000	598	377	63.04
	17000	670	357	53.28
Vehicle A to C	14000	670	534	79.70
	15000	648	514	79.32
	16000	658	423	64.29
	17000	598	342	57.19
Vehicle A to D	14000	658	391	59.42
	15000	598	320	53.51
	16000	670	230	34.33
	17000	648	171	26.39
4 to 2 Repairs per Year	17000	598	501	83.78
	19000	670	341	50.90
	21000	648	122	18.83
	23000	658	110	16.72
4 to 6 Repairs per Year	11000	658	272	41.34
	12000	598	222	37.12
	13000	670	212	31.64
	14000	648	186	28.70
21 to 25 mpg	18000	648	308	47.53
	19000	658	234	35.56
	20000	598	152	25.42
	21000	670	159	23.73
21 to 17 mpg	11000	670	335	50.00
	12000	648	294	45.37
	13000	658	259	39.36
	14000	598	198	33.11

Thus, apart from options, vehicle E was not expected to differ significantly in value from the baseline vehicle and, by omitting it from the survey, the length and complexity of the survey was further reduced. The attribute price in Figure A.1 at the point where $Ln(f/[1-f])=0$ less the baseline price of \$16,000 is equal to the value, ΔV, added by the attribute under consideration.

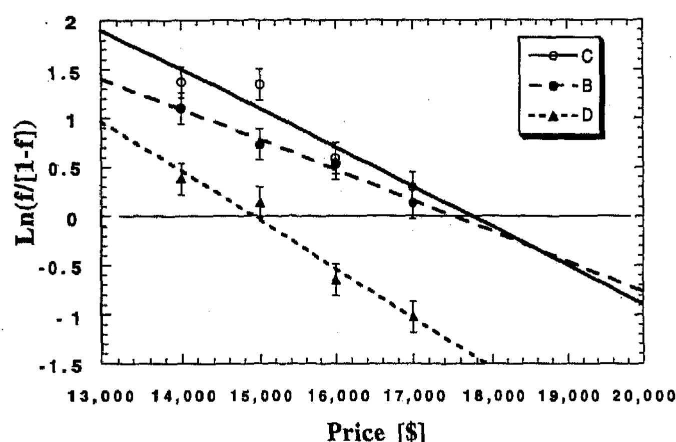


Figure A.1: Graphical determination of name value of manufacturers B, C, and D relative to manufacturer A as determined from the logit model, Eq. 11.

The error bars shown are ± two standard deviations or 0.16, the standard deviation of $Ln(f/[1-f])$ being approximately four times the standard deviation of 2% for f as determined from the binomial distribution for nominal sample size of 600 returned surveys. The agreement of this finding with actual standard deviation of 0.09 computed from the deviations of the points about the line in Figure A.1 is support that the model is representative of the behavior under study.

The DV results computed using both the linear and logit models are summarized in Table A.2 for name value, reliability, and fuel economy. Based upon the R^2 findings shown in column 6, the two models are seen to be statistically equivalent over the range of prices examined. The computed E_2 elasticities in column 7 for the three *simulated* vehicle purchase comparisons are roughly twice the value of unity for this quantity reported elsewhere [9] from the analysis of *actual* changes in vehicle demand with price.

The 90% confidence errors for E_2 shown in the last column of Table A.2 for the three vehicles were computed from the pooled slope errors using a t distribution with 6 df. The E_2 errors for repairs and fuel economy were computed from the respective pooled slope errors using a t distribution with 4 df. Persons may be more price sensitive in the simulated market versus the actual market but considering the error in the measurements of the price elasticity the confidence level is not high.

The difference between the simulated and market elasticities may be due to the choice set in the simulated market being much simpler (the only difference between the vehicles being the name of the manufacturer) than the more complex choice set between vehicles in the real market. Our chief interest in the survey, nevertheless, is the median price persons state that they are willing to pay and not so much as to how this price is distributed about the median.

VALUE OF RELIABILITY -- The difference between the values for reliability of roughly \$3,000 for two fewer repairs over five years or \$300 per repair in Table A.2 seems reasonable to us. The loss of roughly \$7,000 for two more repairs per year over five years (\$700 per repair) is also reasonable when viewed in the context of prospect theory [18] for gains versus losses of comparable sized attributes. This factor of two (roughly) between gains and losses is also seen in the fuel economy values reported in Table A.2.

Table A.2: Results of DVM Survey for Logit (bold) and Linear models.

Attribute Change	Model	Value	90% Conf. +/-	R^2	E_2	90% Conf. +/-
Vehicle A to B	**Logit**	**\$ 1,513**	**\$ 524**	**0.98**	**1.7**	**0.62**
	Linear	\$ 1,612	\$ 475	0.98	1.5	0.55
Vehicle A to C	**Logit**	**\$ 1,749**	**\$ 428**	**0.90**	**2.1**	**0.62**
	Linear	\$ 1,938	\$ 540	0.90	1.8	0.55
Vehicle A to D	**Logit**	**\$ (1,075)**	**\$ 177**	**0.96**	**2.7**	**0.62**
	Linear	\$ (1,057)	\$ 171	0.96	2.5	0.55
Vehicle A to E		Assumed Zero				
4 to 2 Repairs per Year	**Logit**	**\$ 3,382**	**\$ 124**	**0.91**	**3.0**	**0.8**
	Linear	\$ 3,278	\$ 139	0.95	2.7	0.7
4 to 6 Repairs per Year	**Logit**	**\$ (6,824)**	**\$ 708**	**0.99**	**1.0**	**0.8**
	Linear	\$ (7,025)	\$ 818	0.99	0.9	0.7
21 to 25 mpg	**Logit**	**\$ 1,509**	**\$ 292**	**0.93**	**2.0**	**0.57**
	Linear	\$ 1,422	\$ 303	0.92	1.7	0.43
21 to 17 mpg	**Logit**	**\$ (4,902)**	**\$ 457**	**0.99**	**1.3**	**0.57**
	Linear	\$ (4,918)	\$ 479	1.00	1.2	0.43

Although the value per repair results are of interest, they can not be used in the analysis because the repair frequencies of the vehicles are not available. However, the value of reliability is likely factored into the value determined for manufacturer's name. This hypothesis was partially checked by comparing the relationship between name value and the reliability index for these vehicles published by Consumer Reports [36]. A strong linear correlation was found. Vehicles B and C had the same high reliability index and the highest name values in Table A.2. Vehicle D had the lowest reliability index and is seen to have the lowest name value. Vehicle A was in the middle ground for both reliability index and name value.

FUEL ECONOMY -- The value of fuel economy was determined in two different ways: from the survey, Table A.2, and from a direct economic calculation based upon the fuel costs for driving an automobile for 100,000 miles. The EPA fuel economy ratings for the 1993 model year powertrain combinations were obtained from the *Automotive News 1993 Market Data Book* [37]. The average fuel economy, FE_i, for each vehicle was then determined by averaging over the 1993 model year installed powertrain percentages which are listed in the *Automotive News 1994 Market Data Book* [38]. The EPA metro and highway fuel economies were then multiplied by the factors of 0.6 and 0.4, respectively, and added together to arrive at the overall averages shown in line 1a of Table 3.

The economic analysis for the value of vehicle i having fuel economy FE_i relative to the value of the baseline vehicle having fuel economy, FE_0, was set equal to the difference in the amount of money spent for fuel at $1.20 per gallon over 100,000 miles:

$$\Delta V(fuel\ econ.) \equiv 1.20 \times 10^5 \left(\frac{1}{FE_i} - \frac{1}{FE_0} \right)$$

The results of economic analysis determined from this expression are compared with the results of the survey in Figure A.2. The economic calculations are shown for 0, 6, and 10% discount rates over a period of 7 years (14,286 miles/year). The stated willingness to pay for added fuel economy is seen to be about a factor of two greater than the economic value even without discounting. There are two possible explanations for this. Customers could be factoring into their assessment an imputed savings in environmental quality for improved fuel economy or they could be overestimating the economic value of fuel economy when taking the survey but would not do so if the true economic value had been made known to them. Because these questions remain unresolved at this juncture, the stated willingness to pay results for fuel economy in Table A.2 were averaged with the economic estimate (0% discount) to arrive at a final estimate for the value differences in fuel economy. Because the baseline of 21 mpg for the survey was below the fuel economies of all five vehicles, only the slope of the WTP line in Figure A.1 of 355(=1422/4) $/mpg (linear model) was used in arriving at the value of fuel economy based upon the survey results.

The term $\dot{V}_D$, taken as 15 [$/hr], is the estimated value in dollars that a person in the market segment of interest on average places on one hour of his or her time. As evaluated by Simek [40], the factor 6.8 is the time in seconds on average to add one gallon of fuel and 214 is the remaining time in seconds for needed to leave the road and arrive at the fuel pump plus pay the bill and return to the road. The factor of 0.8 represents the fractional amount of the total tank volume filled on average for each fuel stop, the assumption being that the tank is refilled when the fuel remaining is 20% of the total tank volume. Greene [31] calculated the value of range in a similar manner in considering alternate fueled and electric vehicles but used $10 per hour for the value of time versus our $15 here.

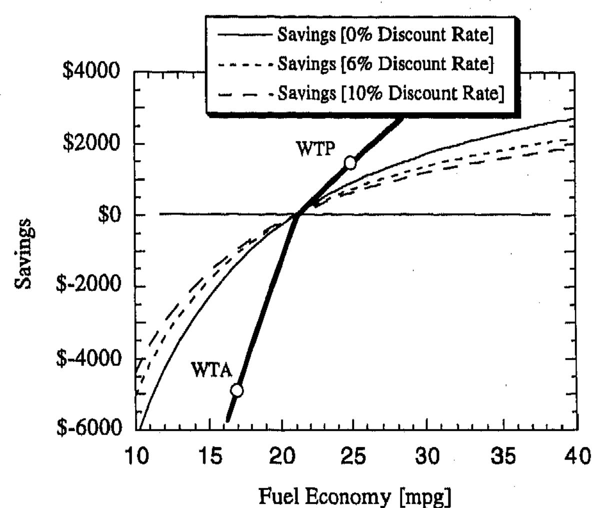


Figure A.2: A comparison of the added value of fuel economy obtained by the economic and survey methods for a 21 mpg baseline.

RANGE -- Increased vehicle range between refueling stops, which is governed by fuel economy and the size of the fuel tank, increases vehicle value as it eliminates lost time in reaching a destination. Fuel tank volumes were taken from the vehicle attribute summary for the 1993 model year given by *Consumer Reports* [39]. The value of a vehicle i having a fuel tank volume of T_i and fuel economy of FE_i was taken as equal to the monetary value of the time lost in refueling the vehicle over 100,000 miles estimated by Simek [40]:

$$V(range) \equiv \frac{10^5 * \dot{V}_D}{3600 FE_i} \left[6.8 + \frac{214}{0.8 T_i} \right]$$

OPTIONS -- When a person buys a car, he or she also generally considers the purchase of several optional features or "options." Here these purchases are considered separately from the larger decision to buy the vehicle in the first place which means that we will compute price elasticities on the basis of the option price instead of the price of the vehicle. Also model option purchases were considered as being independent of each other and the linear model was used for simplicity.

If D persons buy a product annually and half of these also purchase an option, then the purchase price is just equal to the neutral price described in the DV method, the differences between the two prices being equal to the added value of the option. When the percentage is off 50%, we can use the linear model value differences computed from Eq. 4 written in the form:

$$V_{Opt} = V_2 - V_1$$

$$= \frac{2(D_2 - D_1)}{3K_{Opt}} + P_{Opt} \quad (A.1)$$

where D_2 is the number of persons purchasing the option out of the total number of persons $D = D_1 + D_2$ and where V_{Opt} and P_{Opt} are the option value and price, respectively. On substituting for the slope of option demand K_{Opt} in Eq. A.1 taken at the reference point of $D_2 = D_1 = D/2$ the relation:

$$K_{Opt} = \frac{E_{Opt}D}{2V_{Opt}}, \quad (A.2)$$

we have after rearranging that

$$V_{Opt} = \frac{P_{Opt}}{1 - \frac{4[2f_{Opt} - 1]}{3E_{Opt}}} \quad (A.3)$$

where $f_{Opt} = D_2 / D$. If the price elasticity for the option is known then the value of the option can be estimated from its price and its fraction of total sales.

As the fraction of sales of the option increases, its value according to Eq. A.3 increases. This result can be used to arrive at self-consistent relationship for the price elasticity of demand for an option as follows: On summing Eq. 4 over all the products in the segment, we find that the average value of a product in a segment is related to the average price (reference price) by the relation:

$$\overline{V} = \overline{P}\left[\frac{1+E_2}{E_2}\right] \quad (A.4)$$

which is obtained from Eq. 3. As discussed earlier, the elasticity E_2 for automobiles is approximately unity which makes the average perceived in the segment equal to roughly twice the average price for the segment. Therefore, as the purchase rate of an option increases towards 100%, the value of the option should approach twice the option price if the value of the option is to be folded into the overall value of the product in a self-consistent manner. This relationship holds in Eq. A.3 when E_{Opt} is equal to 8/3 which is certainly not unreasonable. With E_{Opt} set equal to 8/3, it also follows that the purchase of an option approaches zero as the option price approaches 1.5 times its value.

The option values estimated using Eq. A.3 for the five vehicles under consideration are shown in Table A.3 and the overall impact of the options on vehicle value are given in Table A.4. The prices represent the option list prices, P_L, for these vehicles obtained from *Edmund's 1993 New Car Prices* [41] except for the 4 wheel disc brake and driver's airbag options which represent estimates as discussed below. The true option transaction price is given by $P_{Opt} = P_L d$ where d is a discounting fraction. In all likelihood, the transaction price lies between the dealer's price (approximately $0.86 P_L$) and option list price.

Blanks in the price listings represent instances where the options were either not available or were bundled together. The latter was typical for vehicle B and almost entirely the case for vehicle C. Because vehicles A and E were built off the same platform, the option values for E were set equal to those of A because the sales of vehicle A were much higher than those for vehicle E which means that results based upon vehicle A should be more reliable. The percentages of installed options were for the 1993 model year were taken from those listed in the *Automotive News 1994 Market Data Book* [38]. The results for import vehicles were determined by averaging the percentages across the vehicles built in North America and built off-shore.

The values of the options for which prices were available were computed from Eq. A.3 using the list prices (d=1) and are shown in bold face in Table A.3. The last column in Table A.3 shows the average of the bold face values for each option. This average value was entered as the value for those options for which we were unable to find a quoted option price.

The list price of $500 for a driver's air bag was taken from *Consumer Reports* for a 1992 mid-sized vehicle [42]. This was necessary because the driver's air bag was standard equipment for vehicles A, B, C, and E but was not available on vehicle D. It was also necessary to estimate the price of the 4 wheel disc brake option because its price was folded into the option price for anti-lock brakes in *Edmund's* [41] summary; whereas, these two options were not bundled for all vehicles.

The average *option content value per vehicle* in Table A.4 was computed by multiplying the value per option in Table A.3 by the percentage installed for each option divided by 100. The differences in option value with respect to vehicle A is shown at the bottom of Table A.4 for no discounting of the list price, d=1, and for discounting equal to the approximate dealer price, d=0.86 [41]. It is likely that the actual average price paid for an option is somewhere between the list price and the dealer price. In subsequent calculations, we have used the results for d=0.86.

INTERIOR ROOM -- The results of Simek and Cook [35] based upon combined considerations of the percentage of the adult population which would be satisfactorily accommodated by the major interior dimensions for these five vehicles and of the stated willingness to pay for added space using the DV method, were used to estimate the value coefficients for interior room in line 5 of Table 3. Polynomial expressions up to fifth order were generated for value as a function of front and rear head room, front leg room, rear knee room, and front and rear shoulder room. Stated willingness to pay results for additional space were determined using the DV survey method only for head room and front leg room.

This method for assessing value of interior room did not generate ideal and critical points which is why such points are absent for this attribute in Table 2. The interior room dimensions used in the computations by Simek and Cook [35] for the value of interior room of these five vehicles were taken from the listing published by the Motor Vehicle Manufacturer's Association [43] for the 1993 model year vehicles.

VC METHODS -- The value coefficients for the remaining performance attributes were determined from Eq. 10 after arriving at estimates for the ideal and critical specification

Table A.3: Analysis of the value of installed options.

Options	List Price [$]					Percent Installed					Value [$]					
1993 Model Year	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	Bold Avg
3 Speed Auto				575		0	0	0	83	0	858	858	858	**858**	858	858
4 Speed Auto		800	750			99	96	73	16.9	100	1233	**1490**	**977**	1233	1233	1233
4 Speed Manual	Base	Base	Base	Base	Base	0	0	0	0	0	0	0	0	0	0	0
5 Speed Manual	Base	Base	Base	Base	Base	1	4	27	0.1	0	0	0	0	0	0	0
Anti-Lock Brakes	395	887.5		450	395	39.4	24	44	64	48	**357**	**706**	485	**521**	**357**	485
4 Wheel Disc Brakes	*200*	*200*	*200*	*200*	*200*	5	78	52	100	48	**138**	**279**	**205**	**400**	**138**	232
Styled Al Wheels	239	410			239	37	16	34	31	91	**212**	**306**	243	243	**212**	243
Full Size Spare	73				73	5	70	1	0.1	8	**50**	50	50	50	**50**	50
Cruise Control	224	230		225	224	91.7	93	96	96	100	**384**	**403**	398	**419**	**384**	398
Driver's Airbag	*500*	*500*	*500*	*500*	*500*	100	100	100	0	100	**1000**	**1000**	**1000**	1000	**1000**	1000
Pass Airbag	488				488	47.6	0	2	0	99	**477**	477	477	477	**477**	477
Keyless Entry	164				164	20	0	0	0	20	**126**	126	126	126	**126**	126
Electric Sunroof	776	900			776	5.9	38	34	0	4	**539**	**802**	626	626	**539**	626
Manual Climate Contrl	841	915		830	841	100	96	96	100	100	**1682**	**1694**	1680	**1660**	**1682**	1680
+ Auto Climate Contrl	183				183	8.8	0	0	0	86	**130**	130	130	130	**130**	130
Rear Defrost	170			170	170	100	100	100	85	100	**340**	314	314	**263**	**340**	314
Power Windows	356			297.5	356	82	93	96	70	99	**524**	473	473	**371**	**524**	473
Power Locks	257				257	81.4	93	96	100	99	**375**	375	375	375	**375**	375
AM/FM Stereo	206				206	100	99	97	100	100	**412**	412	412	412	**412**	412
AM/FM Stereo w/Cass	155	290		146	155	81.9	86	97	77	97	**228**	**450**	276	**199**	**228**	276
AM/FM Other	491	990		396	491	3	12	0	3	0	**334**	**716**	440	**270**	**334**	440
Power Antenna	102				102	12	66	0	0	34	**74**	74	74	74	**74**	74
Leather Seat	618	1030			618	8	9	2	0	12	**435**	**732**	534	534	**435**	534
Driver Power Seat	305			270	305	66.9	6	34	20	97	**367**	314	314	**208**	**367**	314
Pass Power Seat	305				305	4	0	0	0	6	**209**	209	209	209	**209**	209

Table A.4: Average value per vehicle of installed options.

Options 1993 Model Year	Average Option Value in $ per Vehicle A	B	C	D	E
3 Speed Auto	0	0	0	712	0
4 Speed Auto	1221	1436	715	208	1233
4 Speed Manual	0	0	0	0	0
5 Speed Manual	0	0	0	0	0
Anti-Lock Brakes	141	172	212	332	171
4 Wheel Disc Brakes	7	219	107	400	66
Styled Al Wheels	78	48	84	76	192
Full Size Spare	3	35	0	0	4
Cruise Control	352	375	381	403	384
Driver's Airbag	1000	1000	1000	0	1000
Pass Airbag	227	0	10	0	472
Keyless Entry	25	0	0	0	25
Electric Sunroof	32	303	216	0	22
Manual Climate Contrl	1682	1627	1610	1660	1682
+ Auto Climate Contrl	11	0	0	0	111
Rear Defrost	340	314	314	224	340
Power Windows	429	441	453	259	518
Power Locks	305	350	359	375	371
AM/FM Stereo	412	407	400	412	412
AM/FM Stereo w/Cass	186	385	268	153	221
AM/FM Other	10	84	0	9	0
Power Antenna	9	49	0	0	25
Leather Seat	35	68	12	0	52
Driver Power Seat	246	18	108	42	356
Pass Power Seat	8	0	0	0	13
	$6,759	$7,331	$6,248	$5,267	$7,672
Value Difference Comp. Ref.					
d=1.0	$104	$675	($407)	($1,389)	$1,016
d=0.86	$89	$581	($350)	($1,194)	$874

points and the exponent. A feature of the VC method is the ability to make quick, directional estimates of the value improvement on changing an attribute. If an attribute is found to have a significant effect, an improved estimate can be made using DV survey method in which respondents evaluate the attribute change versus a baseline. Both methods have been used here.

Runnion [44] estimated the value curve for turning radius based upon a critical attribute equal to the upper limit on turning radius for making a right angle turn while keeping within a 15 foot wide lane. The critical and ideal attributes determined by him are listed in line 6, Table 2.

The critical attributes and exponents for lines 6 through 10 are rough estimates suitable only for directional analyses. Overall length and overall width in lines 7 and 8 represent dimensions which we judge as making the vehicle too large and cumbersome to be driven and parked effectively by most drivers. The exponents for lines 7 and 8 represent a consensus reached by seven persons engaged in value benchmarking research. For line 9, the 95% male and 5% female bent torso heights were used, respectively, for ideal and critical door-top to ground clearances [45]. Although the exponent chosen of 0.03 is small, it reflects what is certainly an upper limit on the fractional amount of time needed to enter and exit a vehicle. The exponent for luggage capacity, line 10 of Table 2, was set equal to an assumed (4/52) time fraction of weeks out of the year that the full luggage capacity would be used on average. As none of these attributes in lines 7 through 10 are seen to have important impact on the value differences in Table 3, it is not likely that a more careful evaluation of their impact would significantly change the results of the attribute value computations in line 19 of Table 3.

The critical value and ideal values for interior noise in lines 11 through 13 in Table 2 are those obtained by McConville [22] referred to in the body of the text. The interior noise levels, acceleration times from 0 to 60 for vehicles A (=E), B, and C, and the skidpad lateral accelerations times in Table 3 were obtained from *Road and Track* [46]. The 0 to 60 acceleration time for vehicle D was obtained from an earlier model [47] and assumed unchanged for the 1993 model year. The exponent of 0.16 represents McConville and Cook's [13] determination, discussed earlier, based upon an analysis of the value of acceleration performance using a combination of VC and DV methods. The critical and ideal attributes listed for skidpad lateral acceleration are our estimates for these quantities and the exponent of 0.2 is our estimate of the time fraction during vehicle operation that handling is important.

STATION WAGON VALUE -- Station wagons were also treated as an option and their values were estimated using Eq. A.3. The results are shown in Table A.5. Although the price premiums for the station wagons are not shown in column 5 of Table A.5 because they are proprietary, the average premium is given at the bottom of the column. The values for vehicles B and C are seen to be somewhat lower because of their low sales versus vehicles A and E. However, the low sales for the wagons from manufacturers B and C may not be representative of low demand but of limited vehicle availability in which case the assumptions assumed in using the model do not apply.

Table A.5: Station wagon option value results.

Vehicle	2 Dr	Fraction 4 Dr	SW	S/W Price Premium	S/W Value	Average Added Value	Composite Reference
A	N/A	0.78	0.22	*	$1,246	$263	$139
B	N/A	0.95	0.05	*	$874	$54	($70)
C	0.16	0.80	0.04	*	$893	$49	($75)
D	0.19	0.81	N/A	N/A	N/A	$0	($124)
E	N/A	0.78	0.22	*	$1,096	$255	$131
			Avg. =	**$1,392**			

970763

Determining the Value of Vehicle Attributes Using a PC Based Tool

E. M. Monroe and H. E. Cook
University of Illinois at Urbana-Champaign

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Product engineers and product planners are routinely faced with trade-off decisions involving the cost of adding a product feature or modifying an existing feature versus its added value to the customer. The purpose of this paper is to assess the use of a personal computer (PC) for surveying respondents' willingness to pay (WTP) for four options – two-tone color, 4x4 drive, sporty trim package, and extended cab -- available on the base 1997 Ford F-150 truck. The results show that the respondents' stated WTP reflected the value of the options as determined from their prices and fraction of sales.

INTRODUCTION

Product value, cost, and the pace of product innovation are the fundamental metrics that generate profits to companies selling products in highly competitive markets [1]. In making trade-off decisions relative to cost, value engineers consider the performance divided by cost [2] of the various alternatives. Although this approach has been found to be very useful in making trade-offs at the component level of design, it is not effective at the system level of design where the trade-off is often between two attributes which have different measures of performance requiring a comparison to be made between quantities having different units. The trade-offs presented by added trunk space, versus reduced fuel tank volume, and/or rear leg room is a good example of this type of system level problem.

There are other approaches, however, which do not have this difficulty because they convert changes in full system level performance to changes in value to the customer in units of dollars. Once this has been achieved, a direct comparison can be made between the added value and the added costs. The process known as contingent valuation [3] has been widely used for assessing the societal benefits in dollars of potential infrastructural investments such as airports, parks, and recreation facilities by asking potential customers what they are willing to pay (WTP) or accept (WTA) for a proposed change. Although the use of WTP for product features has seen only limited use, the approach, in our view, has considerable merit based upon determinations of the value of vehicle acceleration [4], vehicle interior room [5], the name value of a vehicle manufacturer [6], and the value of vehicle options [7] using what we term the Direct Value (DV) survey method to distinguish it from classical conjoint analysis survey [8].

The purpose here is to report on the use of a PC based tool for assessing the value of product features using the DV method. Although commercial software [9] is available for conjoint analysis surveys, we constructed the software used here so that it would provide the added support needed for the DV method for soliciting questions and analyzing the responses. The survey was used to evaluate four options – two-tone color, 4x4 drive, sporty trim package, and extended cab -- available on the base 1997 Ford F-150 truck. The respondents were adults attending the University of Illinois at Urbana-Champaign Engineering Open House in March 1996. This F-150 truck was chosen because it was a new model with representative photographs readily available from information brochures at dealerships showing the third door open and closed and several exterior options.

THE DV METHOD FOR ASSESSING VALUE

The DV survey method combines elements of conjoint analysis [8], contingent valuation [3], prospect theory [10], and choice theory [11]. Paired comparisons are made between a fixed baseline and one or more alternatives and respondents are asked to make a simulated purchase decision (choice) between them. Usually the alternative differs from the baseline in only one attribute and never more than a few. With the features of the alternative and the baseline fixed, the respondent is asked to choose between the baseline at a fixed price and the alternative over a range of prices. By aggregating the results and plotting the fraction f of all of the respondents surveyed choosing the alternative as a function of price, a neutral (median) price can be found where 50% of the respondents in aggregate prefer the alternative and 50% prefer the baseline. The difference between the neutral price and the baseline price equals the added value generated by the product feature or the change in the feature. The added value could, of course, be negative. If we write the baseline price as P_0 and the neutral price as P_N the added value of the feature or option is given by:

$$V_{Opt} = P_N - P_0$$

This is does not imply that the value determined is the price that should be charged for the option or feature. The price of an option or feature should reflect both its cost and value.

PC TOOL

There are several important advantages of a PC based tool over a written survey in assessing the value of certain attributes of a product including the ability to readily generate color images, animation, sound, and noise. Logic can also be included in the software to help the respondent when questions arise and to detect and thus avoid logically incorrect answers. If the demographic information about the respondent is obtained up-front before the survey is presented, the survey given can be selected from a library that targets the specific demographic profile of the respondent. Moreover, the data acquisition and analysis can be automated. Although a PC tool is not today well suited to delivering a survey to a geographically wide respondent base, it can be used at special locations such as shopping areas and dealerships frequented by persons whose demographics are close to those of interest. However, as the computer internet grows in size and speed, PC based market surveys may well become more important for quantitatively assessing customer needs than mail surveys.

The PC tool was divided into six modules as shown in Figure 1. The four central modules -- Instruction Interface, DVM Product Comparisons, Demographic Interface, and Thank You Interface -- controlled the real time interaction with the respondent. The Design Interface at the start and the Data Analysis Interface at the end were used to set up the survey and to analyze the results obtained for value, respectively.

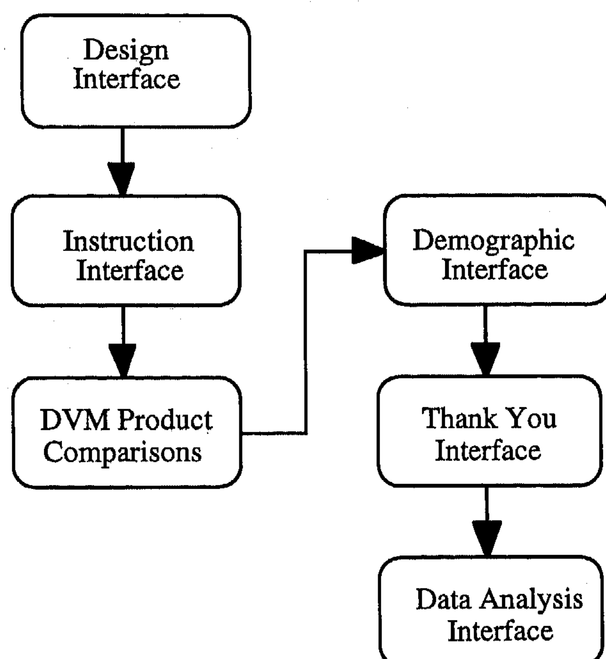


Figure 1: Visual Basic DVM Procedure

INSTRUCTION INTERFACE - The Instruction Interface, the first screen seen by the respondent, is shown in Figure 2. It described what the survey was about and gave general directions to the respondent. The respondent clicked the start survey button to begin.

SURVEY INTERFACE - The PC tool took the respondent through a sequence of screens in which alternatives to the baseline were presented to the respondent and he or she indicated a simulated purchase of the baseline or the alternative at the prices shown by clicking on the appropriate button. One of the survey screens used is shown as Figure 3 in which a baseline 4x4 XLT package at $19,200 is compared against the 4x4 XLT with the extended cab at $24,000. The passenger side access door on the extended cab image could be opened and closed by the respondent using the two buttons just below the picture.

If the respondent clicked the "Buy Alternative" button, the alternative's price was increased and the respondent was again asked to reconsider the choice at the new price level. The baseline price remained fixed throughout. When the baseline was selected, the price of the alternative was reduced. These steps were repeated until the respondent switched from "Buy Alternative" to "Buy Baseline" or vice versa. The highest price the respondent was willing to pay for the alternative was recorded as part of the database and a new feature relative to the baseline was presented to the respondent. This process was repeated until all of the alternatives were considered. Six alternatives formed by combinations of the four features under consideration, Table 1, were presented to each respondent.

Table 1: DV Experimental Design

Vehicle	Base	Two-Tone	4x4 Drive	Sport Trim	Extd. Cab
Base	1	0	0	0	0
Alt 1	1	0	0	1	0
Alt 2	1	0	1	0	0
Alt 3	1	1	0	0	0
Alt 4	1	0	1	1	0
Alt 5	1	1	0	1	0
Alt 6	1	0	1	1	1

The median WTP by the respondents in total and by selected demographics is shown in Table 2. The neutral price was also computed using three other methods: 1) a best fit plot of the WTP prices on normal probability paper, 2) a logit model plot [12] of $Ln(f/[1-f])$ versus price and 3) a linear model plot of $4f-2$ versus price. (The logit and linear model plots yield neutral prices at the point where $Ln(f/[1-f])$ and $4f-2$ are zero.) The results for the normal probability plot were very close to those shown above computed from the MEDIAN function in the spreadsheet and also with those determined from the logit and linear model plots.

Table 2: Summary of Median WTP

Option	All (57)	Own truck (22)	Not Own Truck (35)	Male (35)	Female (22)
Baseline	$19,133	$19,117	$19,150	$19,150	$19,175
Color	$200	$200	$350	$100	$500
4x4 Drive	$2,000	$1.900	$2,300	$1,950	$1,825
Sporty Trim	4433	$467	$400	$500	$250
Extended Cab	$2,433	$2,267	$2,550	$2,400	$3,150

DEMOGRAPHICS INTERFACE -- Upon completing the survey, respondents were asked to complete the demographics interface shown in Figure 4. This information was coded with the WTP response of each respondent. As noted above, if the survey administrator desired to have different surveys for persons in different demographic segments, the demographic information could be asked for initially and logic commands would then present the appropriate survey to the respondent.

THANK YOU INTERFACE - The final interface, shown in Figure 5, thanked the respondent for his or her time. The "Restart Survey" button shown was used by the administrator to restart the survey for the next respondent.

VALUE OF ATTRIBUTES

The WTP results for each of the six alternatives considered were assumed to be represented by a linear model of the form:

$$[X][\lambda]=[Y] \tag{1}$$

where $[X]$ for this problem is the 7 rows of alternate vehicles plus baseline by 5 columns of attributes array of 0's and 1's shown in Table 1. The matrix $[\lambda]$ is the column array of unknown coefficients representing the values that each off-baseline attribute added to the baseline value when present on the vehicle. The array $[Y]$ is one of the WTP column arrays in Table 2 determined for each alternative. The five unknown coefficients $[\lambda]$ were solved for using the regular least squares solution given by:

$$[\lambda]=\left[X^TX\right]^{-1}\left[X^T\right][Y] \tag{2}$$

where $[X^T]$ is the transpose of $[X]$. This solution was readily obtained using spreadsheet matrix operation commands. The baseline coefficient λ_0 is not a value but the price of the baseline product. If desired, the baseline value of a product can be determined from an analysis of the price and demand of the vehicles competing in the market segment and is roughly two times price.

The values calculated for the options are shown in Figure 6 for all participants, for truck versus non-truck owners, and for men versus women. The ± 95% confidence limits computed for each side of the median (assumed to be the same as for the average) are shown in Table 3. A significant difference exists between the value assessed by women for the two-tone paint and the extended cab options.

Table 3: Confidence Range (95%) on One Side of Mean

	All (57)	Own Truck (22)	Not Own truck (35)	Men (35)	Women (22)
Color	120	193	153	153	193
4x4 Drive	315	507	402	402	507
SportyTrim	120	193	153	153	193
Extended Cab	120	193	153	153	193

OPTION PRICING

The addition of an option to a product line can increase profits by the incremental profit generated per unit sold and by increasing the number of products sold. The price for the option which optimizes profits can be estimated from the value and variable cost of the option. To a first approximation, the price of the option is independent of the fixed costs for the option.

The total demand for the product (those with and those without the option) can be written as [13]:

$$D=K\left\{V-P-\frac{N-1}{N}\overline{[V-P]}_{Comp}\right\} \tag{3}$$

where K is the negative slope of the demand with price, V is the average value of the product, P is its average price, N is the total number of products competing, and $\overline{[V-P]}_{Comp}$ is the average value less demand for the N-1 competing products. The average value and price of the product are given respectively by:

$$V=V_0+fV_{Opt} \tag{4}$$

$$P=P_0+fP_{Opt} \tag{5}$$

The fraction of persons taking the option relative to the baseline product according to the linear model is given by:

$$f=0.5+\frac{3E_{Opt}}{8}\left[1-\frac{P_{Opt}}{V_{Opt}}\right] \tag{6}$$

which should be valid over the range of f from 0.25 to 0.75. The expression for the logit model is given by [6]:

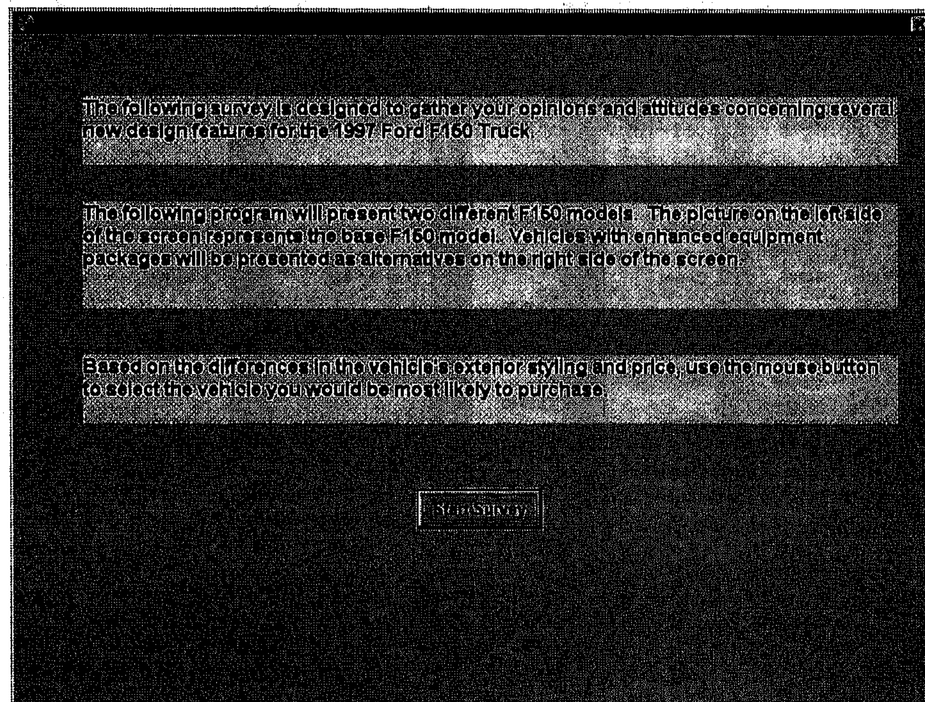


Figure 2. Instruction Interface.

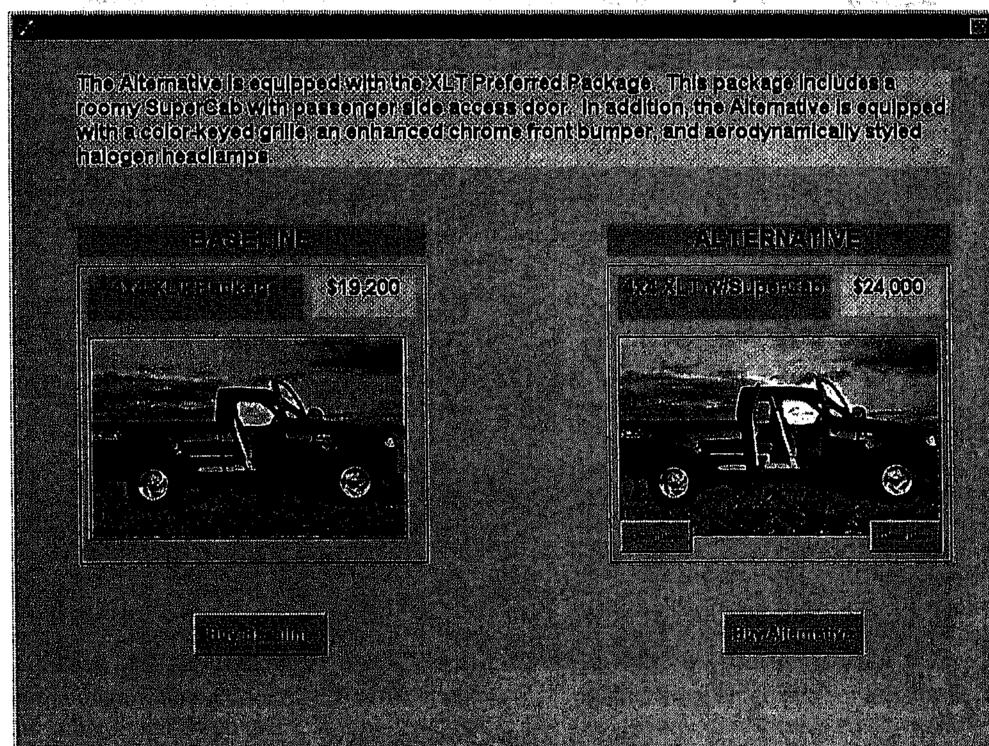


Figure 3. DVM Product Comparisons.

Please complete the following demographic questions.

GENDER
Female
Male

MARITAL STATUS
Single
Married
Divorced / Widowed

DO YOU OWN A TRUCK?
Yes
No

CHILDREN
None
One
Two
Three
Four
Five
Six or more

AGE
under 16
18-23
24-29
30-39
40-49
50-59
60-69
70+

EDUCATION
Some High School
High School Degree
Some College
Bachelor's Degree
Master's Degree
Advanced Degree
Other

HOUSEHOLD INCOME
under $25,000
$25,000 - $49,999
$50,000 - $74,999
$75,000 - $99,999
$100,000 - $124,999
$125,000 - $149,999
$150,000 or more

Finish

Figure 4. Demographic Interface.

Thank you for taking the time to complete the previous survey.

The information from this survey will be used to develop vehicles that better serve your needs.

Figure 5. Thank You Interface.

$$Ln\left(\frac{f}{1-f}\right)=\frac{3E_{Opt}}{2}\left[1-\frac{P_{Optt}}{V_{Opt}}\right] \quad (7)$$

The two models closely agree for f in the range 0.25 to 0.75. The term E_{Opt} is price elasticity for the option.

Gross total profit (GTP) equal to cash flow excluding charges for fixed cost and investments is given by:

$$A=D\{[1-f][P_0-C_0]+f[P_0+P_{Opt}-C_0-C_{Opt}]\} \quad (8)$$

The behavior of the GTP/GTP$_0$ versus the option price is shown in Figure 7 where GTP$_0$ is the gross total profit if the option were not offered. The ratio for the 4x4 Drive option has been computed for illustrative purposes. A family of curves are shown for variable costs over the range of 0.2 to 0.8 times the value of this option. For simplicity, it was assumed that competitors did not change the value or price of their products and the price elasticity of 1.9 for the 4x4 option in Table 4 was rounded off to 2.0.

The price elasticities shown in Table 4 were computed in two steps. First, the elasticities for each of the six alternatives were computed from the slopes of $Ln(f/(1-f))$ and f versus price for the logit and linear models, respectively, with E_{Alt} set equal to E_{Opt} in Eqs 6 and 7. The R^2 statistic ranged between 0.93 to 0.98 for these plots. The linear model slopes were computed from points lying within the range of f from 0.25 to 0.75. The elasticities for the individual options were then computed from the E_{Alt} using a linear elasticity model for the experimental design given by Table 1 , the elasticity for the baseline being set equal to the average of the six E_{Alt}. Except for the extended cab, the elasticities for each of the options are seen to be independent of the range used for the logit model and in agreement with those computed from the linear model over its range of applicability.

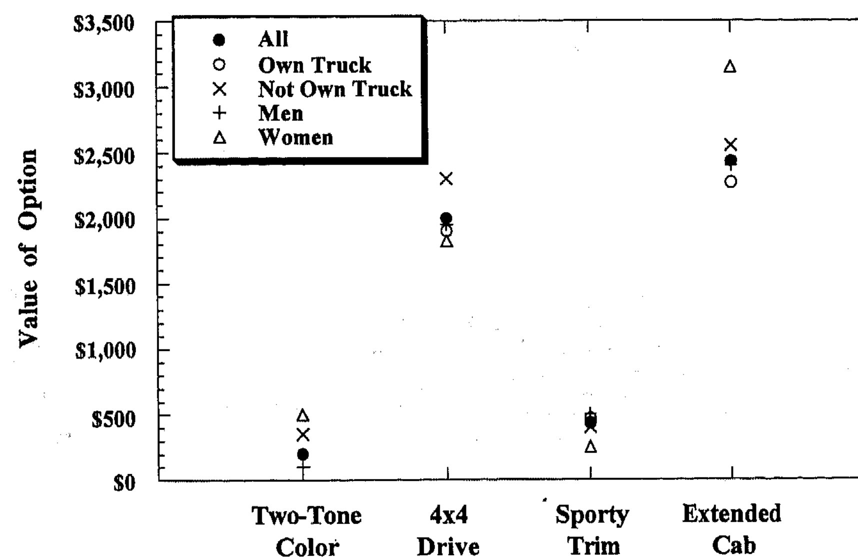


Figure 6: Values of Options

DISCUSSION

Our purpose here was to evaluate the use of a PC tool to assess the values of product options using the DV method. The tool was shown to be very useful for estimating the value of options such as the two-tone color and sporty trim whose appeal is primarily visual as well as options such as the 4x4 drive whose appeal is primarily functional. Although the extended cab option is also primarily functional, it also has significant impact on the overall appearance of the vehicle and thus has an important subjective component. We believe the PC tool, because of its ability to show color images, allowed both of these options to be properly considered by the respondents and in a manner better than could be achieved with a written survey. Of course, having respondents view and drive full scale prototypes would be superior but this approach is not feasible in the early design stages where only conceptual renderings of the product are available. The results obtained here confirm that the PC tool is capable of delivering a DV survey,

Of some surprise to us was the finding that persons in the local survey valued the options the same, within the statistical error of the experiment, independently of whether they did or did not own trucks. This suggests that deep product knowledge was not a factor in influencing the values of the options considered. This finding plus the fact that the different segments of respondents agreed that two-tone paint and sporty trim were of considerable less value than either the 4x4 drive or the extended cab supports the belief that stated WTP results are not capricious but reflect reasoned responses

which can be used for assessing the intrinsic value of product attributes.

If desired, the value of an option already in production can be determined from the average price of the option and the percentage of persons choosing the option provided that the price elasticity of the option is known. Using the prices and the number of options sold on the F150 truck for the 1995 model year (the latest results available to us), we also computed the market values for the same options, the results being shown in column 4 of Table 5. The last column, for comparison, shows the option values computed from the neutral prices determined from the logit plots of the WTP prices of all respondents. (These values differ slightly from the values computed from the median prices in Table 2. If the option elasticities are not known, we suggest using 2 as an estimate of the elasticity.) The results of the comparison give further support to the use of WTP to determine the value of an option. Nevertheless, the results here do not necessarily reflect an assessment of the average value of these options for the national market because of the local nature of the survey.

The forecasted profitability calculations shown in Figure 7 for the 4x4 Drive option were made, as already stated, by assuming for simplicity that competitors did not make any price or value changes. Because of this, the gross added profits shown for placing this option in production are a considerable overestimate.

If the 4x4 Drive were truly a new option which was introduced by one manufacturer ahead of its competitors, the latter would likely react over the near term by lowering their prices once they saw its impact on the demand for their products and develop their own 4x4 Drive over the longer term because, as can be seen from Eq. 3, the advantage gained by a net increase in value minus price resulting from the introduction of a new product feature can be offset by a reduction in price of the competing products. However, competitors will generally not lower prices to totally offset the feature advantage because of the fear of starting a price war. Likewise, the manufacturer with the feature or option advantage should offer it at a fair price rather than at a low price because such action could start a price war [13], the fair price being roughly the sum of the option's variable cost plus value divided by two.

SUMMARY

- The PC tool worked well in estimating the value of visual as well as functional options.
- Respondents seemed comfortable with the PC as a medium for the survey.
- For the options surveyed, truck owners and non-owners valued the options approximately the same.
- Values of the options obtained from option price and option fraction of total sales agreed very well with the option values determined from survey.

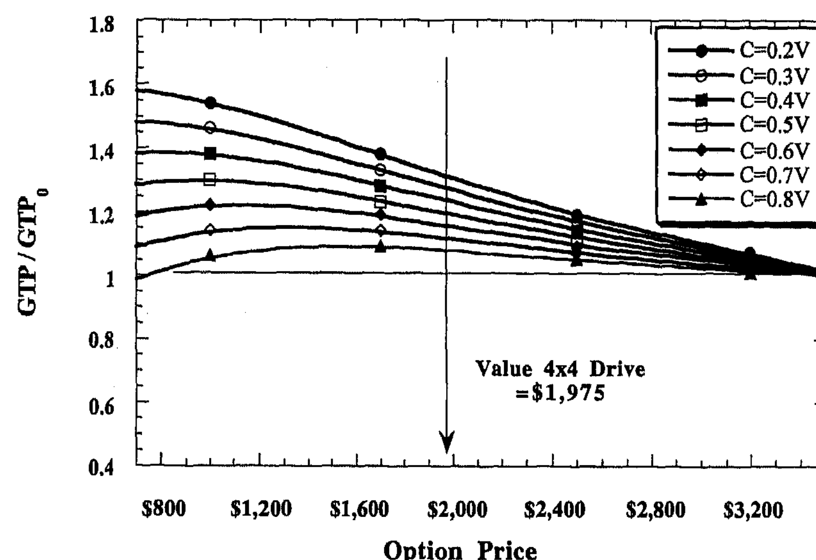


Figure 7 Idealized ratio of profits with and without 4x4 Drive option as a function of option price.

Table 4: Elasticities computed for the Individual Options.

	f = 0.05 to 0.95	f = 0.25 to 0.75	
	Logit Model	Logit Model	Linear Model
Color	0.7	0.8	0.8
4WD	1.9	1.8	1.8
Trim	1.3	1.5	1.5
Extnd Cab	4.0	3.2	3.0

Table 5: Estimated Price and Value for Options

	Est Price	1995 f	1995 Market Value	Survey Value
Color	$188	0.51	$191	$100
4WD	$1,828	0.35	$1,494	$1,950
Trim	$511	0.51	$519	$500
Extnd Cab	$2,480	0.459	$2,337	$2,400

ACKNOWLEDGEMENTS

The authors are deeply grateful to Ford Motor Company for support of this research through unrestricted grants.

REFERENCES

1. Cook, H.E. and DeVor, R.E. (1991) "On Competitive Manufacturing Enterprises I: The S--Model and the Theory of Quality," Manufacturing Review, **4**, 96-105.
2. Fowler, T. C. (1990) *Value Analysis in Design*, Van Nostrand Reinhold, New York.
3. Randall, A., B. C. Ives , and C. Eastman, (1974) "Bidding Games for Valuation of Aesthetic Environmental Improvements," *Journal of Environmental Economics and Management*, **1**, 132-149.
4. McConville, G. and Cook, H. E., (1996) "Estimating the Value Trade-off Between Automobile Acceleration Performance and Fuel Economy," SAE Paper 960004, Warrendale, PA, Society of Automotive Engineers, Inc.
5. Simek, M.E. and Cook, H. E., (1996) "A Methodology for Estimating the Value of Interior Room in Automobiles," SAE Paper 960002, Warrendale, PA, Society of Automotive Engineers, Inc.
6. Donndelinger, J. and Cook, H.E. (1995) "Methods for Analyzing the Value of Vehicles," SAE Paper 970762, Warrendale, PA, Society of Automotive Engineers, Inc.
7. McConville, G. and Cook, H. E., (1996) "Using Mail Surveys to Determine the Value of Vehicle Options," SAE Paper 970763, Warrendale, PA, Society of Automotive Engineers, Inc.
8. Green, P.E. and Ward, Y. (1975) "New Way to Measure Consumers' Judgments," *Harvard Business Review*, July-August, 107-117.
9. Sawtooth Technologies, 1007 Church Street, Evanston, IL.
10. Tversky, A. and Kahneman, D. (1981) The Framing of Decisions and Psychology of Choice, *Science* , **211**, 453-458.
11. Louviere, J. L. and G. Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, **20**, 350-357.
12. Ben-Akiva, M. and S. R. Lerman (1985), *Discrete Choice Analysis: Theory and Applications to Travel Demand*, The MIT Press, Cambridge, MA.
13. Cook, H. E. and Kolli, R. P. (1994) Using Value Benchmarking to Plan and Price New Products and Processes, *Manufacturing Review*, 7, 134-147.

970764

Evaluating Mail Survey Techniques for Determining the Value of Vehicle Options

G. P. McConville and H. E. Cook
University of Illinois at Urbana-Champaign

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

In developing new products and improving existing products, engineers make numerous trade-offs between the cost of a new or modified feature and its value to the customer. One method for estimating value is to ask potential customers their willingness to pay (WTP) for the product change. Their stated WTP may, however, depend upon how the question is framed. Mail survey techniques based upon simulated choice experiments were used for estimating value to the customer. The main objective was to explore how the framing of the survey questions affected the WTP response and if one or more of the methods provided simulated responses in reasonable agreement with actual buyer behavior. It was found that the best way to frame the questions was to give respondents multiple choices for price of the alternative versus the baseline product as opposed to having a choice of only one price or having to write in a price representing their WTP.

INTRODUCTION

Automotive engineers working in diverse areas of the product realization process -- manufacturing, vehicle system design, component design, assembly, and service – are routinely faced with trade-off decisions between value to the customer and cost. Value trade-off issues can also occur when cost is not a factor. A typical problem of this nature might involve the desire to lower the roof-line of a future vehicle to improve aerodynamics and visual appeal but it compromises head room. How much value is added by the better visual appeal and improved aerodynamics and how much value is lost by the reduction of head room?

In all cost/value and value/value trade-offs, the alternative chosen should be the one that is forecasted to give the highest profit or internal rate of return over the life of the vehicle. To use forecasted profitability as a basis for decision making requires that estimates be made of value to the customer and costs to the company for each alternative. Most companies have procedures in place to arrive at cost estimates and marketing research tools can be used to arrive at an estimate of value.

The purpose of this paper is to consider the use of mail survey techniques based upon simulated choice experiments for estimating value to the customer. Respondents were asked to state their willingness to pay (WTP) for an alternative vehicle (pay more or pay less) with the option installed versus a baseline vehicle without the option. Our main objective was to explore how the framing of the survey questions affected the WTP response and if one or more of the methods provided simulated responses in reasonable agreement with actual buyer behavior.

The market survey method which we used combined elements of conjoint analysis [1], contingent valuation [2], prospect theory [3], and choice theory [4]. The approach has been termed the Direct Value (DV) method [5] because it asks respondents, as in contingent valuation, their WTP in contrast to classical conjoint analysis where price is simply as another product attribute whose utility is to be determined. A constant baseline is used in the DV method because prospect theory [3] has shown that a fixed baseline is important in quantifying value differences between alternatives. A neutral price is determined from the responses to the survey where one-half of the respondents choose the baseline product and one-half choose the alternative. The value of the alternative relative to the baseline is equal to the neutral price less the baseline price. This value difference can, of course, be positive or negative. The method has been used, for example, to estimate the value of vehicle acceleration [6], the interior room of automobiles [7], and the value of certain attributes for mid-sized vehicles [8].

We are focusing on the value of options here because it allows us to test DV survey results against actual buyer behavior. The ideal vehicle of choice for this experiment would have the options available as separate choices, the buyer being unencumbered by having to take a package of options to obtain a desired single option. Unfortunately, none of the vehicles we reviewed were ideal. The Ford Mustang was chosen because it was simply the vehicle which more nearly approximated the ideal for this study. Buyers of the V6 Mustang could choose the anti-lock brakes (ABS) as an option without having to choose it as part of a package and , by comparing the make-up of different option packages, we were able to arrive at estimates of the price, if they had been offered separately, of the V8 engine (versus V6), ABS (across all models), convertible, automatic transmission, air-conditioning, and leather interior.

VALUE

The word value is used across many disciplines but it is seldom defined in the same way. As shown by Cook and DeVor [9], for example, value, or the lack thereof, is related in Taguchi's cost of inferior quality which when added to manufacturing costs yields his quality loss function. Value engineers on the other hand define value as function divided by cost [10]. This approach has been very helpful at making trade-offs at the component level of design. However, when trying to make cost/value trade-offs across subsystems, this definition gives different units for value depending on what performance variable is being measured. It also does not account for the distinct likelihood of diminishing returns on performance as it improves.

To overcome these limitations and link value in the engineering contexts to value and utility defined in other disciplines, we will use value as it appears in the linear demand expression by Cook and Kolli [11] because this gives an operational definition which is ideally suited to handling the problem of interest, namely forecasting demand and profits for product alternatives. In the absence of a price change, the change in value, δV, associated with the change in the product attribute g is equal to the change it produces in demand δD divided by the negative slope of the demand curve with price, K:

$$\delta V = \frac{\delta D}{K}. \tag{1}$$

This results in value having the units of price (dollars) independent of the system, subsystem, or component being considered. Moreover, based upon the definition of K Eq. 1, it follows that a change in value of δV will have the same effect on demand as a change in price of $-\delta P$.

There are two important ways in which this definition of value can provide information to support decision making in developing new products. Both are obtained through the application of the equation for demand written in terms of the value and price of the product i and the values and prices of competing products $j \neq i$ given by [11]:

$$D_i = K\left\{V_i - P_i - \frac{1}{N}\sum_{j \neq i}\left[V_j - P_j\right]\right\} \tag{2}$$

where N is the number of products competing in the market segment. This expression describes how demand for a product is influenced by its value and price and by the values and prices of the products competing against it. Because Eq. 2 is based upon a Taylor's expansion, the values and prices of the products should not be too different which limits the application of the expression generally to the top four to six competitors in the market segment of interest. When the demand and prices of the competing products are known, the values for the products can be solved for from Eq. 2.

When an option for a product is being considered in a simulated paired choice experiment, N=2 and Eq. 2 can be rewritten in the form:

$$f = 0.5 + \frac{3E_{Opt}}{8}\left[1 - \frac{P_{Opt}}{V_{Opt}}\right] \tag{3}$$

where f is the fraction of persons selecting the option and $1-f$ is the fraction selecting the baseline. The term V_{Opt} is the value of the option, P_{Opt} is the price of the option, and E_{Opt} is the price elasticity of demand for the option evaluated at f=0.5 and $P_{Opt} = V_{Opt}$. We see from Eq. 2 that a plot of f should be linear (in the vicinity of f=1/2) and that the option price at the point where f =0.5 is equal to the value of the option. The linear model given by Eq. 3 is expected to be accurate within the region $f = 0.5 \pm 0.25$ but become increasingly inaccurate outside this region. As a result, plots of f versus the option price can be used to determine the value of the option by setting it equal to the option price at which f=0.5.

The non-linear, logit model [12] for the logarithm of the ratio of the two fractions extends the range of applicability of the demand model at the expense of additional assumptions regarding buyer behavior and, for this problem, is given by [8]:

$$Ln\left(\frac{f}{1-f}\right) = \frac{3E_{Opt}}{2}\left[1 - \frac{P_{Opt}}{V_{Opt}}\right]. \tag{4}$$

The equivalent form for the linear expression for direct comparison to the logit model is given by:

$$4f - 2 = \frac{3E_{Opt}}{2}\left[1 - \frac{P_{Opt}}{V_{Opt}}\right]. \tag{5}$$

The two models are in close agreement in region $4f - 2 = 0 \pm 1$ When using Eq. 4 for determining the value of the option, plots of $Ln(f/[1-f])$ should be made as a function of the option price, the option value being the price at which $Ln(f/[1-f])$=0.

SURVEYS

Shown in Figure 1 are portions taken for entry number 9 from the three types of surveys used in determining the value of Mustang options. In the first survey, Figure 1a, respondents wrote-in a price; in the second, Figure 1b, they responded to a single price; and in the third, Figure 1c, they responded to a sequence of three of prices, a paired choice being made in each of the three lines. The write-in and one price formats, respectively, asked for and showed the full price for the vehicle alternative. The three prices format, however, showed the increase in price for the option versus the baseline.

A total of 5000 surveys were sent out in July 1995 to people who purchased 1995 Mustangs in various body styles and trim levels during March and April of 1995. The response rate for the one and three price versions was 27% versus 22% for the write-in version. The actual fractions of options purchased by respondents to the survey are shown in Table 1. The sample error in the points is within ± 0.07 (± two standard deviations).

9	Mustang with **3.8L V-6** engine $17,000	Mustang with **5.0L V-8** engine Maximum Price ________

Figure 1.a: Write-in Example

9	☐	Mustang with **3.8L V-6** engine $17,000	☐	Mustang with **5.0L V-8** engine **$19,000**

Figure 1.b: One-Price Example

9a	☐	Mustang with **3.8L V-6** engine $17,000	☐	Mustang with **5.0L V-8** engine **+$3,000**
9b	☐	Mustang with **3.8L V-6** engine $17,000	☐	Mustang with **5.0L V-8** engine **+$2,000**
9c	☐	Mustang with **3.8L V-6** engine $17,000	☐	Mustang with **5.0L V-8** engine **+$1,000**

Figure 1.c: Three Price Example

Figures 1a-c: Examples of three types of surveys used.

Table 1: Fraction of Options Actually Purchased by Respondents to the Surveys. (COBRA % also included in V-8.)

Option	One Price Set 1	Set 2	Set 3	Fill-In	Three Prices	Average
V-8	0.56	0.57	0.50	0.55	0.56	0.548
(COBRA)	0.06	0.06	0.06	0.05	0.04	0.053
Convertible	0.23	0.27	0.20	0.20	0.22	0.225
ABS	0.79	0.79	0.76	0.77	0.78	0.779
Auto. Trans.	0.47	0.46	0.51	0.51	0.42	0.475
Air	1.00	0.99	1.00	1.00	1.00	0.997
Leather	0.35	0.33	0.28	0.35	0.29	0.318

Graphs of f versus option price are shown as individual plots in Figure 2 for each of the six options considered. For a direct comparison, the cumulative results for the write-in respondents were tallied at the same three price points used in the three price version which were the same as the three prices used in the combined sets of the one price versions. For the V-8 and the convertible, the two highest priced options considered, the write-in and three price methods are seen to give close agreement but both are in poor agreement with the one price result because of its much lower price sensitivity. The intercepts of the lines with the horizontal line drawn at f=1/2 is the value of the option according to the linear model.

The largest discrepancies between the write-in and three prices version occurred for the automatic transmission and the air conditioner. Based upon the results from the three prices version, the automatic transmission was not viewed as a good in the economic sense as only 49% of its respondents said that they would take the automatic at no price increase over the manual. This is in close agreement with the 51.5 % respondents in the single price version who said they would take the automatic transmission at no price increase. But both are in marked disagreement with the write-in result of 89%. Part of this discrepancy (roughly one-half) results from a bias due to a statistical fluctuation in which 54% of the fill-in respondents selected the automatic transmission for their actual purchase; whereas, only 42% of the respondents in the three prices survey actually purchased the automatic, Table 1.

For options that are not considered as a good by a large fraction of the segment, there is little benefit in making plots of the type shown in Figure 2. The reason is that such plots return only one estimate of value for an option when there is clearly a bimodal distribution of those who will pay for it and those who will need to be paid to take it. With such options, the fraction of persons that view the option as a good should be determined and the WTP distribution should then be developed for them only. For options suspected as being goods for only a portion of respondents, one of the alternate prices for the option in the survey should be the same as the baseline. The fraction of persons, f_0, choosing the alternative at no price difference can then be used to generate a "corrected" fraction equal to f/f_0 representing the distribution of the population that views the option as a good. For the automatic transmission, dividing each of the three data points shown by 0.49 shifts the value intercept from 0 to $1,800, which represents its value to the segment that sees it as a good.

The air conditioner was not a true option here as people who might not otherwise want it would get it if they chose any option package because it was included in every package. If the write-in price intercept of $916 is used for the value of this option, we see that it is about as valuable as leather interior or ABS which we feel, by comparison, results in an underestimate of air conditioner value.

Although the above comparisons are helpful in judging the merits of the different surveys, comparisons with actual purchases of these options should be more enlightening. For this, we have plotted the point for the percentage of each type of option actually purchased in each of the graphs Figure 2. The actual percentages could also be determined from the survey as respondents were asked whether they had purchased each of the six options. The actual option price was determined from published pricing of the options. Of the three types of surveys, the three price version is seen to give the best agreement with the actual options purchased. However, the most rigorous test, as already stated above, is to compare the survey findings for ABS against the percentage of persons that actually chose the ABS option among those who purchased the V-6, Figure 3. The three prices method is seen to give the best agreement with actual purchase behavior.

OPTION VALUE COMPUTATIONS

In Table 2, we show the values of the six options computed from the intercepts of the best fit lines using the linear and logit models, Eqs. 2 and 3. The logit model values are shown first because they are preferred as several of the data points were outside the range of applicability of the linear model. This is most dramatic for the air conditioning where f for the lowest price is so near 100% that it is also very likely out of the range of applicability for the logit model. So only two data points were used for each model in making the extrapolations which means that the error in the result is large.

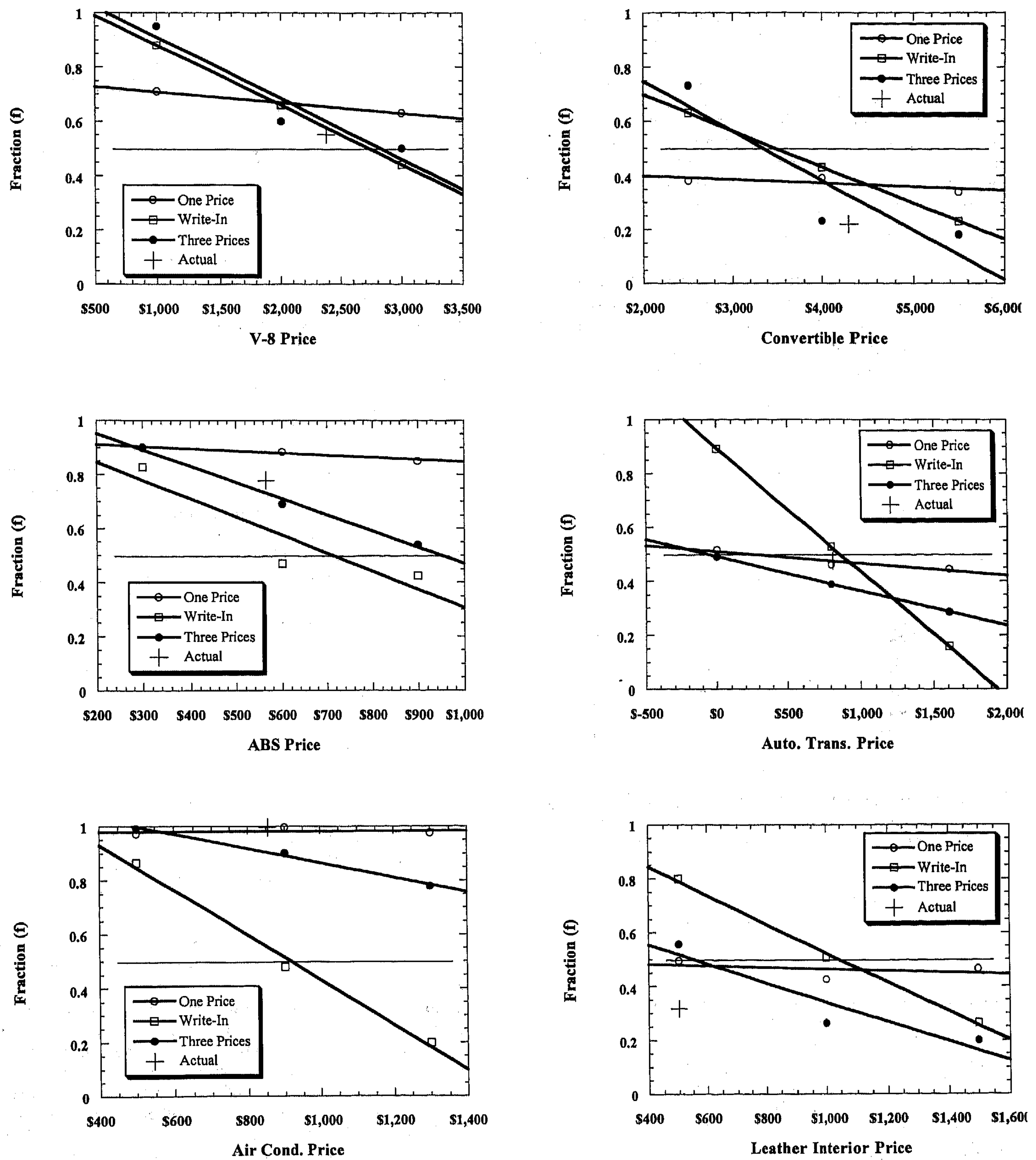


Figure 2: Linear Plots of the fraction of respondents selecting the alternative versus WTP price for the six options evaluated.

PRICE SENSITIVITY -- A plot of $Ln(f/[1-f])$ versus P_{Opt}/V_{Opt} according to the logit model, Eq. 4, is shown in Figure 4 for all of the three prices data. A line with a slope of 4 is seen to fit the scatter reasonably well in the positive region except for the highest point for the air conditioning option which was noted earlier as likely lying outside the range of both the linear and logit models. For a slope of 4, the option elasticity is seen from Eq. 4 to be equal to 2/3 of the slope or 2.7. It is not necessary that all of the options should have the same price elasticity. It is only roughly true for the options considered here. We note for comparison that on average the price elasticity E_1 of cars when only one manufacturer changes price is approximately 6 and the price elasticity E_2 when all change price is approximately 1 [13]. Because the alternative to buying the options in question is not to buy (a true substitute was not offered) we expect the elasticity to be around 2 (two times E_2 in theory). Thus the finding of 2.7 is not unreasonable. We will consider this point further in the discussion section.

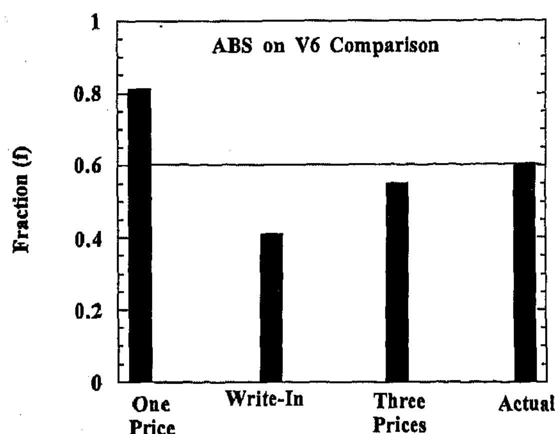


Figure 3: The fractions f at the purchase price for the ABS option as computed from the three WTP surveys are compared with the actual fraction of respondents who purchased the V6 and the ABS option.

Table 2: Option Values Determined from the Three Prices Survey for the Logit and Linear Models.

	Model	
Option	**Logit**	**Linear**
V-8	$2,751	$2,800
Convertible	$3,290	$3,360
ABS	$908	$948
Auto. Trans.	N/A	N/A
Air*	$1,826	$2,200
Leather Interior	$549	$552

* Two Data Points

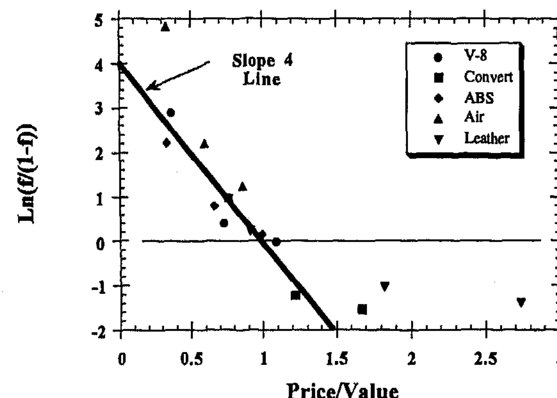


Figure 4: $Ln(f/[1-f])$ versus P_{Opt}/V_{Opt} according to the logit model, Eq. 4, as determined from the three prices survey results.

OTHER SURVEY QUESTIONS

OTHER VEHICLE ATTRIBUTES -- The respondents were also asked to force rank the ten vehicle attributes shown in Table 3 where the results are compared between V-8 and V-6 buyers. The results enclosed within borders are for attribute rankings which differed by more than two points between V-6 and V-8 buyers. The V-6 buyers ranked fuel economy and good price higher than V-8 buyers and acceleration performance lower which is not surprising.

Table 3: Attribute Rankings by V-8 and V-6 Buyers

Attribute	V-8	V-6
Styling	1	3
Reliability	3	2
Good Price	4	1
Fun to Drive	2	4
Ergonomics	6	5
Acceleration	5	9
Resale Value	7	8
Operating Costs	8	6
Fuel Economy	10	7
Interior room	9	10

OTHER VEHICLES CONSIDERED -- Use of the demand expression, Eq. 2, to solve for the values of competing products, requires that, N, the effective number of products competing against each other be known as well as their demands and prices. The names of other vehicles considered by Mustang buyers in making their purchase decision, as obtained from responses to the survey, are shown in Figure 5. The number of other vehicles considered is shown in Figure 6 and the number of other vehicles test driven is shown in Figure 7. The average number of vehicles considered computed from Figure 6 is 4.36 including the Mustang vehicle purchased.

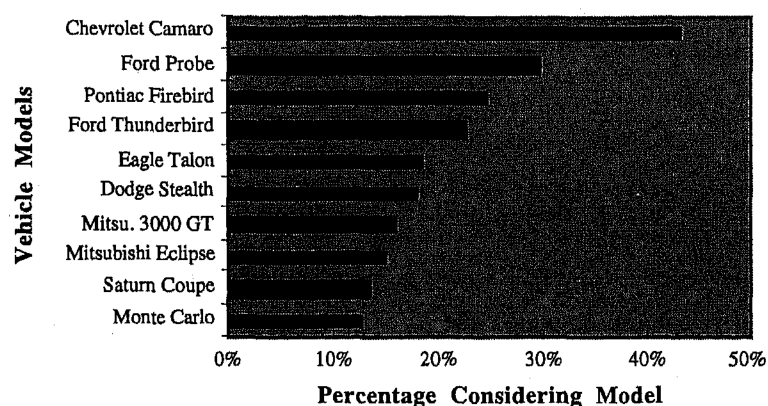


Figure 5: Other vehicles considered by respondents in making their purchase decision.

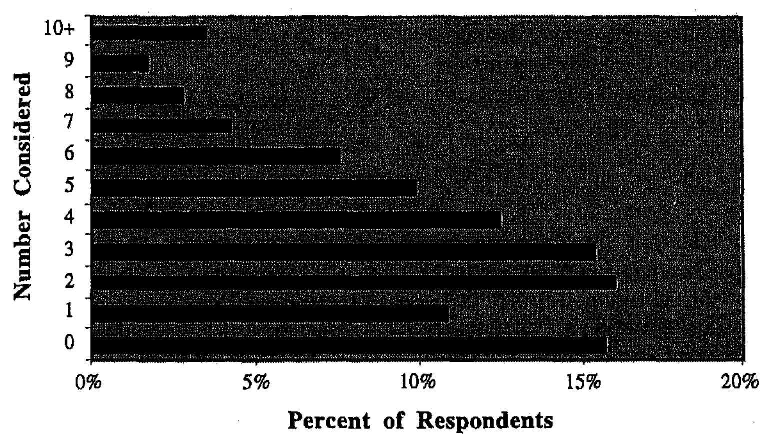


Figure 6: The number of other vehicles considered by respondents in making their purchase decision.

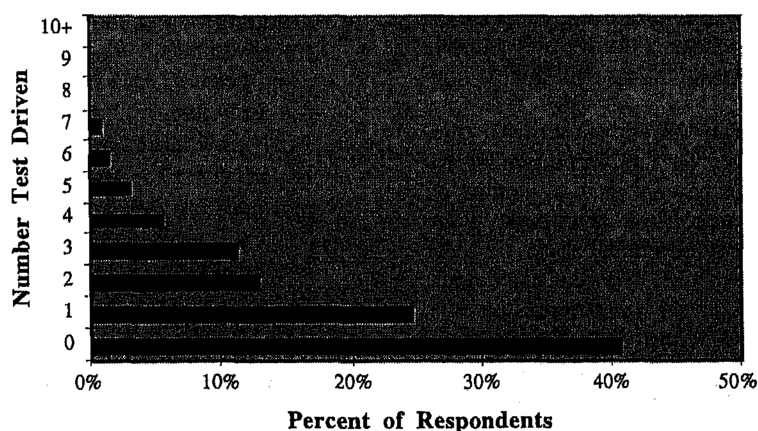


Figure 7: The number of other vehicles test driven by respondents in making their purchase decision.

PROFIT FORECASTING

The basis for making a decision regarding improving an attribute or an adding an installed option can be made, as stated earlier, on forecasted annual profitability given by:

$$A = D\left\{[1-f][P_0 - C_0] + f\left[P_0 + P_{Opt} - C_0 - C_{Opt}\right]\right\} \quad (6)$$

where C_0 and P_0 are the baseline cost and price of the vehicle in the absence of the option. The value and price expressions for substitution into Eq. 2 for demand D are given by:

$$V = V_0 + fV_{Opt} \quad (7)$$

and

$$P = P_0 + fP_{Opt} \quad (8)$$

where V_0 is the value of the vehicle without the option. The option fraction f according to the logit model is given by:

$$f = \frac{\exp\left\{\frac{3E_{Opt}}{2}\left[1-\frac{P_{Opt}}{V_{Opt}}\right]\right\}}{1+\exp\left\{\frac{3E_{Opt}}{2}\left[1-\frac{P_{Opt}}{V_{Opt}}\right]\right\}}. \qquad (9)$$

Substitution of Eqs. 7, 8, and 9 into Eq. 6, allows the forecasted profitability of the option to be computed as a function of its price and cost. The same expression can be used for an alternative such as a higher horsepower engine which would likely increase value due to improved acceleration with an offset for the loss of fuel economy. This combined value problem has been considered elsewhere [6]. A computation of the profitability of an option as a function of price has been given by Monroe and Cook [14] using the 4x4 drive option on a truck as an example for a range of costs.

DISCUSSION

The major finding here is that the mail survey having the sequence of three prices gave better agreement with actual buyer purchases for the Mustang options than either the write-in or single price surveys. The major concern with the single price survey results was that the fraction of persons selecting the alternative was approximately constant with price. A possible explanation for this behavior is that the choice between the baseline and any alternative considered as a good (i.e. excluding automatic transmission) was made by respondents to the one price surveys according to whether or not they had actually chosen the option when they purchased the vehicle. To fully account for the one price results using this line of reasoning, 10 to 40% of those persons who did not actually buy the option also had to select the alternative at the lowest price offered.

When respondents were asked to write-in the maximum price they would pay for the alternative or to judge the alternative versus the baseline across a range of three prices for the alternative, they exhibited price sensitivity of approximately the same amount as can be seen in Figure 2. We have already argued that the elasticities determined from the three prices and the write-in surveys were reasonable. The greater price sensitivity for the three prices surveys yields less statistical uncertainty in estimating the value of the option from the neutral price intercept. This is one more reason to prefer this approach. Unfortunately, we do not know the price elasticities of the options studied here from actual market behavior to make a direct comparison to the elasticities determined from our simulated buyer behavior.

Although we have concentrated on options in this study, the DV procedure as already noted can be used, in principle, to determine the value of any vehicle attribute. In For example, in our study of the value of vehicle acceleration each of the 86 respondents drove three vehicles having different acceleration performance through a city/suburban test route [6]. At the end of testing, they were asked their willingness to pay for the two higher performing vehicle versus the baseline vehicle.

CONCLUSIONS AND RECOMMENDATIONS

When respondents were asked to choose sequentially between a baseline Mustang at a fixed price and an alternative Mustang having an option added and offered at three different prices, their simulated purchase behavior was in much better agreement with the actual behavior of Mustang buyers than when a single price was given for the alternative or when the respondents were asked to write in their WTP for the option. Thus we recommend that the respondents in DV surveys be asked to make a series of choices between a baseline at a fixed price on the left and the alternative under consideration on the right as shown in Figure 1.c. Although Figure 1.c used three alternative prices, the desire to have the price range chosen for the alternative to bracket the neutral price for good statistics supports using a sequence of four or five prices for the alternative. To counter possible anchoring effects, the price sequence for one-half of the surveys should be decreasing and the other half should be increasing.

ACKNOWLEDGEMENTS

The authors are deeply grateful to Ford Motor Company for support of this research through unrestricted grants.

REFERENCES

1. Green, P.E. and Ward, Y. (1975) "New Way to Measure Consumers' Judgments," *Harvard Business Review*, July-August, 107-117.
2. Tversky, A. and Kahneman, D. (1981) "The Framing of Decisions and Psychology of Choice," *Science*, **211**, 453-458.
3. Randall, A., Ives, B. C. and Eastman, C. (1974) "Bidding Games for Valuation of Aesthetic Environmental Improvements," *Journal of Environmental Economics and Management*, **1**, 132-149.
4. Louviere, J. L. and Woodworth, G. (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, **20**, 350-357.
5. Cook, H.E. (1992) "New Avenues to Total Quality Management," *Manufacturing Review*, **5**, 284-292.
6. McConville, G. and Cook, H. E., (1996) "Estimating the Value Trade-off Between Automobile Acceleration Performance and Fuel Economy," SAE Paper 960004, Warrendale, PA, Society of Automotive Engineers, Inc.
7. Simek, M.E. and Cook, H. E., (1996) "A Methodology for Estimating the Value of Interior Room in Automobiles," SAE Paper 960002, Warrendale, PA, Society of Automotive Engineers, Inc.
8. Donndelinger, J. and Cook, H.E. (1995) "Methods for Analyzing the Value of Automobiles," SAE Paper 970762, Warrendale, PA, Society of Automotive Engineers, Inc.
9. Cook, H.E. and DeVor, R.E. (1991) "On Competitive Manufacturing Enterprises and the Theory of Quality," *Manufacturing Review*, **106**, 96-105.
10. Fowler, T. C. (1990) *Value Analysis in Design*, Van Nostrand Reinhold, New York.

11. Cook, H. E. and Kolli, R. P. (1994) Using Value Benchmarking to Plan and Price New Products and Processes, *Manufacturing Review*, 7 (2), 134-147.
12. Ben-Akiva, M. and Lerman, S. R. (1985), *Discrete Choice Analysis: Theory and Applications to Travel Demand*, The MIT Press, Cambridge, MA.
13. McCurdy, R., Ford Motor Company, *Personal Communication*, January, 1993.
14. Monroe, E.M. and Cook, H.E., (1996) "Determining the Value of Vehicle Attributes Using a PC Based Tool," SAE Paper 970764, Warrendale, PA, Society of Automotive Engineers, Inc.

970773

Advanced Finite Element Analysis in the Structural Design of Airbag Modules

Simon Xunnan He
AlliedSignal Safety Restraint Systems

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper demonstrates how to use advanced finite element modeling techniques such as airbag modeling, material plasticity with consideration of failure and effects of temperature and strain rates, surface to surface sliding contact treatment, nonlinear transient dynamics, and so on, to perform structural analysis of airbag modules. A technique to calculate the transient dynamic loads on the structure of an airbag module is first demonstrated. The dynamic transient loads on the structure are calculated by an integrated airbag modeling and sliding contact treatment with the structural model of the airbag module.

Other techniques for modeling special design features of the airbag module, such as tear seam, tear tie, and hinge on a cover, press fit of a gas diffuser, housing, connections of each component, and so on are also illustrated and discussed. All modeling techniques are illustrated by analysis examples in product development. Some tests have been done to verify the modeling techniques and good correlation has been achieved.

INTRODUCTION

The structure of an airbag module is subject to transient dynamic loads during airbag inflation and deployment. When finite element analysis is applied to assist the structural design, the unknown dynamic loads on the structure, the complex contacts among the components, the large deformation, material yielding and failure, and the effects of temperature and strain rates to the materials make the analysis very challenging.

A passenger or driver airbag module normally contains an inflator, a gas diffuser, a bag (cushion), a housing with mounting brackets, a cover, and mechanical connections. A side impact airbag module may not have the cover and the diffuser. The inflator and the bag designs are determined by the performance requirements of the safety restraint systems. All other components are designed in the airbag module development that ensures the required airbag performance and meets the interior trim and packaging requirements.

The cover must meet a structural requirement that it must open quickly when the airbag starts to inflate and must stay with the airbag module during the airbag deployment. For the reduction of cost, a cover is usually made by thermoplastic material molded into one piece. A tear seam or tear tie and a hinge are designed on the cover to fulfill the opening function. They are normally designed by reducing the material thickness and adding ribs in these areas. When the inflator is fired and the bag is pushing the cover, the tear seam or tear tie must break first, then the cover opens and rotates along the designed hinge by material yielding at the hinge. No fracture at the hinge area is allowed because a broken piece of cover could threaten the safety of the occupants. The finite element analysis can significantly reduce the time in the material selection and the structural design of the cover, especially the tear seam/tie and the hinge.

The gas diffuser is designed to even the gas flow from the inflator into the bag. The gas diffuser of a passenger airbag is usually a separate piece from the inflator. Due to a high gas pressure, some structural problems need consideration and analysis. For example, for a cylindrical type of diffuser design with a press fit onto the inflator, the finite element analysis can assist to find the thickness of the cylinder and the interference of the press fit that will ensure the integrity of the diffuser with the inflator during the assembly and inflator firing.

The housing not only provides a storage space for the inflator and the cushion, but also provides a structural support to the entire module and a mounting to the structure of an instrument panel. The complex connections with other components make the structural analysis very challenging, especially for the determination of the loads and restraints as well as the plastic deformation and failure.

The challenges in the finite element analyses of all those structures lie in the calculation of the transient dynamic loads, the prediction of material and structural

failure, and the modeling of complex contact between components.

METHODOLOGY

The first challenge in the finite element analysis of an airbag module is how to determine the loads on the structures. It is conceptually clear that all airbag components are subject to a transient dynamic loading during airbag inflation and deployment. However, it is difficult to determine these loads prior to some prototype tests. For example, an experimental study was done trying to catch the pressure loading on a passenger airbag cover during bag inflation [1]. This is usually costly in the early stage of product development and only applicable to the similar cover designs with the same inflator.

Analytical method to calculate the dynamic loads on airbag components during bag inflation is highly desirable and is possible with advanced modeling techniques of airbag inflation and contact interaction in some FEA codes. For example, the transient dynamic pressure on the cover and the housing during the airbag inflation and deployment can be obtained by integrating an airbag model into a structural finite element model of the airbag module. The load source comes from the airbag inflation in the airbag model, and the pressure loads on the components are calculated from the contact interfaces defined between the airbag cushion and the components. Other types of transient dynamic loads, such as the load on the brackets, the pressure on the gas diffuser, etc. can also be calculated in the advanced finite element modeling in the similar methods.

Another challenge in the finite element analysis of an airbag module is how to model material plasticity and failure of airbag components, especially the cover. The tear seam or tie on the cover is modeled with membrane elements with plasticity and plastic strain failure criteria. A modeling feature of failed element deletion is used to simulate the fracture of the tear seam or tie. The hinge, another highly interested area in the cover, is modeled by shell elements with the same features as the tear seam. The large plastic deformation and possible fracture in the hinge area during the cover opening can be studied in this model. The effects of the temperature and strain rate on the material properties of thermoplastics, such as elastic modulus, yield stress, as well as the failure plastic strain are modeled based on test data and the formulations in Reference [2].

Connections, such as spot welts, rivets, bolts, are all modeled by discrete elements with post yield properties and stress or plastic strain failure criteria because they are also the possible failure areas and absorb some energy before the failure. Contact interfaces are defined among the connections of the components themselves, such as the cover to the housing. A contact interface is also defined between the cover and the instrument panel because the moment that the cover hits the IP may be the worst scenario for the hinge. When the entire model is analyzed, the time history of reaction forces at mounting points of the brackets can also be used as the dynamic loading for the IP structure.

The analysis can be done in one FEA run or by multiple step runs to save run time. For example, in the study of the cover, after the cover is fully opened and no longer in contact with the bag cushion, the airbag model and contact interfaces related to the cushion can be deleted from the entire model in a restart run to save run time. All time history of dynamic loads on each component generated from the contact interfaces and connections can be used in separate component analyses for design optimization.

MODELING PROCEDURE, NUMERICAL STUDIES AND TEST CORRELATION

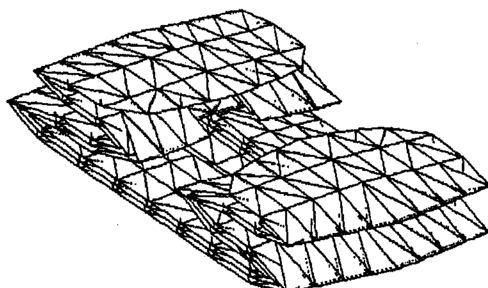
This section is to discuss in details about the finite element modeling of key structures of airbag modules. A modeling of transient dynamic loads induced by airbag inflation and deployment is described first.

MODELING OF THE TRANSIENT DYNAMIC LOAD INDUCED BY AIRBAG INFLATION AND DEPLOYMENT The most important modeling technique in the finite element analysis of the airbag module structures is the modeling of the transient dynamic load induced by the airbag inflation and deployment. The load source is the airbag inflation and deployment. The dynamic load is transmitted to each component through different path. Some are due to the direct gas pressure, such as the diffuser. Some are due to the momentum change of the cushion and the cover. The most are through the cushion impact under the gas pressure. For those subjected to direct gas pressure, analytical methods with a close form can probably be found to calculate the pressure history during the inflation. An example is a gas diffuser analysis that follows. For those subjected to the momentum change of the cushion and the cover as well as the impact of the cushion, such as the cover, the housing and brackets, a method that integrates an airbag finite element model and a structural finite element model of an airbag module can be used to calculate the dynamic load.

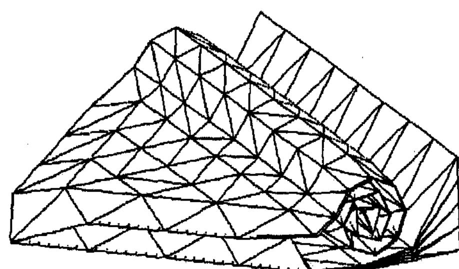
In the past years, many studies have been done in the airbag modeling [3] [4] [5]. All of them are for airbag performance analysis in the occupant protection. Sophisticate modeling features such as math models of inflation, cushion materials and elements, as well as the single contact interface are available to model airbags. With a modeling feature of automatic surface to surface contact interface that several FEA codes provide, it is possible to integrate an airbag finite element model into the structural finite element model of the airbag module and calculate the dynamic load on the structures induced by the airbag inflation and deployment.

During the airbag inflation, the gas puts the direct pressure on the components forming the gas flow path, such as the diffuser and cushion. Then the motion of these components transmits the load to other components through the impact contact, such as the contacts of cushion to the housing, to the cover, etc. Therefore, the contact interfaces play a very important role in the dynamic load calculation.

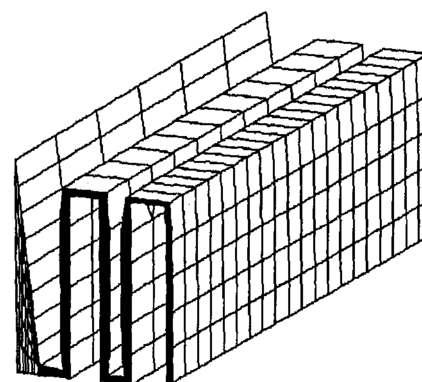
Airbag Modeling The airbag model used in the structural analysis of airbag module includes a mesh of the cushion, a math model of the inflation, and a rigid body of the inflator. The rigid body of the inflator is to provide the space, the mass and moment of inertia of the inflator. The cushion may be modeled as a directly folded mesh, a mapped and folded mesh, or a mapped and scaled mesh with a technique of initial matrix [6] or reference geometry [7] on the fully filled bag, as shown in Figure 1. The deference in the dynamic load calculation by the last two meshes is insignificant, while the mapped and scaled mesh is much easier to create. The gas jetting effect should be included in the inflation model because this jetting effect is significant on the dynamic load calculation. Figure 2 shows these differences on the resultant contact forces on the cover of a passenger airbag module in analysis.



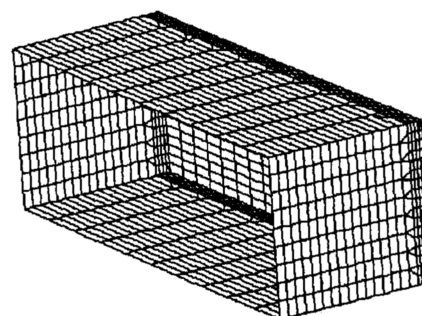
Directly folded cushion mesh
of a dirver airbag



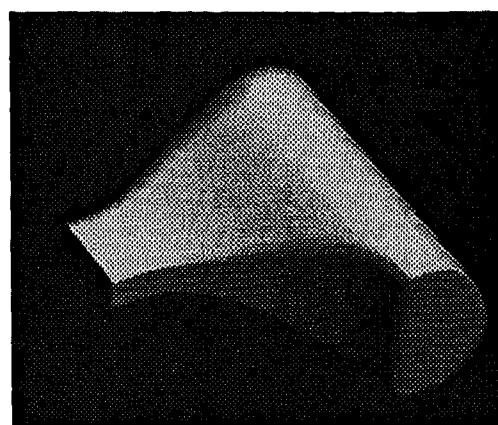
Directly folded cushion mesh
of a side impact airbag



Mapped and folded cushion mesh
of a passenger airbag



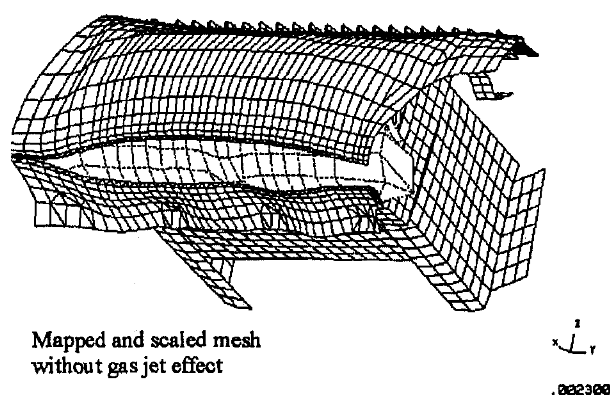
Mapped and scaled cushion mesh
of a passenger Airbag

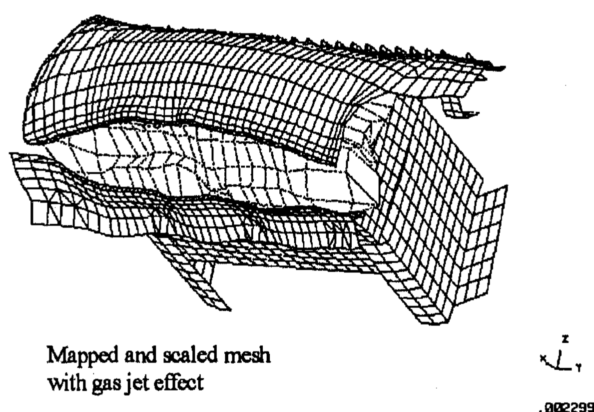


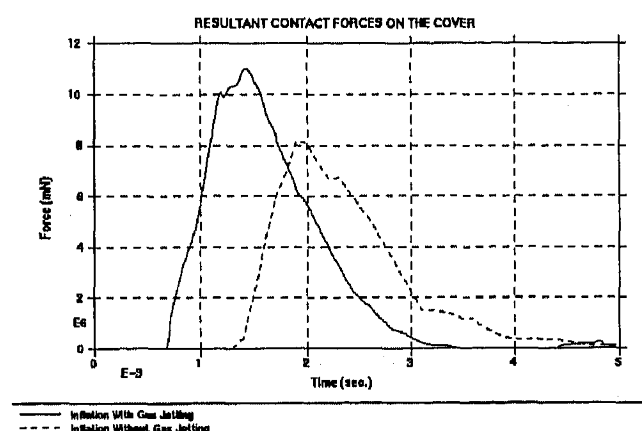
Reference geometry of a passenger airbag cushion

Figure 1. Finite Element Meshes of Airbag Cushions

Another type of dynamic load on the structures of airbag modules is due to the momentum change in high speed motion of components, such as the cushion and the cover. The maximum loads on the brackets, cushion fasteners, as well as the hinge of the cover come from this motion. The integrated structural FEA model with the airbag model can also model this type of dynamic load. A few examples are included in the following sections.







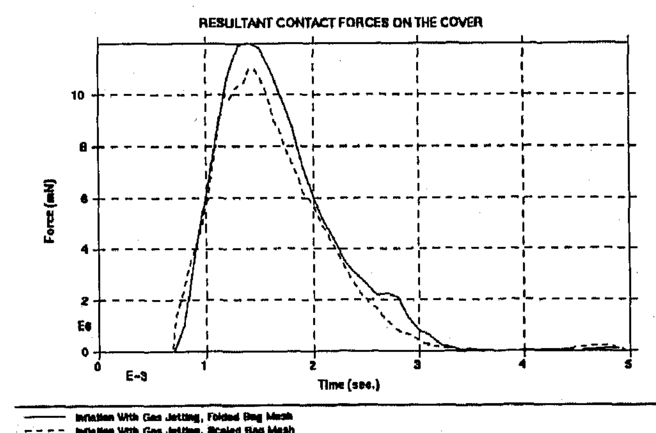


Figure 2. Effects of Defferent Airbag Models on the Dynamic Load on the Cover

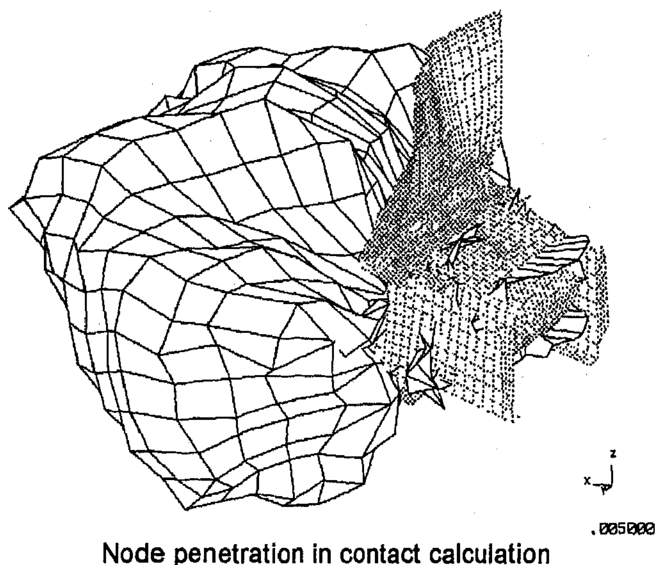
Contact Interface Definition During airbag inflation and deployment, multiple contacts between the cushion and other components as well as among other components themselves occur. These contacts play an important role in load transmission. Contact interfaces must be defined in the finite element model of an airbag module to handle these contacts. First of all, a contact interface between the airbag cushion and all surrounding components must be defined. If dynamic impact forces on particular portions of these components are of interest, several contact pairs that are exclusive can be defined and then the contact forces can be output for component analyses. For example, the impact forces on the front panel of the cover and on side panels of the housing can be used to do separate component analysis for design optimization. Then, for the meshes of airbag cushion, single surface contact should be defined to treat the cushion contact itself.

Contact interfaces among other components may need to be defined, such as the contact of the cover to instrument panel due to the large rotation of the cover that may be the worst scenario for the hinge of the cover, the contact of the housing to instrument panel due to the large plastic deformation of the side panels so called bell mouthing, as well as contact between the cover and the housing, and so on. It is impossible for a FEA code without a sophisticated contact treatment to perform accurate structural analysis for airbag modules.

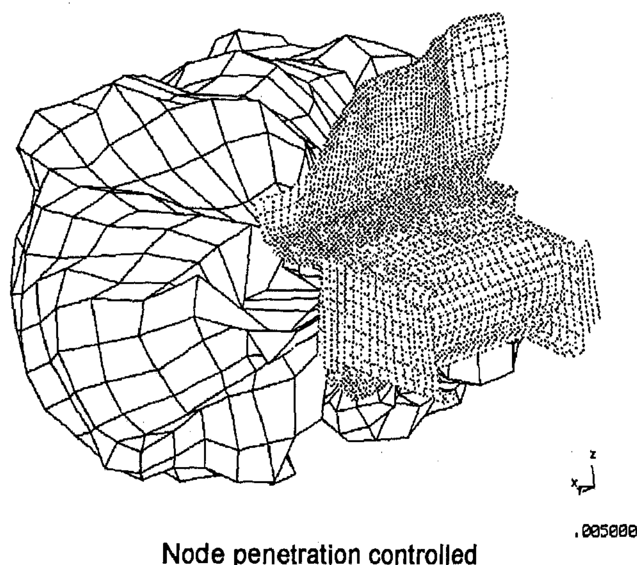
Even for a FEA code with contact modeling, it sometimes runs into trouble in implemention. In practice, a common problem in interface definitions is the unrealistic node penetration that is caused by a penalty contact algorithm used in some FEA codes. For example, in an automatic surface to surface contact of LS-DYNA3D, the airbag nodes could largely penetrate other components defined in the contact interfaces by the default parameters. The experience shows that in this type of contact definition this problem could be solved by adjusting the stiffness penalty factor of the cushion side so that its contact time step is in the same magnitude as that of the other contact side. Figure 3 shows a finite element analysis of an entire passenger airbag module by LS-DYNA3D in which large node penetrations occur between the cushion and the housing in the automatic surface to surface contact definition with default parameters. The time steps in both surfaces are 0.3E-4 second and 0.5E-5 second that have a difference of one order of magnitude. By adjusting the contact stiffness penalty factors to 20 in the cushion side to achieve a time step of 0.6E-5, the node penetrations are totally avoided, as shown in Figure 3.

GAS DIFFUSER MODELING A challenge in the finite element modeling of a gas diffuser is to find the structural behavior under the dynamic load due to the inflation with pre-stresses due to the assembly. An example is a modeling of a cylindrical press-fit type of airbag gas diffuser, as shown in Figure 4. The purpose of the analysis is to find a feasible interference tolerance

and the thickness of the cylinder to ensure that during the inflation the press fit can hold the diffuser and there is no structural failure and no gas leak through the press fit surfaces.



Node penetration in contact calculation



Node penetration controlled

Figure 3. Node Penetration in Contact Calculation and Penetration Control

During the press fit assembly, the diffuser body is pressed onto the inflator. That means that the diffuser cylinder is already pre-stressed when it is subject to the inflation pressure. The inflation pressure will further stress the diffuser and reduce the pressure in the press fit surfaces. If the pressure on the press-fit surfaces drops to zero, gas leak may occur. If the friction forces in the press-fit surfaces become smaller than the axial throttle force caused by the inflation pressure, the diffuser may be pushed out of the inflator.

Considering the thin shell structure of and the loads on the diffuser, the diffuser can be meshed by shell elements with material plasticity. The inflator can be modeled as a rigid surface because of the large stiffness ratio between the inflator and the diffuser. The press-fit is modeled by contact interfaces with interference.

The analysis is done by two steps. The first step is to calculate the stresses and plastic strains in the diffuser after the press-fit. Depending on the FEA codes, this can be done by defining an initial interference in the contact interface, such as in ABAQUS [8]. The initial interference simulates the upper and lower tolerance limits of the press fit. This can also be done by pushing a rigid shaft with a diameter of nominal value plus upper and lower tolerances into the diffuser in LA-DYNA3D. Any other FEA codes with similar features should also be able to do so.

The second step is the analysis of the diffuser under the inflation pressure starting with pre-stresses calculated from step one. This can be done by a restart run from the end of the first step. The dynamic pressure on the diffuser, P_d, during the inflation can be calculated by an analysis of thermodynamics in a control volume for the diffuser if the exit gas properties, mass flow rate $\dot{M}_{in}$ and temperature T_{in}, of the inflator are known. The governing equations for the calculation of pressure in the diffuser are

$$\left.\begin{aligned} \dot{P}_d &= \frac{Rk}{V_d}(T_{in}\dot{M}_{in} - T_d\dot{M}_{out}) \\ \dot{T}_d &= \frac{1}{M_d}[k(T_{in}\dot{M}_{in} - T_d\dot{M}_{out}) - T_d(\dot{M}_{in} - \dot{M}_{out})] \\ \dot{M}_d &= \dot{M}_{in} - \dot{M}_{out} \end{aligned}\right\} \quad (1)$$

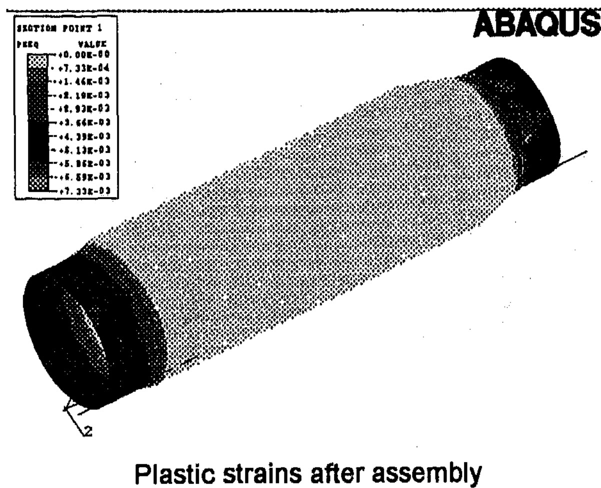
The details of the derivation of the equations and calculation of each term are in Appendix.

For the FEA codes with multiple chain airbag modeling features such as LS-DYNA3D, the dynamic pressure can also be calculated within the code by defining a chain of control volumes for the diffuser and the airbag with an appropriate inflation model. The additional pressure on the diffuser caused by the exit gas jetting from the inflator can also be estimated by considering gas momentum [5]. It should be pointed out that the gas flow will cause a pressure gradient inside the diffuser. The best method to get the dynamic pressure is to use computational fluid dynamics. For structural analysis, however, the method used in this paper with an estimation of gas jetting effect is appropriate.

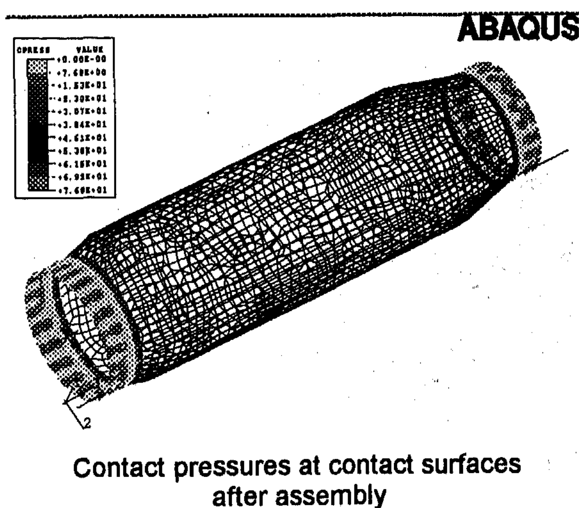
Figure 4 first shows the effective plastic strain and the press-fit pressure of the diffuser after the press fit assembly. The maximum press-fit pressures occur at the inboard circumferences of the press-fit area. The

analysis also found that for a feasible interference tolerance in manufacturing, the diffuser is already yielded at the press-fit area after the assembly. The result implies that the thickness of the diffuser is the only key factor to control the holding force and a loose tolerance can be used to reduce manufacturing cost.

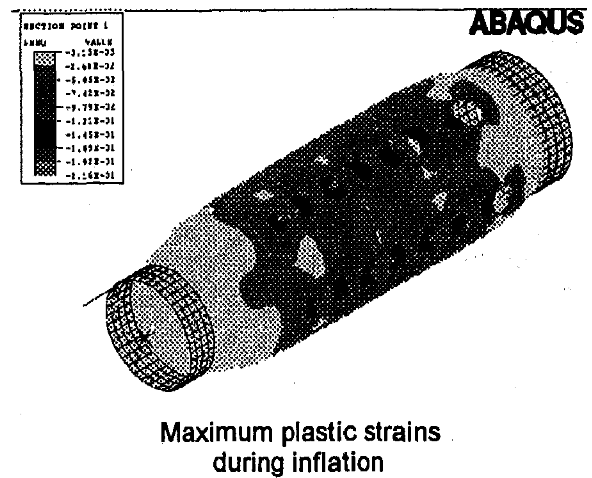
Figure 4 then shows the maximum effective plastic strain and the minimum press-fit pressure of the diffuser during the inflation for the same tolerance as above. It can be seen that the maximum plastic strains moved to the holes on the diffuser body and the remaining press-fit pressure shifted towards outboard circumferences and reduced by 65% for this case. The strength of the diffuser, the minimum holding force, and the leaking possibility can be judged from the results of this analysis. It should be noted that for the structural strength of the diffuser, the largest interference tolerance should be used to see the worst scenario in the structure, while for the minimum holding force and leaking concern, the least interference tolerance should be used.



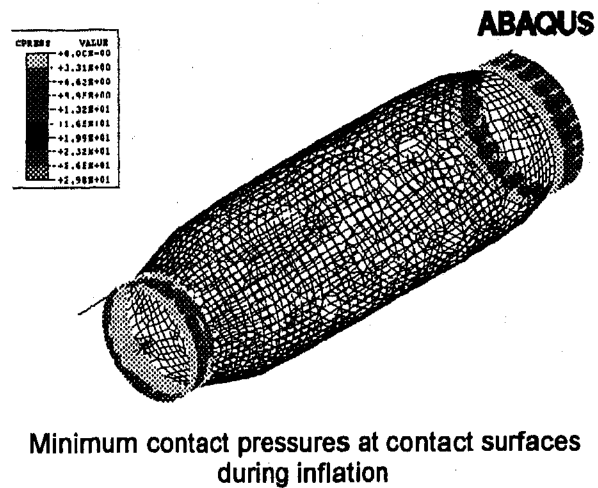
Plastic strains after assembly



Contact pressures at contact surfaces after assembly



Maximum plastic strains during inflation



Minimum contact pressures at contact surfaces during inflation

Figure 4. FEA Modeling of A Gas Diffuser in Assembly And in Airbag Inflation

HOUSING AND BRACKET MODELING An airbag housing plays a very important role in the airbag module. It not only provides a storage space for the inflator and cushion assembly, but also provides a structural support to all other components. For example, the cover can be mounted on the housing, the cushion and the diffuser can also be mounted on the housing in a passenger airbag module, and the entire airbag module is mounted on the IP frame through brackets on the housing. Permanent deformation is allowed in extreme deployment condition but fracture should be avoided. The complex dynamic loading and connections with other components and material plasticity form the characteristics of the housing modeling.

A housing and brackets are normally made by stamped sheet metal with spot welding or riveting, die cast non-ferrite metal, extruded non-ferrite metal, and even high temperature and high strength plastics. Shell elements with plasticity and failure are appropriate to model the housing structure. All mechanical connections such as spotwelds, rivets, bolts, and so on are modeled

with discrete elements or node constraints with post yield properties and failure criteria. Besides, contact interfaces are also needed to model the on-off boundaries between the connected components. The dynamic loads on the housing and brackets during the airbag deployment include the gas pressure, impact pressure from cushion, inertial forces due to the momentum change of the cushion and the cover, and so on. To calculate the dynamic loads, am integrated finite element model of the airbag and housing, brackets, as well as the cover is normally needed depending on the interests.

Figure 5 shows a finite element model of a passenger airbag housing and brackets with the cushion and the diffuser mounted on it (one side panel of the housing is not shown). The interest of this analysis is to find the strength of the structures and appropriate designs of the connections of the diffuser to the housing. Shell elements with material plasticity are used to model the housing. An airbag model is integrated in the model. The cushion is hooked on the diffuser that is connected to the housing. Contact interfaces are defined to model the hooking. The housing is restrained on the brackets that provides the mounting of the module to the IP frame.

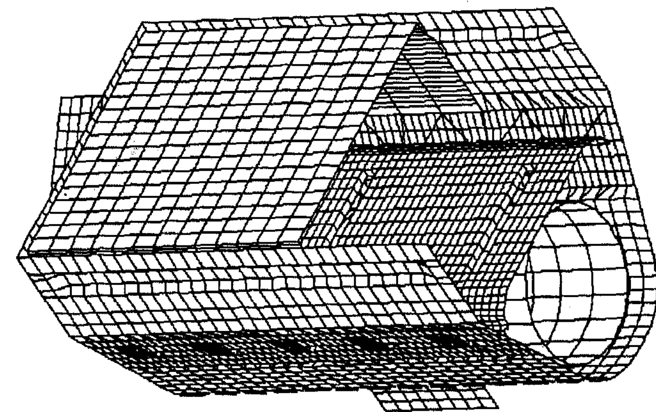
Figure 5 also shows the analysis results that the hooking failed and the diffuser and the cushion were separated from the housing during the airbag inflation. This was verified in tests and redesigns have been made.

Figure 6 is an analysis example of a driver airbag housing and bracket. Excessive plastic deformation of the housing and bracket may alter the facing direction of the bag during the deployment as shown in the figure. The dynamic load on the bracket is due to the change of momentum of the cushion during the deployment and can be calculated in the analysis. The reaction forces at the mounting points were measured in tests for the verification of the analysis model. The analysis results correlate with the test results quite well.

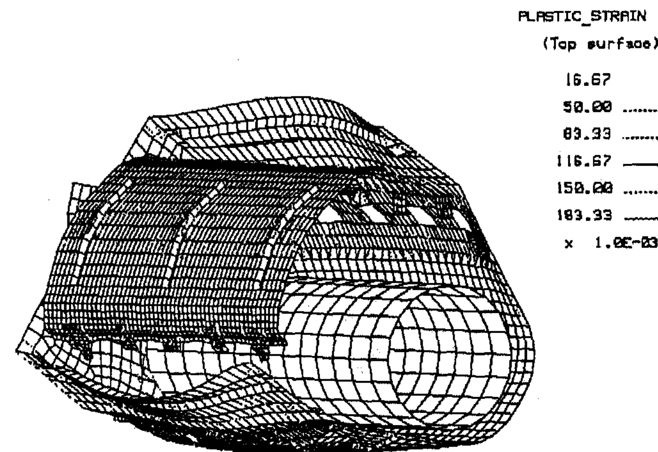
Figure 7 is another analysis example of a side impact airbag housing and bracket. The integrity of the housing, bracket, as well as connections are studied in the analysis. The load comes from the inflation and deployment of the airbag. To calculate the dynamic load, a finite element model of the side impact airbag is integrated in the structural model. Folded cushion is used to catch more accurate loading. The reaction forces at the mounting points were measured in tests for the verification of the analysis model. The analysis results are well correlated with the test results.

COVER MODELING Airbag covers are often made by thermoplastic material molded into one piece including two crucial features, a tear seam or tie and a hinge. The tear seam is designed to break during the airbag inflation. It must be the first place on the cover to break when the airbag starts inflation. It also needs to break at a designed pattern such that the cover opening will not partially block the deployment of the cushion. To prevent deployment delay, it should also absorb as less energy as possible during the breaking. The hinge is designed such that the material will yield along a line but not break during the airbag deployment. This will let the cover to open along the hinge axle and keep the cover staying together with the module. Any fracture in the hinge could be of significant danger to the safety of the occupants.

The effects of temperature and strain rates on the properties of the thermoplastic material are significant, as shown in Figure 8 for a typical material used in production and must be considered in the modeling. The cover structure can be well modeled by shell elements with plasticity and failure. Special attentions should be paid to the modeling of the tear seam, tear tie, and the hinge.

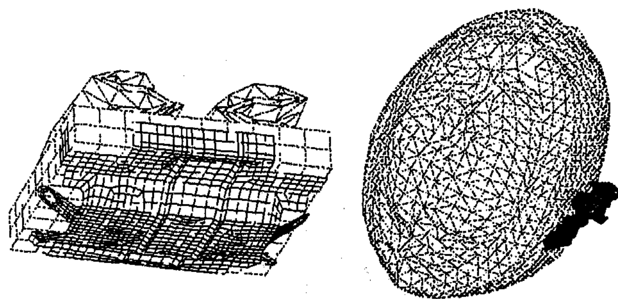


FEA model of a passenger airbag housing

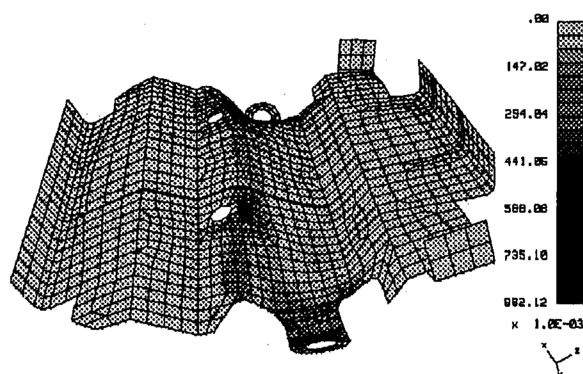


Failure of the connections at diffuser to housing during inflation

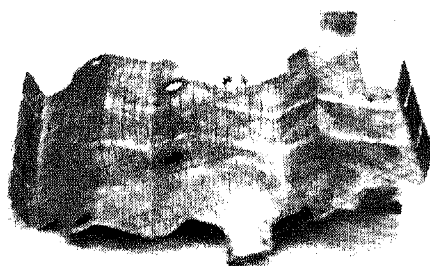
Figure 5. FEA Modeling of A Passenger Airbag Housing



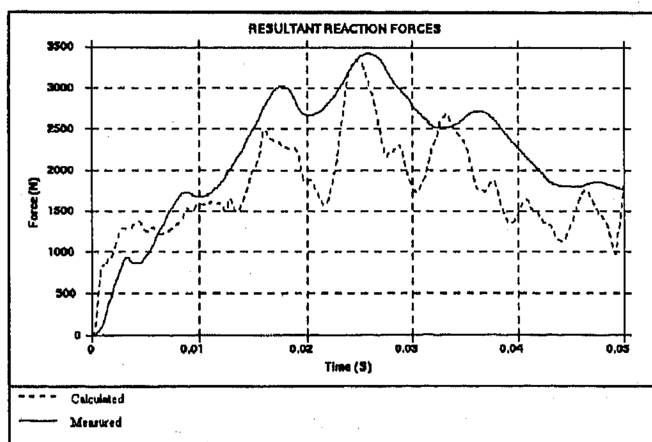
FEA model of a driver airbag module and the simulation of bag deployment



Plastic strain and permanent deformation of the bracket after airbag deployment

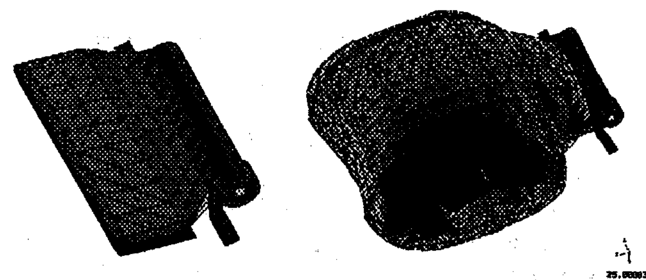


Permanent deformation of the bracket in test

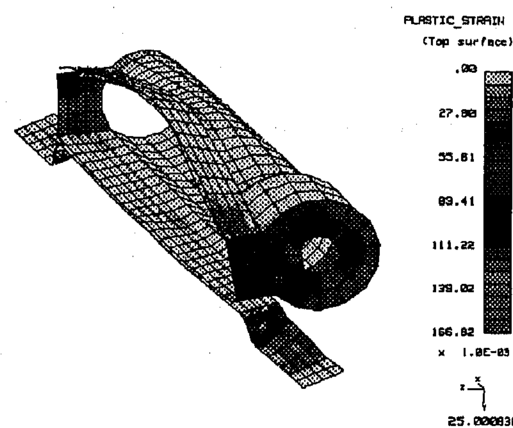


Comparison of reaction forces in analysis and test

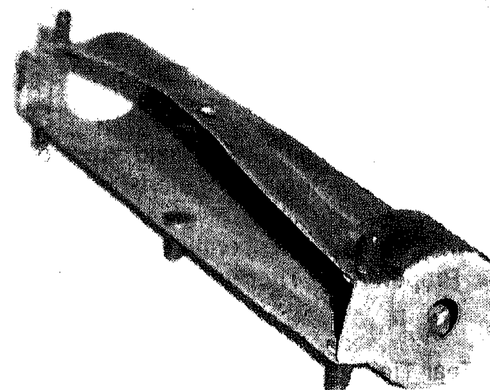
Figure 6. FEA Modeling of A Driver Airbag Bracket



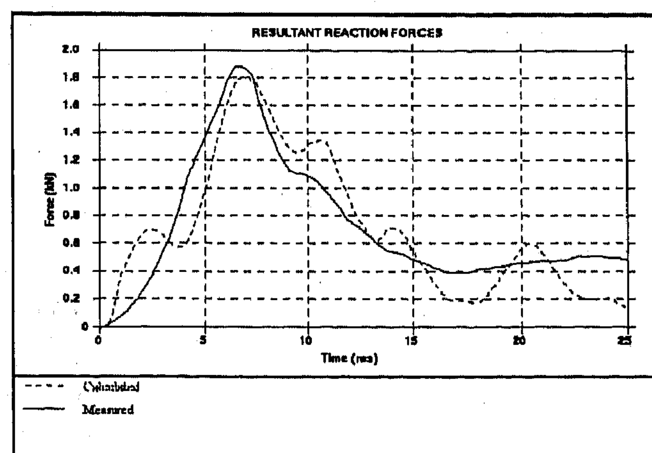
FEA model of a side airbag module and the simulation of bag deployment



Plastic strain and permanent deformation of the housing and bracket after airbag deployment

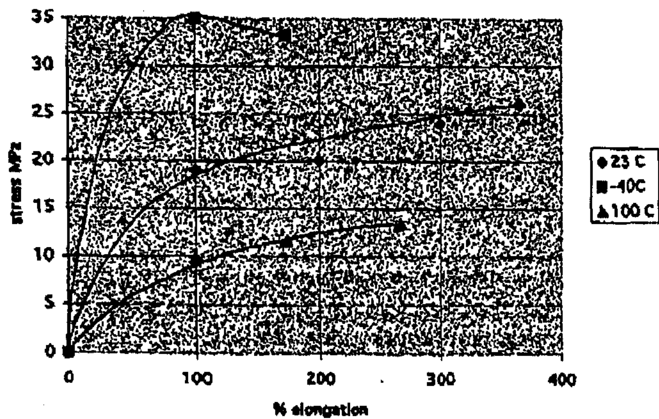


Permanent deformation of the housing in test

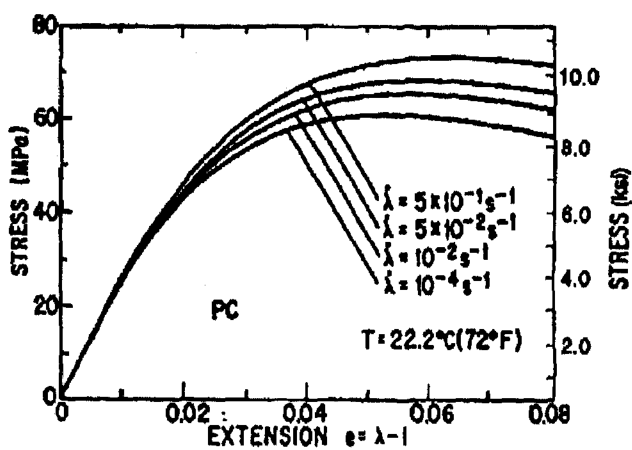


Comparison of reaction forces in analysis and test

Figure 7. FEA Modeling of A Side Airbag Module



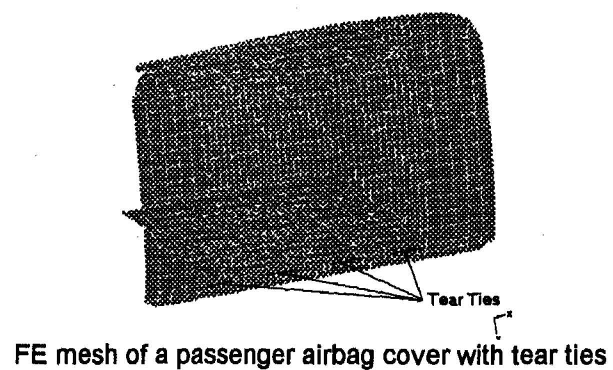
Effects of temperature on the elastic modulus, yield stress, and elongation of a thermoplastic material



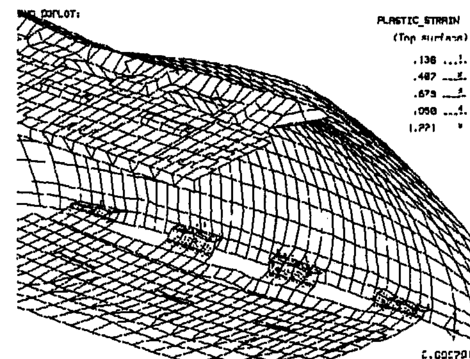
Effect of strain rate on yield stress of a thermoplastic material [2]

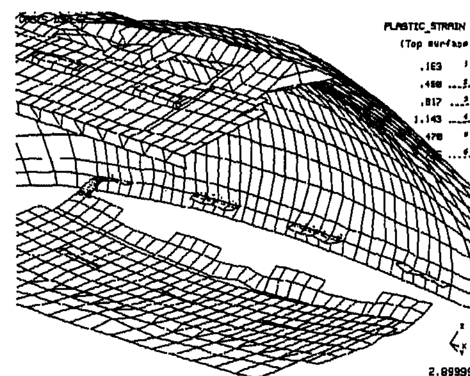
Figure 8. Effects of Temperature and Strain Rate on Properties of Thermoplastic Materials

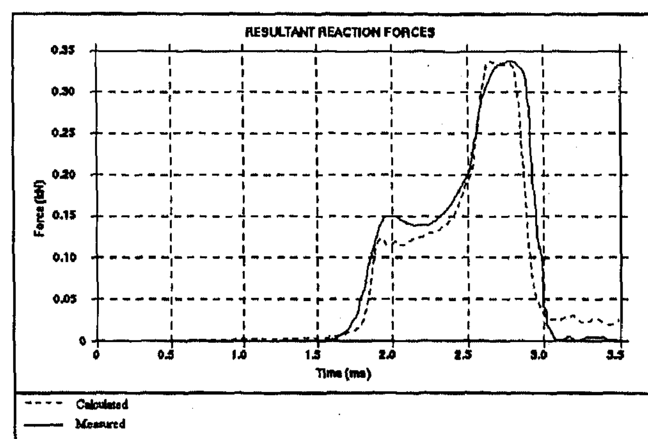
Tear Tie and Tear Seam Modeling From tests, the failure of the tear tie is found as stretching failure. It is also found that some permanent deformation exists after the fracture. This means that some energy has been consumed during the breaking. Since the thickness at the tear ties is several times thinner than that of the cover, it can be modeled with membrane elements with a plastic strain failure criterion. Figure 9 shows tear tie modeling in a cover design of a passenger airbag module and the element failure and deletion during the airbag inflation (cushion and housing are not shown). The dimensions follow the geometry and the thickness uses the minimum thickness. Note that the time unit is millisecond and the failure process lasts less than 1.0 millisecond. A test to verify the modeling was done by measuring the reaction forces at the two bottom plate restraints. Figure 9 also shows the calculated and measured reaction forces. They are well correlated. The reaction forces also indicate the sequences of the yielding and break of the tear ties and the total duration of the fracture.



FE mesh of a passenger airbag cover with tear ties


Elements of tear ties start to fail


Elements of tear ties totally fail



Comparison of reaction forces in analysis and test

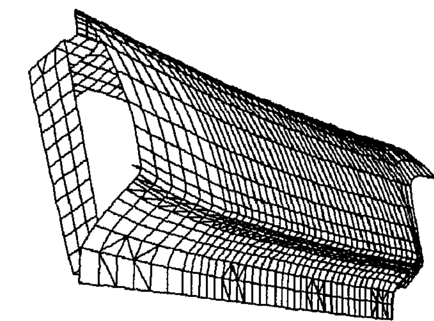
Figure 9 FEA Modeling of Tear Ties of A Passenger Airbag Cover

The failure of a tear seam is different from the tear ties. Before break, a tear seam is stretched as that of the tear ties. Once a fracture is initiated at a point (sometimes this point is particularly designed at a specific location), the failure propagates very rapidly along the entire tear seam. The best modeling method is the elements using fracture mechanics and adaptive meshing. The implementation of this technique in commercial FEA codes is still under development at present. To solve this problem, membrane elements with strain failure and a feature of failed element deletion are used to model tear seam failure. In this way, the first failed and deleted tear seam element transfers its load to the connected tear seam elements and these connected elements fail and are deleted again until all the tear seam elements fail and are deleted. This is similar to a crack propagation with an approximation.

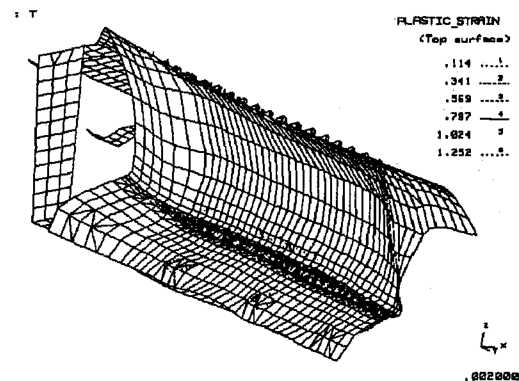
Figure 10 shows a horizontal bottom tear seam design. It is modeled by a series of membrane elements with a plastic strain failure criterion. The dimensions follow the geometry and the thickness uses the minimum thickness. Figure 10 also shows the initiation and propagation of the element failure and deletion along the tear seam during airbag inflation (airbag cushion not shown). The failure starts from the center of the tear seam and develops towards both ends. The effects of temperature and strain rates are found on the time of failure initiation and the duration of the failure propagation. The tear seam design modeled in Figure 10 has a non-symmetrical geometry and so the opening of the cover is also biased to one end. This was also observed in tests. A design modification has been done and a symmetrical opening is achieved.

The advantage of using membrane elements to model tear seams and tear ties is that during the failure, tear seams and tear ties absorb energy that is a part of the reason for the delay of the airbag deployment. This has been proved in many tests.

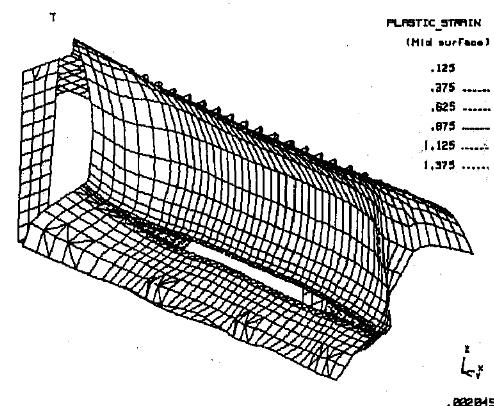
Hinge Modeling Hinge is often designed with a thinned material with some rib stiffeners to force a slender yielding area on the cover when the cover is opening. The tests show that the failure at this area is due to bending. Shell elements with a plastic failure criterion are used to model the hinge area, as shown in Figure 11 for a typical production design. The effects of temperature and strain rates on the yield stress, elastic modulus, and failure strain must be considered in the hinge modeling, because the function of the hinge will cover a wide temperature range from -30 °C to 80 °C. Under low temperature and higher strain rate thermoplastic materials behave more brittle. The hinge design shown in Figure 11 that only yields under room temperature fractures under a temperature of -30 °C, as shown in Figure 11 for the FEA results and a test result.



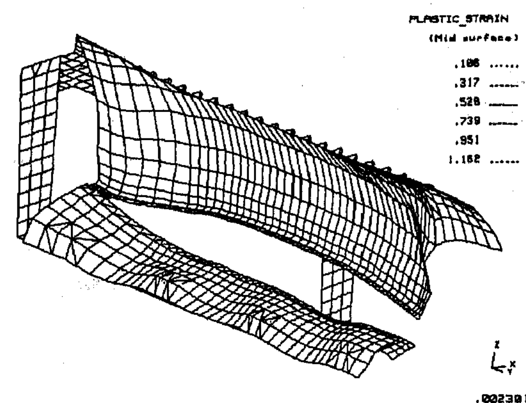
FE mesh of a passenger airbag cover with a tear seam



Elements of tear seam start to fail

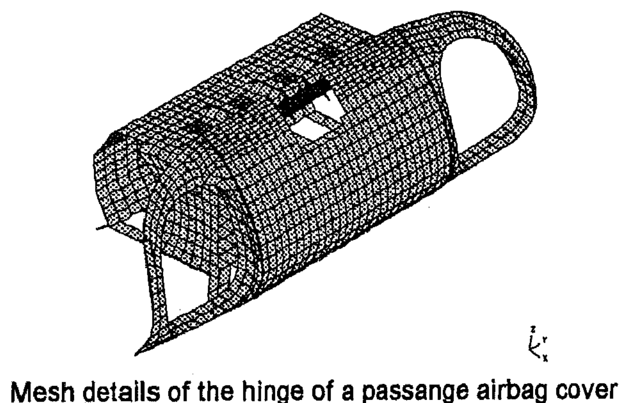


Elements of tear seam continue to fail

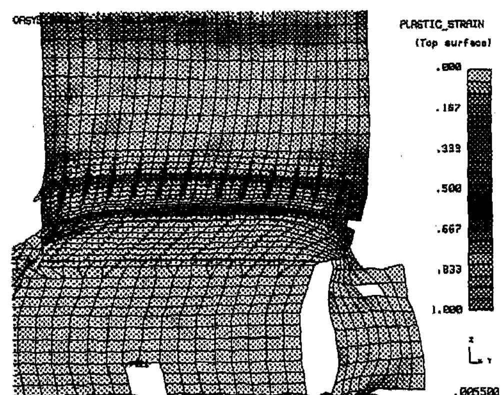


Elements of tear seam totally fail

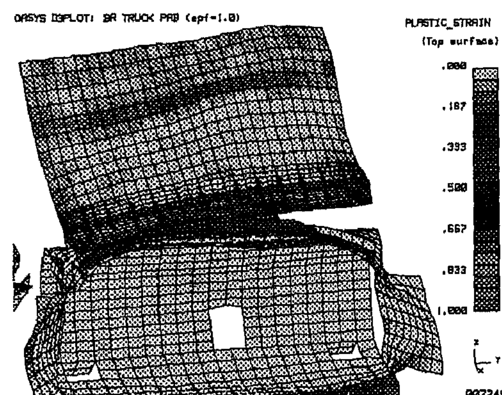
Figure 10. FEA Modeling of The Tear Seam of A Passenger Airbag Cover



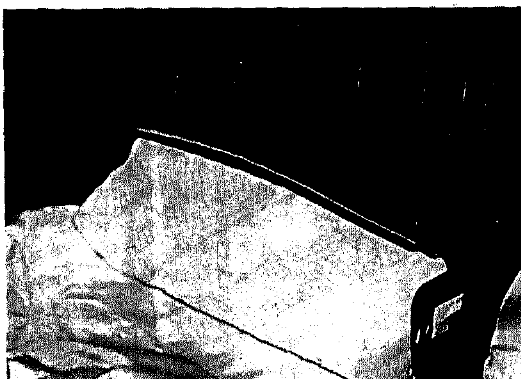
Mesh details of the hinge of a passange airbag cover



Hinge starts to fracture in a cold temperature



Hinge continues to fracture



Hinge fracture in a cold temperature test

Figure 11. FEA Modeling of The Hinge of A Passenger Airbag Cover

SUMMARY AND CONCLUSIONS

The challenges and solutions in finite element analysis of the structures of airbag modules are presented in this paper. The finite element modeling of the structures of airbag modules requires advanced finite element modeling techniques, such as airbag modeling, highly non-linear material property with consideration of effects of temperature, strain rates, and failure, capability to treat surface to surface sliding contact, multi-step restart runs, and so on. The modeling techniques of key components are also presented with test correlation.

The finite element analysis in structural design of airbag modules can be implemented by integrating an airbag model into the structural finite element model. The transient dynamic load induced by airbag inflation and deployment can be calculated from the airbag model and the contact algorithm during the analysis. To achieve accurate dynamic loads, the most sophisticate inflation model with jetting effect and carefully defined contact interface should be used. Mapped and scaled cushion mesh does not have significant deference from the folded cushion mesh. Most key design features, such as tear ties, tear seam, hinge, and so on can be evaluated with non-linear material property with failed element deletion. The strengths of all components, i.e., housing, brackets, mechanical connections, as well as cushion enforcement can all be predicted in the integrated finite element model.

REFERENCES

1. G. Bayley, et al., Experimental Measurements of Deploying Pressure Distribution on a Passenger-Side Airbag Door, SAE Paper 950339.
2. G. Trantina and R. Nimmer, Structural Analysis of Thermoplastic Components, McGraw-Hill, Inc. 1994.
3. J.T.Wang, A New CAL3D Airbag Inflation Model, SAE Paper 880654
4. T.B.Khalil, et al., Finite Element Simulation of Airbag Deployment and Interactions With an Occupant Model Using DYNA3D, SAE Paper 916144
5. H.A. Lupker and W.E.M. Bruijis, Gas Jet Model for Airbag Inflators, SAE Paper 930645
6. A. Tanavde,et al., Airbag Modeling Using Initial Metric Methodology, SAE Paper 950875
7. LSTC, LS-DYNA3D User's Manual, 1995
8. HKS, ABAQUS User's manual, 1995
9. Van Wylen, G. and Sonntag, R.E., Fundamentals of Classical Thermodynamics, John Wiley & Sons, 1978

APPENDIX DERIVATION OF GOVERNING EQUATIONS OF DIFFUSER PRESSURE

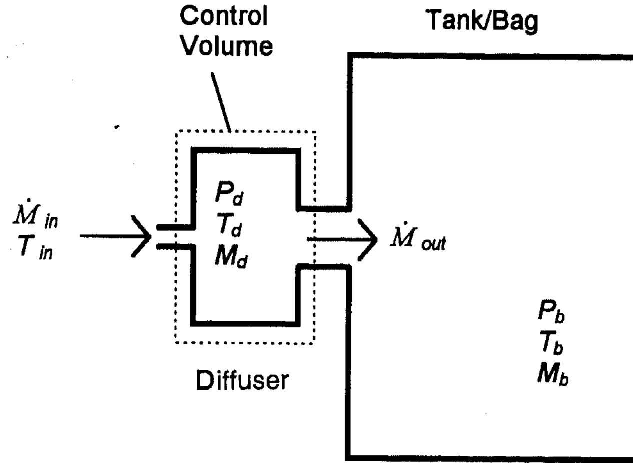


Figure A1. Thermodynamics Model to Calculate Diffuser Pressure

Assuming an isentropic process for the inflation, one has the following:

By the assumption of ideal gas property inside the diffuser

$$P_d V_d = R M_d T_d \quad \text{(A1)}$$

By the conservation of mass

$$\dot{M}_d = \dot{M}_{in} - \dot{M}_{out} \quad \text{(A2)}$$

By the conservation of energy

$$\frac{d}{dt}(M_d C_v T_d) = C_p T_{in} \dot{M}_{in} - C_p T_d \dot{M}_{out} \quad \text{(A3)}$$

Differentiating equation (A1) and rearranging equation (A3), we have

$$\begin{bmatrix} V_d & -RT_d & -RM_d \\ 0 & 1 & 0 \\ 0 & C_v T_d & C_v M_d \end{bmatrix} * \begin{Bmatrix} \dot{P}_d \\ \dot{M}_d \\ \dot{T}_d \end{Bmatrix} = \begin{Bmatrix} 0 \\ \dot{M}_{in} - \dot{M}_{out} \\ C_p T_{in} \dot{M}_{in} - C_p T_d \dot{M}_{out} \end{Bmatrix} \quad \text{(A4)}$$

This is the matrix form of simultaneous equations (1) in the text. By taking an inverse of the matrix at the left side, one has the simultaneous equations (1).

The mass flow rate out of the diffuser $\dot{M}_{out}$ can be calculated based on one dimensional orifice flow [9]

$$\dot{M}_{out} = CA \frac{P_d}{R\sqrt{T_d}} \left(\frac{P_b}{P_d}\right)^{\frac{1}{k}} \sqrt{2g_c\left(\frac{kR}{k-1}\right)\left(1-\left(\frac{P_b}{P_d}\right)^{\frac{k-1}{k}}\right)} \quad \text{(A5)}$$

where

$$\left(\frac{P_b}{P_d}\right) = \begin{cases} \dfrac{P_b}{P_d} & \text{if } \dfrac{P_b}{P_d} > \left(\dfrac{2}{k+1}\right)^{\frac{k}{k-1}} \\ \left(\dfrac{2}{k+1}\right)^{\frac{k}{k-1}} & \text{if } \dfrac{P_b}{P_d} < \left(\dfrac{2}{k+1}\right)^{\frac{k}{k-1}} \end{cases} \quad \text{(A6)}$$

970774

The Use of Magnetostrictive Sensors for Vehicle Safety Applications

Tony Gioutsos
Artistic Analytical Methods, Inc.

Hegeon Kwun
Southwest Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

New sensor approaches termed magnetostrictive sensor (MsS) and nonlinear harmonics (NLH) for vehicle safety applications such as crash detection and occupant seat weight sensing are described. Both sensor approaches utilize the changes in the magnetic properties of ferromagnetic materials that occur when a stress (or a strain) is applied to the material. The MsS is a passive device suitable for vehicle crash sensing. The NLH is an active device suitable for occupant seat weight sensing. Technical features of these sensors are also discussed together with preliminary results of ongoing testing.

1.0 INTRODUCTION

Vehicle safety sensing applications have increased dramatically with the increased use of airbag technology. Of the various vehicle safety sensing applications, there are three areas which have generally produced inadequate performance, they are: 1) single point crash detection for frame/body vehicles; 2) side impact detection and; 3) occupant seat weight sensing.

In this paper, we describe two sensor approaches termed "magnetostrictive sensor (MsS)" and "nonlinear harmonics" (NLH)" for the above mentioned application areas. Both sensor approaches utilize the changes in the magnetic properties of ferromagnetic materials that occur when a stress (or a strain) is applied to the material. The MsS [1,2] is a passive device suitable for vehicle crash sensing. The NLH [3] is an active device suitable for occupant seat weight sensing

In the following, an overview of the currently proposed sensor solutions and their limitations for the three application areas mentioned above is first given in Section 2. Then a description of the MsS and NLH is given in Section 3 including their physical principles, sensor design configurations, and their general properties and technical advantages. Specific applications of these sensor approaches to crash detection and occupant seat weight sensing are then given in Section 4 followed by conclusions in Section 5.

2.0 BACKGROUND ON VEHICLE SAFETY APPLICATIONS

The use of electronic accelerometer based single point sensing for vehicle airbag deployment has increased substantially in the last few years. Their use has been slower than anticipated but still climbing. An overview of this technology is given in [4] and [5]. A natural extension of their use is for other applications including side impact detection and frontal impact detection for frame/body vehicles.

2.1 FRAME/BODY VEHICLE CRASH SENSING - For many frame/body vehicles, acceleration values near zero can be encountered until the required Time-To-Fire (TTF) for a single point module located in the passenger compartment. This leaves even the best algorithm designers with no chance of achieving the desired requirements. A robust sensing concept capable of producing information at a faster rate from the passenger compartment is desired.

2.2 SIDE IMPACT DETECTION - Similarly, side impact sensing with accelerometer based approaches has produced at best marginal results. The inherent problem with an accelerometer based approach is the variation encountered for ON/OFF crashes across the side of the door. For example, a pole crash located at the B-pillar (assume an accelerometer is mounted at the B-pillar), will produce a substantially different sensor output than a similar crash at the A-pillar. Yet, the crashes to the occupant are similar in terms of potential injury. A detailed discussion of the disadvantages of crush zone sensing is given in [5].

There are other approaches to side airbag deployment as well (Autoliv [6] and Siemens [7]). In [6], a mechanical firing pin approach is described. This

sensor approach suffers from the same inherent weakness as an accelerometer based approach, namely, the large variation in sensor output based on location of the crash on the side of the vehicle. The idea behind both the accelerometer based approach and the mechanical based approach is to reduce this effect by creating an "array" sensor from a "point" sensor.

In both [5] and [8], the benefits of array sensors are described . In [8] a crush zone sensor is described. Unfortunately, this array sensor is truly an array causing other problems. In [5], frontal crash detection using a single point module approach is described as the best of both worlds (assuming an appropriate algorithm): array sensing from a point sensor. The front part of the vehicle acts as the array, the sensor simply produces similar waveforms for similar crashes (e.g. an offset pole crash on either side of the vehicle).

Therefore, it is beneficial to create an array sensor for side impact detection. By stiffening the vehicle or adding a crossbeam, the accelerometer/firing pin approaches begin to look more like a vehicle side array sensor. However, the inherent weaknesses still exists.

In [7], an air pressure approach is described that again expands on the notion of array sensing from a point. However, disadvantages of this approach include: variation over life, mounting in the door and the limited capability of the approach in sensing only those crashes impacting the door. Without dwelling on this concept, it does not provide the robustness that a magnetostrictive sensor approach can provide.

2.3 OCCUPANT WEIGHT SENSING - With the recent attention on fatalities due to airbag deployment, increased emphasis has been placed on passenger sensing. In particular, weight has been deemed the appropriate parameter for disabling the airbag. Several approaches have been proposed and are being proposed. Without getting too specific, several problems exist with current and proposed approaches including:

- Cost
- Implementation without disrupting the seat
- Temperature compensation
- Robustness to variability in seat occupancy scenarios
- Area coverage

These issues and potential solutions will be addressed further in section 4.

3.0 TECHNICAL BACKGROUND ON MsS AND NLH APPROACHES

3.1 MAGNETOSTRICTIVE SENSOR (MsS) - The MsS is a passive sensing approach which relies on a specific physical phenomenon, that exists in ferromagnetic materials, called "inverse magnetostrictive (or Villari)" effect [1,2]. The Villari effect refers to a change in magnetic induction (B) of material with application of stress, in comparison to the magnetostrictive (or Joule) effect which refers to changes in physical dimension of the material with magnetization. Being a passive device, the MsS is limited to detection of only time-varying or transient stresses (or strains) in the material such as those produced by mechanical impacts or those acoustic emission signals produced by cracking [1,2]. The MsS has a very broad frequency response, ranging from a few Hz to a few 100 kHz. The sensor can also be applied to nonferrous material such as plastics, if a thin layer of ferromagnetic material is plated or bonded to the material surface in a local area where the sensor is to be placed. In addition, the MsS requires neither a direct physical contact to the material nor a couplant for sensing.

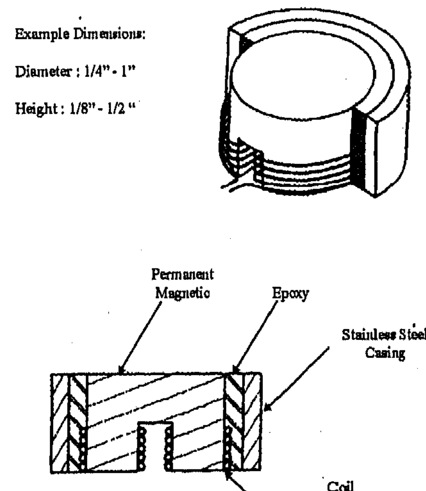


Figure 1 : Typical MsS Sensor

A typical MsS is depicted in Figure 1. It consists of an inductive coil and a permanent magnet. When stress in the material changes with time, the resulting changes in B due to the Villari effect induce an electrical voltage in the coil. The permanent magnet provides a static bias magnetic field to the material which will enhance the sensor sensitivity. The transient stresses produced by a crash or impact will propagate through the material. The attenuation of such impact stress waves in vehicles is negligibly small. Therefore, the MsS can sense the crash or impact event at a location far away from the exact impact site as long as the path

for the wave to propagate is good. The MsS can survey a large area and, thus, functions like an "array" sensor from a point and is ideally suited for vehicle crash detection.

3.2 NON-LINEAR HARMONIC SENSOR (NLH) - The NLH sensor operates in a similar fashion to the MsS, but it is an active device [3]. An example NLH sensor which consists of a U-shaped ferrite core with an excitation coil and a detection coil wound on each leg is depicted in figure 2. When a ferromagnetic material such as steel is excited by a sinusoidal magnetic field (H), the corresponding magnetic induction (B) of the material is no longer sinusoidal but shows a distorted waveform. This distortion is due to nonlinear magnetic permeability and the magnetic hysteresis of the material. The distorted B-waveform contains harmonic frequencies of the applied H. In addition, the magnetic hysteresis curves of a ferromagnetic material change significantly when the material is subjected to mechanical stress or strain. The NLH sensor transmits a sinusoidal H to a ferromagnetic material and then detects the resulting B waveform.

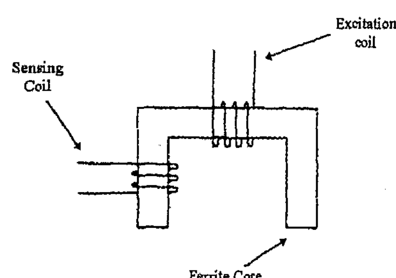


Figure 2 : Typical NLH Sensor

Typically, a coil is used for applying H and detecting B. The H is achieved by sending a sinusoidal current of a given frequency through an excitation coil (e.g. 1kHz). The resulting time-varying B induces an electrical voltage in the detection coil. The stress on the material of interest is determined by harmonic analysis of waveform. In general, the magnitude of the third harmonic of the resultant waveform is related to the stress on the material. In steel, up to about 50 percent of the yield strength, the relationship is linear (i.e. the greater the magnitude of the third harmonic the greater the stress).

The chief differences between the NLH sensor and the MsS are:

- The MsS is a passive device, whereas the NLH requires power
- The NLH sensor can sense constant stress

The NLH approach is similar to those approaches described as magnetostrictive sensors used for torque measurements [9] except that the NLH utilizes nonlinear harmonic components, whereas, the other utilizes the fundamental component.

3.3 TECHNICAL FEATURES - The two sensing approaches have the following features in common:

- Temperature insensitivity
- Non-contacting
- No moving parts
- Small and inexpensive

The MsS also has the following features:

- Passive (produces a signal without power)
- No DC offset (i.e. detects transient stresses)
- An "array" sensor from a point

The NLH also has the following features:

- Active
- Detects DC or constant stress

4.0 VEHICLE SAFETY APPLICATIONS

4.1 FRAME/BODY VEHICLE CRASH SENSING - The MsS can be used in the same fashion as an accelerometer for vehicle crash sensing [10]. Since the frontal crash sensing field is more mature, we will focus on the application of using a single point module for frame/body vehicles (e.g. trucks).

Because of the structure of frame/body vehicles, accelerometer signal values are near zero until the required TTF for passenger compartment mounting locations. This is due to the fact that the frame encounters the crash object well before the body has. By cutting a ferromagnetic hole (e.g. rubber or plastic) in the body above the frame and then facing the MsS at the frame, we are able to detect stress waveforms in the frame from the body. This will allow faster response from a single point sensor module located in the passenger compartment.

4.2 VEHICLE SIDE IMPACT DETECTION - The MsS is ideally suited for side impact detection. The device can be mounted on or near the door without necessarily being mounted on a surface. For example, one could mount the sensor on the crossbeam and face it at the door skin. Therefore, stress on the skin can be measured. Again, the "array" nature of the sensor will allow the sensor to produce waveforms even if the sensor is not contacted directly or is nearby the "hit" area. A good propagation path is all that is needed. There are other locations on or near the door that the

sensor can be placed to provide a good signal with ease of mounting. For example, placing the sensor on the sill facing the base of the door will allow the door to be sensed but the sensor to be mounted within the passenger compartment. Various isolated side impacts can be detected with this arrangement. If a pole hits the A or B pillar or is on-line with the sill mounting location, the signal strength should be similar for the same impact speed. This response should be differentiated from that of an accelerometer, where depending on the mounting location (e.g. A/B pillar location or the middle of the door) substantially different waveforms will be produced.

These two features (i.e., non-contacting and array sensing from a point) make the MsS ideally suited for this application.

4.3 OCCUPANT WEIGHT SEAT SENSING

The occupant seat weight sensing problem has become a very important issue in the automotive safety community. The key to the problem is to determine the occupant's or object's weight with a very high resolution. The goal is to change the airbag deployment characteristics (e.g. do not deploy) if the weight is less than a given threshold(s). There are many issues involving this approach but basically there are three main concerns: performance, cost, and ease of implementation.

The NLH sensor is suited to this application. By placing the sensor under the seat on either the pan, springs or mounting points; stress upon the seat can be measured. For example, for seat designs which feature an array of wires holding the seat cushion in place, stress on the wires can be measured as follows. As an object is placed in the seat the foam causes stress on the wires. The stress applied to the wire by the occupant is detected by using a NLH sensor placed on the wire. The NLH output can then be converted to weight. The sensor location, design and placement will be related to the given seat, but should not affect the seat's deign. This will allow a seat weight sensor that is inexpensive, robust, and easy to implement.

5.0 CONCLUSION

We have shown that the MsS and NLH sensor have good potential for automotive safety applications. There are several Tier One suppliers testing these devices with so far excellent results. In addition, the MsS has also shown good potential for condition monitoring of combustion engines such as knock and misfire.

REFERENCES

[1] Kwun, H., "Back in Style: Magnetostrictive Sensors," *Technology Today* (Southwest Research Institute), March 1995, pp. 2-7

[2] Kwun, H. and Teller, C.M., Patent #5456113, "Nondestructive Evaluation of Steel Cables and Ropes Using Magnetostrictive Induced Ultrasonic Waves and Magnetostrictively Detected Acoustic Emissions"; also Patent #5457994 and other pending patents. Assignee: Southwest Research Institute

[3] Kwun H. and Burkhardt G. L., "Nondestructive Measurement of Stress in Ferromagnetic Steels Using Harmonic Analysis of Induced Voltage", *Non-Destructive Testing International*, June 1987, pp. 167-171

[4] Gioutsos, T. and Gillis, Ed, "Testing Techniques for Electronic Single Point Sensing Systems," *SAE International Congress and Exposition*, Paper # 940803, 1994

[5] Gioutsos, T. and Gillis, Ed " Tradeoffs and Testing for Frontal Crash Sensing Systems," *SAE Worldwide Pass. Car Conf. and Expo.*, Paper # 932911,1993

[6] Dahlen, M. "Side Airbag Systems: Seat-mounted vs. Door-mounted", *SAE Side Impact Protection Toptec*, 1994

[7] Hartl, A. Mader, G. Pfau, L. and Wolfram, B., "Physically Different Sensor Concepts for Reliable Detection of Side-Impact Collisions", ," *SAE International Congress and Exposition*, 1995

[8] Breed, D., et al. ,"Performance of a Crush Sensor For Use With Automobile Air Bag Systems," *SAE International Congress and Expo.*, Paper # 920122, 1992

[9] Klauber, R., et. al., "Miniature Magnetostrictive Misfire Sensor," Paper # 920236

[10] Gioutsos, T., Patent # 5,580,084, "System and Method for Controlling Vehicle Safety Device", Assignee : Artistic Analytical Methods, Inc.

970777

Development Methodology of an Airbag Integrated Steering Wheel in Order to Optimize Occupant Protection Balanced Against Out-of-Position Risks

Oliver Spiess, Thomas Marotzke, and Matthias Zahn
PETRI, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

With the increasing number of air bags being used in vehicles there has been a corresponding increase in the number of occupants injured or even killed due to air bag deployment, particularly under out-of-position conditions. Current publications addressing the dangers of air bag deployment has led to worldwide public uncertainty on the safe use of these passive restraint components.

A main goal of the air bag module development must be to balance the out-of-position risks with the overall occupant protection performance to reach the highest biomechanical quality.

Further considerations must be given to the development and production costs, quality and styling of the air bag module and steering wheel. Design, function, cost and styling have competing impact during the product development phase. Economic and efficiency factors require new production and manufacturing of systems that use pre-assembled components as assembly groups, such as the so called "high level assemblies."

Cooperation of system and module design, testing and numerical simulation experts is required to successfully address these concerns. An existing study of an air bag integrated steering wheel as a component of the entire restraint system shows that a strong integration of the CAD, CAE and testing disciplines is the most methodical and effective procedure for restraint system development. The goal of the study is to show and discuss modern, and above all, effective methods for increasing occupant protection. The emphasis lies on the out-of-position occupants.

This study shows the methodology of a development and assessment of a concept for an out-of-position optimized air bag integrated steering wheel. Furthermore, solutions are shown to reduce the out-of-position risks for an air bag integrated steering wheel. The developed solutions are built as functional prototypes. Static deployment tests with a 5th %ile HYBRID III Dummy are performed to analyze and evaluate the effects of the new technology.

INTRODUCTION

Today the driver air bag is standard equipment for a passenger vehicle in the US. In contrast to Europe, the air bag in the US was introduced as a passive restraint system which had and still has to protect the occupant according to the FMVSS 208 along with the kneebolster and the seat design. At this time, however, the study was conducted with an unbelted occupant. The reason for such a decision from NHTSA was a low belt rate of 20% for occupants.

Today the safety potential of an air bag is not in question, with studies showing that 1500 occupants have been saved by the air bag in the last years. Another study of frontal crashes showed an 18% reduction of fatalities due to the air bag and other NHTSA studies showed a reduction of severe injuries to the head, thorax and abdomen [12].

The optimal use of air bags for the ride down of the occupant is only guaranteed if they are fully inflated before contact with the occupant occurs. If the occupant is in contact with the air bag during the unfolding process injuries can occur. These injuries are abrasions, fractures, head injuries and injuries of the inner organs. The situation where the occupant is close to the air bag - or even the worse case - has direct contact with the

module is referred to as out-of-position condition. The possibility of negative side effects from the air bag system in some accident conditions has been known since its introduction to the market. With the rapidly rising number of air bag equipped vehicles, the incidence of accidents with air bag related injuries or fatalities is increasing. Due to the physics, especially small and incorrectly belted occupants have shown the highest risk of severe injuries caused by the interaction with the air bag during its deployment and unfolding process. Press-reports of such accidents, particularly low speed impacts, are the focus today for public discussion of the out-of-position topic.

As a consequence, the following demands for the air bag are in discussion:

- "depowered air bags"
  Reduction of the energy of driver - and passenger air bags

- "smart air bag technology"
  An air bag system which is adjusted to the crash severity and occupant position

The simple change toward "depowered bags" reduces the safety potential for the in-position occupant with no guarantee for an increased safety potential for out-of-position occupants [1]. The "smart air bag technology" necessitates the development of new sensor technology with reliable detection and decision algorithm. The implementation into the vehicle is time consuming and costly. This demands system knowledge and high innovative technology from the safety system supplier. The aim is to reduce the development time and costs and to be flexible for shifting demands.

The development of new solutions should not be a coincidence-product. To reach this goal means to effectively use the available technical and research tools and methods of the system development combined with the product development of the components. The ideal goal must be to develop optimal solutions for a component or the system with no increased cost. That means the supplier has to identify the areas of cost savings and effectiveness of the system in the vehicle. An essential requirement for this is that the supplier is already included as a development partner during the concept phase.

METHODOLOGY

Three fundamental departments of the supplier add up to an engineering team for system development:

- Concept design (CAD)
- Occupant crash simulation (CAE)
- Crash testing (crash facility, dummy-, measurement- and film technology, component facility)

The CAD department is the link to the product development, however, they have enough capacity to develop product innovations out of ongoing research and development activities. The CAD engineers provide vehicle interior geometry and component design for numerical models to be used by the occupant crash simulation. Furthermore, they provide concepts for test rigs and modifications of serial components of occupant protection systems.

The occupant crash simulation develops a parameter matrix of the restraint system components. This allows for a first adjustment of the restraint system to be tested using a sled test facility with high speed measurement, film and video technology.

The occupant crash simulation and the crash test department are connected by a close development loop as shown in Figure 1. Both development tools are well coordinated to efficiently develop the restraint system and to minimize the risks during the development.

SYSTEM DEVELOPMENT METHODS - CAD, CAE and crash testing are connected by a close development loop as shown in Figure 1.

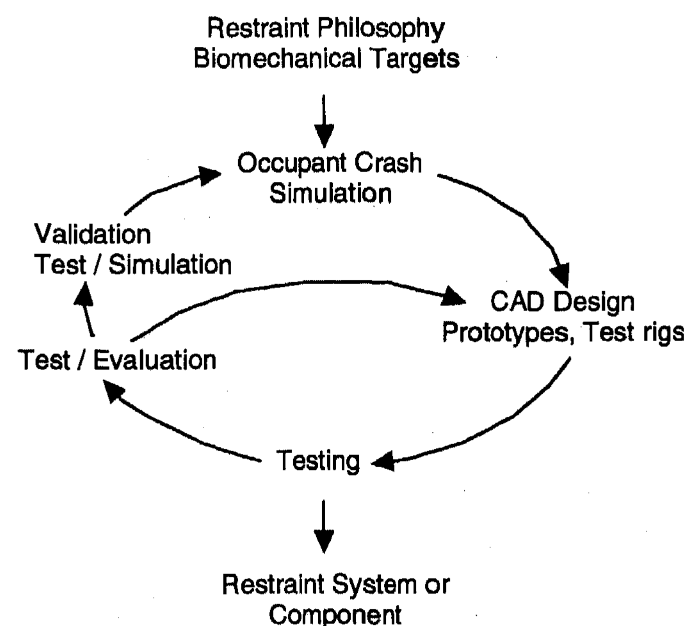


Figure 1: System development method [2]

The evaluation of crash and component tests allows for a step by step validation of the occupant crash simulation model of the vehicle and of the restraint system components. The goal of the occupant crash simulation is to evaluate interesting parameter variations before extensive testing is done.

The main focus of the development results from the examined accident type and determined injury mechanism.

The restraint system of the driver side in frontal impacts includes the following components and subsystems.

- crossmember with interface to body structure
- mounting bracket of steering column
- steering column
- IP and kneebolster
- steering wheel
- driver air bag module
- floor and toe panel, pedals
- seat system (incl. mounting brackets / rails)
- seat belt system
- A-pillar and roof liners
- door panel

Independent of the development of the entire restraint system the optimization of one component must always take into account the entire vehicle system. This method of the scientific and systematic overall view of occupant safety is necessary to effectively develop function and cost optimized products. The following describes the principle steps of the development of an air bag integrated steering wheel as a restraint system component beyond the first prototype.

1. Automotive safety analyses

   1.1 Accident analyses

   Correlation between injury mechanism and the effect of vehicle subsystems or single restraint system components are the result of the evaluation of accident statistics and test results.

   1.2 Basic requirements

   All geometry data and characteristic values of the restraint system are determined. This includes all characteristic values of the restraint system which can not be altered, such as dependent geometry data and characteristic values of existing components.

2. Concept fixation

   First the occupant load limits are determined. This presupposes an extensive, realistic assessment with practicable development solutions based on the scientific expertise of the engineering team. Derived from the first step, the first design concepts are developed.

3. Development

   The three main departments, occupant crash simulation, CAD and crash testing, develop the component using the development loop shown in Figure 1. The development starts with the occupant crash simulation and design of the first functional samples.

4. Prototyping

   As a result of the development ,prototypes are built which meet or exceed the performance criteria.

5. Product development

   The performance criteria as well as design hardpoints for the air bag integrated steering wheel are handed over to the product development. The product development uses the numerical models and the test rigs from the concept phase for further refinement. The main task during this phase is the fine tuning of the component as well as the final design for a fully mechanized assembly line. The final prove out and validation of the developed air bag integrated steering wheel takes place in a full scale crash test.

OOP - PROBLEMATIC NATURE

During a vehicle collision, air bag modules have the task to fully inflate an energy absorbing cushion as quickly as possible before the occupant's forward excursion and subsequently support the occupant. The necessary restraint load is transferred on the occupant using the large area of the air bag. Compared to the seat belt, the restraint load application is essentially reduced. This fact is shown in accident analyses.

In the aggregate, the share of head and cervical spine injuries adds up to no more than 10%, and the statistically relevant AIS ratings of these injuries range between 1 and 2, while the proportion of chest injuries among belted occupants varies between 25 and 30%, at AIS ratings between 1 and 3 [10]. This accident analyses confirms known test results. At identical dummy load levels, belt related loads carry a greater injury risk than air bag related loads. [9]. The main task of occupant safety is the minimization of occupant loads during most accident types and to control the localization, the direction and the application area of the restraint load on the occupant to guarantee a high biomechanical quality and to prevent unnecessary injuries caused by the restraint system [3].

The air bag as a component of the restraint system has the potential, in cooperation with the seat belt, to fulfill this task. The problematic nature has to be analyzed and evaluated before the requirements can be defined.

How can the out-of-position (OOP) or Inflation Induced Injury (I^3) be defined?

OOP describes a situation where the occupant is positioned in the deployment space of the air bag and therefore an uncommon interaction between the head, neck and thorax region and the air bag cushion and module cover can be seen. The direct contact between

the occupant and the module cover is referred to as the worst case scenario. Severe injuries or even fatalities can be the consequence of this interaction. OOP situations in real world accidents can have following causes:

- delayed ignition of the air bag during the vehicle collision
- close by seating position
- forward movement of the occupant due to pre-impact braking

What type of loading mechanism can be seen in OOP situations?

Loading mechanism during the contact phase under OOP conditions have been broken down by Horsch et al. into "punchout" and "membrane loading" phenomena [4]. Punchout occurs in an early phase of the inflation, when the module swells because of the deploying bag until only a small portion of the bag has escaped from the module. Membrane loading denotes the phase when the fabric wraps around a body region as the bag becomes full, thereby exerting a relatively distributed force on the occupant. A third mechanism related with OOP injuries, the so called "bag slap" [11,13] occurs when the occupant is struck by a small but rapidly moving portion of the bag. Even if this effect does not result in critical injuries it can cause skin abrasions or eye injuries.

The following injury criteria are selected and further evaluated:

- Viscous Criterion (VC, thorax)
- Neck Extension Moment
- HIC

OUT-OF-POSITION (OOP) ANALYSIS

PRELIMINARY STUDY - A preliminary study was conducted by Malczyk and Adomeit [5]. This study analyzed the air bag folding pattern as a means for injury reduction for OOP occupants. Static OOP tests were carried out with driver air bag modules mounted to a rigid test rig and folded either according to the PF- or LF-pattern. A Hybrid III small female was placed on a seat in front of the module with two main test configurations: "chest centered on module" and "forehead centered on module" with varying separations from the module ranging from 0 to 100 mm. The test setup for the configuration "chest centered on module" is shown in Figure 2; Figure 3 shows the test setup for the configuration "forehead centered on module".

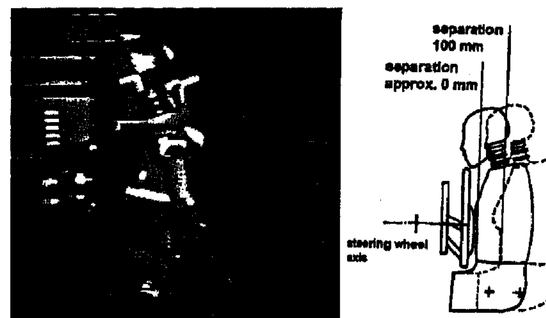


Figure 2: Test setup for configuration "chest centered on module"

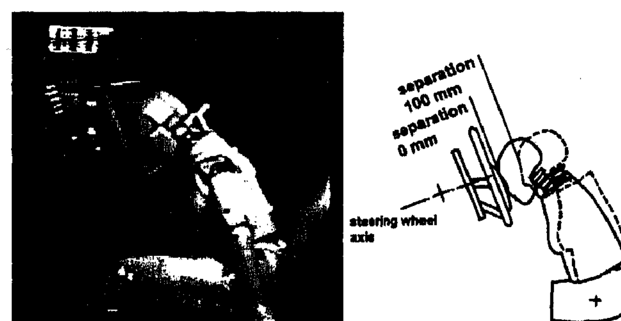


Figure 3: Test setup for configuration "forehead centered on module"

The following main conclusions have been drawn from this study. For the configuration "chest centered on module" values of all relevant injury criteria for the chest and the neck were significantly lower with the PF-pattern. For the configuration "forehead centered on module" neck loadings were lower with the PF-pattern in almost all cases except for neck tension and extension moment at separations of 0 mm and 20 mm. The geometry of the module cover tear seam was found to have negative influence here. Cosmetic markings applied to the dummy's face left far less smeared remains on the P-folded bag in comparison to the conventional folding. This indicates a reduced relative motion between the bag fabric and the skin during inflation. Tests conducted until now concentrated on extreme OOP configurations, with direct contact of parts of the body with the module cover. These configurations are based on the assumption of an unbelted occupant. It is certainly feasible, however, to imagine OOP scenarios with restrained occupants, although reliable information is not available with case figures or the distances between vehicle occupants and the air bag module [4,6]. In this sense, the possible configurations of occupant - module studied in OOP tests should include not only the "worst case position" aspect, but also greater separations between dummy and module.

The tables 1 and 2 provide details on the measured biomechanical loads (VC, neck extension moment, HIC, bold values indicate exceeding of injury threshold). For evaluation of occupant loading the scaled IARVs (Injury Assessment Reference Value) for the small female

Hybrid III given in [6] and [7] were applied. Calculation of the Viscous Criterion (VC) took place in accordance with the scaling formula presented by Mertz [8].

Table 1: VC and neck extension moment for configuration "chest centered on module"

	Separation from module [mm]	Chest VC [m/s]	Upper neck M_y Extension [Nm]
LF-pattern	100	0.28	**59.1**
	50	0.71	**65.5**
	20	**1.16**	**71.4**
	0	**1.78**	**62.6**
PF-pattern	100	0.11	2.4
	50	0.37	9.9
	20	0.9	13.1
	0	**1.46**	19.0
scaled IARV		1.0	31.0

Table 2: Neck extension moment and HIC for configuration "forehead centered on module"

	Separation from module [mm]	Upper neck M_y Extension [Nm]	Head HIC
LF-pattern	100	20.8	60
	50	23.6	96
	20	20.8	250
	0	19.2	359
PF-pattern	100	8.3	75
	50	8.8	81
	20	23.5	190
	0	30.1	235
scaled IARV		31.0	1113

The values are mean values from two tests conducted for each separation from the air bag module.

Regarding the measured injury criteria, the following conclusions for future developments can be drawn to furthermore reduce OOP occupant loads:

- Air bag module design concept to be adjusted to OOP condition (priority folding pattern).
- Distribute OOP loads to sufficient occupant body parts and mass.
- A "minimum distance" between occupant and air bag module of 30 to 40 mm must be provided.

SOLUTION APPROACHES - In order to reach optimal occupant protection the component function has to be developed from the entire restraint system point of view. To reach this goal the performance of the restraint system must not be lowered for the majority of cases to gain performance for the minority.

Taking that into account, the following fundamental solution approaches are possible:

1. Energy reduction of the air bag or "depowered air bags"

   Requirement: - reduction only in case of OOP configuration (advanced sensor system)
   Examples: - dual stage inflator
   - sensor system to detect accident severity
   - sensor system to detect occupant position

2. Optimized occupant load and load distribution on the occupant

   The ideal solution is an adjustable restraint system which can be controlled during the collision. Technically such a restraint system necessitates an inflator with an adjustable mass flow during the gas generation and a corresponding sensor system.

On the assumption that the injury risk for the occupant is a result of partial overload, the demand for the solution approach is specified as follows:

⇒ Reduction of occupant loads during air bag inflation and OOP conditions **without** lowering the occupant protection during in-position conditions using today's technology.

That means that the air bag performance and function is unchanged during all possible accident types. All requirements regarding the restraint system such as air bag inflation time, air bag volume and energy absorbing characteristics of the air bag are adjusted to optimize the restraint system for in-position conditions.

Derived from the OOP analysis and the solution approach analysis the following requirement for the air bag performance is specified:

⇒ Realization of a "minimum distance" between occupant and air bag module in all accident types through the retraction of the air bag module.

DEVELOPMENT OF A RETRACTING MODULE

The preceding analyses admit the conclusion that optimized design solutions can essentially contribute to reduce OOP risks.

In order to retract an air bag module it must be accelerated quickly over the necessary distance.

The following options are possible:

1. Passive retracting module - passive PRM

   The module is retracted by the reaction forces during the cover opening and deployment of the air bag in a direction toward the steering column axis.

2. Active retracting module - active PRM

   The retraction of the module is completed before the cover opening. To retract the module, the energy from the inflator gas escaping or a separate energy source is used.

A separate energy source must be adjusted to the inflator ignition and the complexity increases as the reliability decreases. Therefore the goal must be to use the available energy of the inflator.

Before the product development can begin the realization of the selected concepts must be proven.

TEST REQUIREMENTS - During the development of an air bag integrated steering wheel, numerous requirements of components of the restraint system as well as the entire restraint system must be taken into account, especially the peculiarities of the air bag integration and the retracting module.

The main requirements to be addressed are listed in the following:

1. Restraint System and Components

- FMVSS 201 (ECE-R21)
- FMVSS 203 (ECE-R12)
- FMVSS 208 (ECE-R94)

2. Design

- Environmental tests (includes the steering wheel)
- DOT, BAM (air bag and inflator)

3. Retracting Module

- Guaranteed fixation of the module in the initial and retracted position against head impact, reaction forces and misuse.
- OOP optimized module concept (PF-pattern, design of module regarding optimized deployment trajectory of air bag)

4. Generic requirements

- Availability of developed solution (long / short term)
- Low cost (development and product)
- Flexible design

VALIDATION - Before the product development begins, the realization of the retracting module concept regarding the function must be proven. In the first step prototypes are developed which represent the demanded function and can be used in static OOP tests as described in [5]. To compare the test results the same components are used as in the earlier study and the same test setup.

TESTS CONDUCTED - The first test series is used to verify the possible occupant load reduction under OOP conditions using a passive retracting module. The following tests were selected to be conducted with the retracting module.

1. Configuration "chest centered on module"
   Separation 0 mm and 50 mm

2. Configuration "forehead centered on module"
   Separation 0 mm

Each test is conducted twice.

The idea is to use the reaction forces during the cover opening and deployment of the air bag acting in the direction of the steering column axis, as an energy source to retract the module. The usable force is only available during the cover opening phase because it is the sum of inertia forces from the generated gas mass, the cover and the air bag fabric. The measurable reaction forces are in direct context with the characteristics of the inflator mass flow.

The available air bag steering wheel is modified to allow an unrestrained deformation of the module mounting. In order to hold the module in its initial position, two deformable brackets represent the mounting of the module to the steering wheel. The necessary energy absorption behavior of the deformable brackets was determined through an analyses of measured steering wheel reaction forces during a static air bag deployment test. In accordance with this analysis, simple deformable brackets have been developed and integrated in the air bag steering wheel. The deformable bracket is shown in Figure 4. Steel with a thickness of 1 mm was used to built the brackets. The maximum possible deformation of the brackets is 45 mm. The measured deformation force of one deformable bracket during a static compression test is shown in Figure 5. The possible energy absorption is approx. 36 Nm with two deformable brackets used as the module mounting. Figure 6 shows the mounted air bag module used for the configuration "chest centered on module". The same setup was used for the configuration "forehead centered on module".

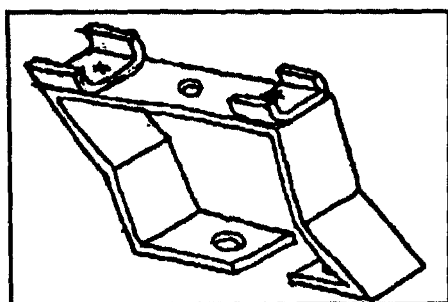


Figure 4: Deformable brackets to mount air bag module on steering wheel

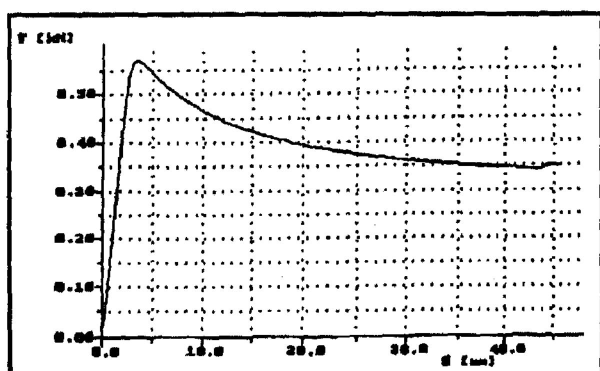


Figure 5: Deformation force of one deformable bracket during static compression test

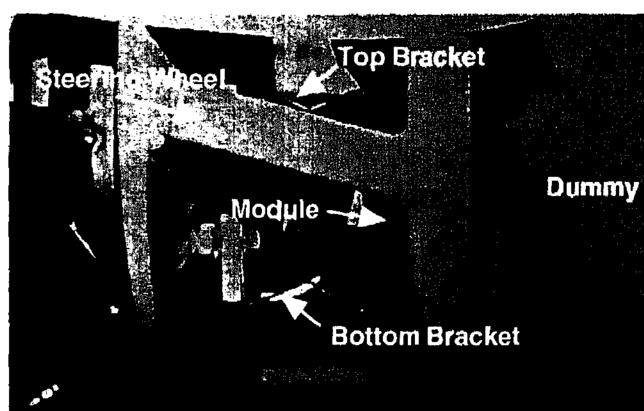


Figure 6: Test setup "chest centered on module" with deformable module mounting brackets

TEST RESULTS - "Chest centered on module" - During the tests with 0 mm separation between occupant and module the top bracket deformation starts directly after the cover opening at approx. 5 ms after air bag ignition. The deformation is finished after 3 ms. The bottom bracket deformation starts at approx. 10 ms after air bag ignition. The end of the deformation can not be seen due to the air bag coverage of this area. The entire deformation distance was used in both tests. The air bag uses the available area caused by the module retraction and the region behind the upper steering wheel rim for deployment.

During the tests with 50 mm separation between occupant and module the top bracket deformation starts shortly after the cover opening at approx. 7 ms after air bag ignition. The deformation is finished after 4 ms. The bottom bracket deformation starts at approx. 16 ms after air bag ignition. The end of the deformation can not be seen due to the air bag coverage of this area. The deformation of the top bracket was 35 mm and the deformation of the bottom bracket was 15 mm. Both brackets have not used the entire available deformation distance. The air bag uses the available area caused by the deformation of the top bracket and the region behind the upper steering wheel rim for deployment.

Table 3 and Figure 7 show the results for the VC and neck extension moment (bold values indicate exceeding of injury threshold) compared between the L-folding pattern (LF), the P-folding pattern (PF), and the P-folding pattern with the retracting module (PF / PRM).

Table 3: Comparison of VC and neck extension moment for configuration "chest centered on module" for LF-pattern, PF-pattern and PF-pattern with retracting module PRM

	Separation from module [mm]	Chest VC [m/s]	Upper neck M_y Extension [Nm]
LF	50	0.71	**65.5**
	0	**1.78**	**62.6**
PF	50	0.37	9.9
	0	**1.46**	19.0
PF /	50	0.30	13.5
PRM	0	**1.07**	21.8
scaled IARV		1.0	31.0

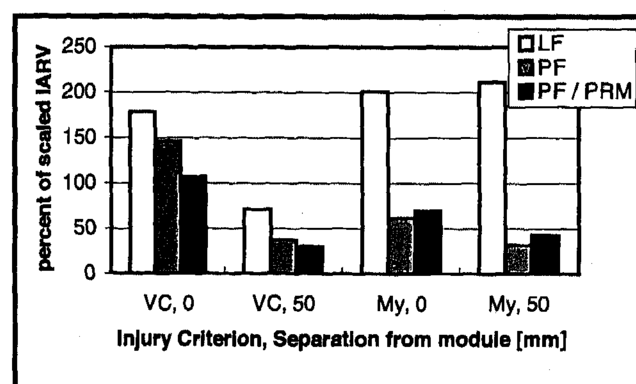


Figure 7: Comparison of VC and neck extension moment for configuration "chest centered on module" for LF-pattern, PF-pattern and PF-pattern with retracting module PRM

A comparison of the air bag deployment and occupant kinematics for the configuration "chest centered on module with 0 mm separation is shown in Figure 8 and for 50 mm separation in Figure 9.



Figure 8: Comparison of air bag deployment and occupant kinematics for PF-pattern with and without PRM, "chest centered on module", 0 mm separation



Figure 9: Comparison of air bag deployment and occupant kinematics for PF-pattern with and without PRM, "chest centered on module", 50 mm separation

"Forehead centered on module" - After the cover opening the air bag inflates towards the upper steering wheel rim for the first 10 ms. During that time the top bracket is deformed. In the beginning the air bag can escape behind the upper steering wheel rim. After approx. 30 ms the entire air bag is located in front of the steering wheel rim. The bottom bracket is deformed between 12 and 15 ms after the air bag ignition. The deformation of the top bracket was 35 mm and the deformation of the bottom bracket was 15 mm. Both brackets have not used the entire available deformation distance.

Table 4 and Figure 10 show the results for the neck extension moment and the HIC compared between the L-folding pattern (LF), the P-folding pattern (PF), and the P-folding with the retracting module (PF / PRM).

Table 4: Comparison of neck extension moment and HIC for configuration "forehead centered on module" for LF-pattern, PF-pattern and PF-pattern with retracting module PRM

	Separation from module [mm]	Upper neck M_y Extension [Nm]	HIC
LF	0	19.2	359
PF	0	30.1	235
PF/PRM	0	22.5	165
scaled IARV		1.0	1113

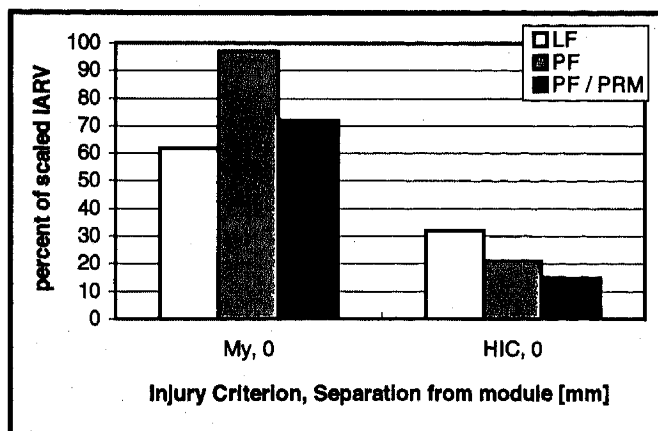


Figure 10: Comparison of neck extension moment and HIC for configuration "forehead centered on module" for LF-pattern, PF-pattern and PF-pattern with retracting module PRM

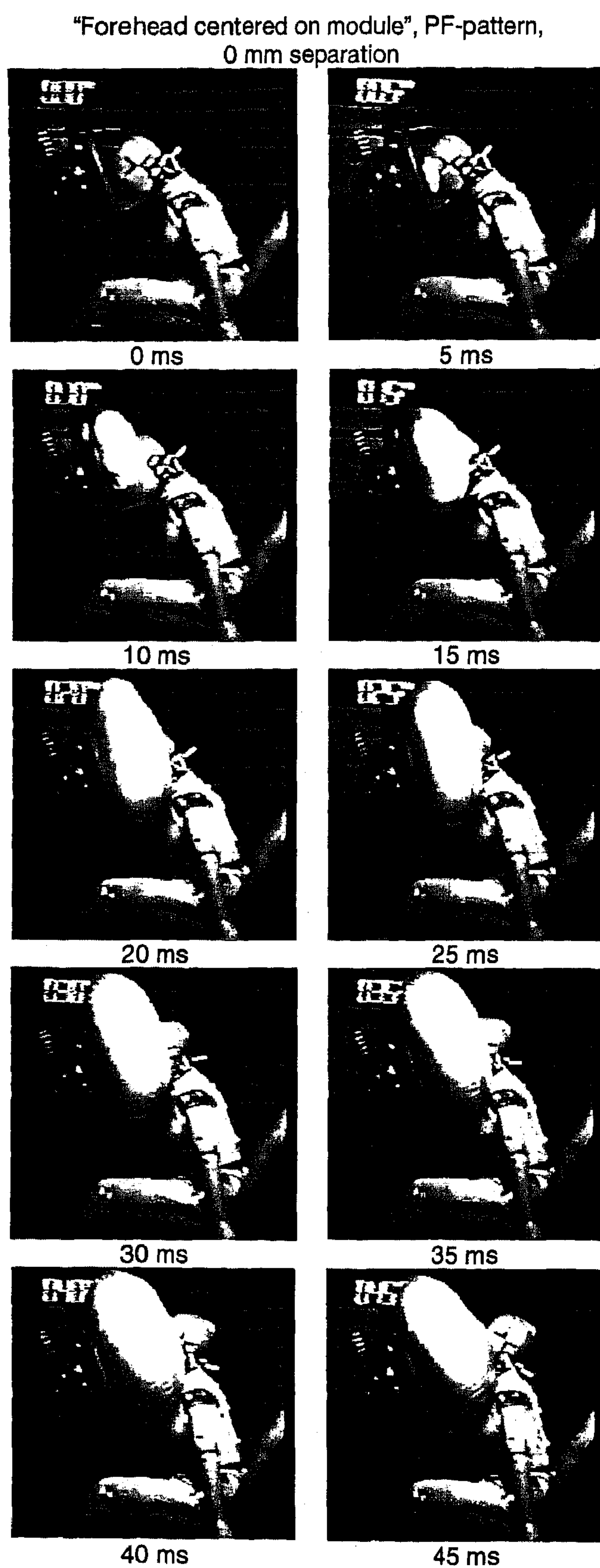
A comparison of the air bag deployment and occupant kinematics for the configuration "forehead centered on module" with 0 mm separation is shown in Figure 11.

TEST RESULT SUMMARY - "Chest centered on module" - The test configuration "chest centered on module" with the PF-pattern and the retracting module showed a significant reduction of the viscous criterion VC compared to the LF-pattern and the PF-pattern without the retracting module. With 0 mm separation the reduction was 26% and with 50 mm separation the reduction was 20%. With 0 mm separation all air bag module configurations exceeded the IARV, however, both tests with the retracting module resulted in VC values slightly above and under the IARV.

The test results with the LF-pattern far exceeded the IARV for the neck extension moment in all tests. The test results with PF-pattern remained far under the IARV for the neck extension moment in all tests. This benefit could not be increased with the retracting module. The neck extension moment in these tests increased slightly, however, they remained far under the IARV of 31 Nm.

"Forehead centered on module" - The test configuration "forehead centered on module" with the PF-pattern and the retracting module showed a significant reduction of the neck extension moment and HIC values compared to the PF-pattern without the retracting module. However, the neck extension moment with the retracting module is slightly higher than with the LF-pattern but remains 30% under the IARV. The increase in neck extension moment for the PF-pattern compared to the LF-pattern can be traced to the interaction of the folding scheme with the tear-open mechanism of the module cover. Analysis of the high-speed films reveals that the torn-open cover becomes wedged in between the deploying air bag and the dummy head situated in front of the module. When this cover finally slides away from the dummy chin and then flips downward, the potential energy of the bag portion located beneath the cover is suddenly transferred to the underside of the chin. As a result of the different bag folding scheme, this wedging of the module door was not so pronounced with the LF-pattern and the rearward rotation of the dummy head was likewise gentler [5]. This phenomenon was mitigated when using the retracting module and the PF-pattern. The neck extension moment with the PF-pattern and the retracting module was decreased by 25% compared to the tests without the retracting module and the PF-pattern. Furthermore, tear seam design may reduce the neck extension moment.

All tests showed HIC values far below the IARV, however, the test with the PF-pattern and the retracting module showed a 20% reduction of the HIC value compared to the test with the PF-pattern and without the retracting module.



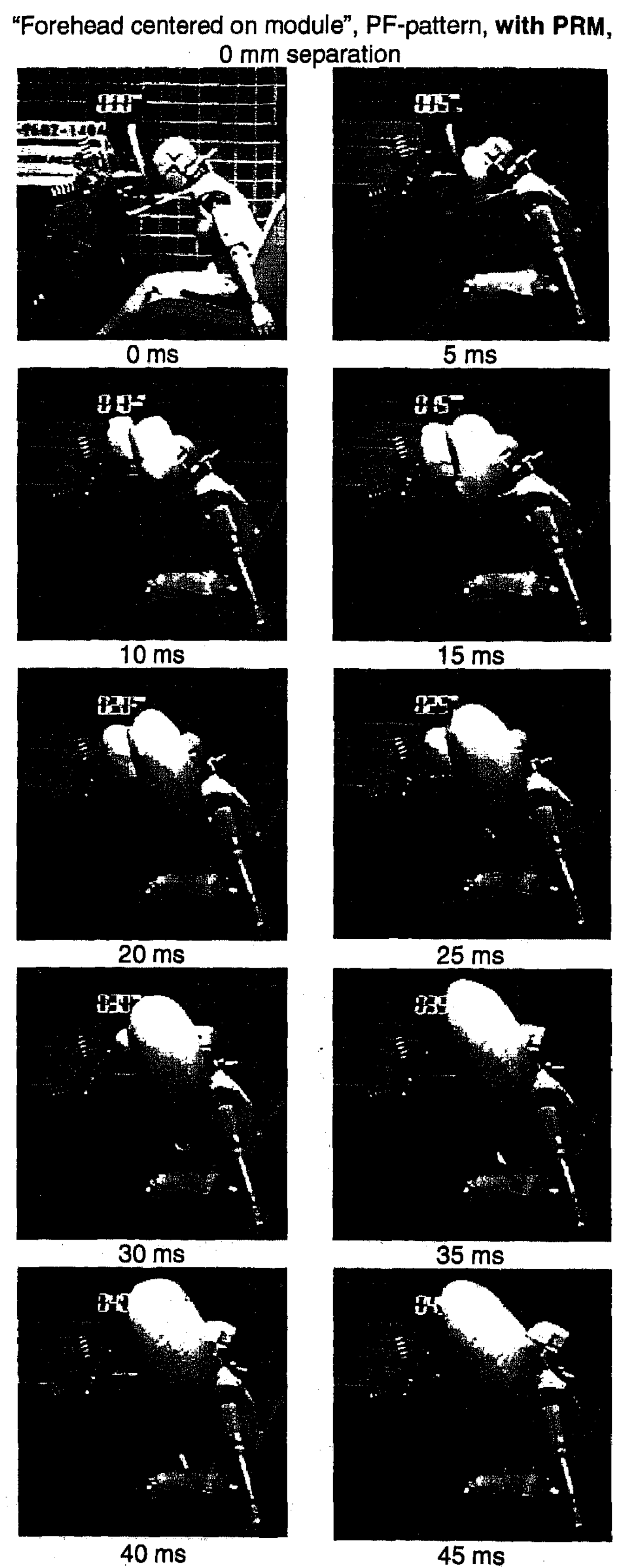


Figure 11: Comparison of air bag deployment and occupant kinematics for PF-pattern with and without PRM, "forehead centered on module", 0 mm separation

SUMMARY AND CONCLUSION

The effective usage of available scientific and systematic tools in the restraint system development requires methodical procedures. This study describes the methodical procedure to develop an air bag integrated steering wheel in order to optimize occupant protection balanced against OOP risks.

The close cooperation between the concept design, occupant crash simulation, crash testing and product development is emphasized. This guarantees the development of a restraint system component from the system point of view.

The evaluation of the OOP topic, analysis of preliminary OOP studies and a methodical solution approach lead to the development of a new component for an air bag integrated steering wheel. The development of a passive retracting air bag module - PRM - was achieved during the methodical procedure and performance of OOP tests.

Compared to OOP tests in [5] the usage of the retracting module - PRM showed a significant reduction for the configuration "chest centered on module" for the viscous criterion VC. The value for the neck extension moment was slightly increased, however, it remained far below the IARV. For the configuration "forehead centered on module" HIC value was reduced by 20% with the usage of the retracting module - PRM. The neck extension moment was reduced by 25%, however, the value was slightly higher compared to the LF-pattern.

It remains to be noticed that the energy absorption of the deformable module brackets was on a very low level. Increasing that level and producing a more even deformation for the module shows another safety potential of the passive retracting module for OOP conditions.

An additional advantage of the retracting module is the use of the mechanism to satisfy the head impact standard (FMVSS 201, ECE-R21). The mechanism is then functioning as a crash can implemented in the steering wheel.

These features included in an air bag integrated steering wheel in addition to the P-fold allow a seamless design of the steering wheel and a smaller package to gain more available distance for the retracting module.

In the future, the active retracting module may furthermore result in reduced occupant loads compared to the passive retracting module due to a mitigated punchout effect. The retraction of the module before the swelling and opening of the module cover may result in reduced punchout loads on the occupant.

Due to an optimized function for all occupant positions, the retracting module concept guarantees a safety potential increase for OOP occupants and no reduction for in-position occupants.

In summary, the test results showed the effectiveness of the retracting module concept - PRM and the analyses admit the conclusion that optimized design solutions can essentially contribute to reduce OOP risks.

The present study suggests that the use of the retracting module - PRM - leads to load reductions for OOP occupants. For the foreseeable future further developments regarding air bag module design, sensor systems and inflator technology show an increase for the safety potential of restraint systems

ACKNOWLEDGMENTS

The authors would like to thank H.-D. Adomeit, A. Malczyk, D. Meißner and H. Ruck for their editorial support and performance of the tests.

REFERENCES

[1] Adomeit, H.-D., Wils, O., Heym, A.; "Adaptive Airbag-Seatbelt-Restraints - An Analysis of Biomechanical Benefits"; SAE Paper # 970776, Society of Automotive Engineers, Warrendale, PA, 1997

[2] Adomeit, H.-D.; "Neue Methoden und neue Ziele bei der Entwicklung von Insassenschutzsystemen"; ATZ Automobiltechnische Zeitschrift 7/8, 1995

[3] Adomeit, H.-D.; "Konzepte von Rückhaltesystemen und Aspekte der Schnittstellen zum Fahrzeuginnenraum"; Haus der Technik Conference Rückhaltesysteme", 1993

[4] Horsch, J. et al; "Assessment of Air Bag Deployment Loads", SAE Paper # 902324, Society of Automotive Engineers, Warrendale, PA, 1990

[5] Malczyk, A., Adomeit, H.-D.; "The Airbag Folding Pattern as a Means for Injury Reduction for Out-of-Position Occupants"; SAE Paper # 952704, Society of Automotive Engineers, Warrendale, PA, 1995

[6] Melvin, J.W. et al; "Assessment of Air Bag Deployment Loads with the Small Female Hybrid III Dummy"; SAE Paper # 933119, Society of Automotive Engineers, Warrendale, PA, 1993

[7] Mertz, H.J.; "Anthropomorphic Test Devices", in "Accidental Injury", Chapter 4; Nahum, A.M., Melvin, J.W. (Editors); Springer Verlag, New York, 1993

[8] Mertz, H.J., "V*C Formulas", ISO Document ISO/TC 22/SC 12/WG 6 N383, N384; October 29, 1993

[9] Nash, C.E.; "Life and Love, Bags and Belts in America"; Bag & Belt 1994, pp. 34-51

[10] Otte, D.; "Aufnahme unfallstatistischer Daten aus In-Depth Untersuchungen am Unfallort / Gurt-spezifische Verletzungen; BAST

[11] Patrick, L.M., Nyquist, G.W.; "Air Bag Effects on the Out-of-Position Child"; SAE Paper # 720442, Society of Automotive Engineers, Warrendale, PA, 1972

[12] Winston, F.K., Reed, R.; "Air Bags and Children: Results of a National Highway Traffic Safety Administration Special Investigation into Actual Crashes"; SAE Paper # 962438, Society of Automotive Engineers, Warrendale, PA, 1996

[13] Zinke, D.T.; "The Development of Air Cushion Restraint Systems for Small Car Front Seat Occupants"; SAE Paper # 800294, Society of Automotive Engineers, Warrendale, PA, 1980

DEFINITIONS, ACRONYMS, ABBREVIATIONS

OOP	= Out-of-Position
P-Fold	= PETRI-Fold
PRM	= PETRI Retracting Module

970812

Study on the Flow in a Typical Seat Valve of Mobile Hydraulics

H. Nguyen-Schaefer, P. Sprafke, and N. Mittwollen
Robert Bosch Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Vibrations induced by the internal flow in a typical seat valve used in mobile hydraulics are investigated under various extreme flow conditions. To reduce the noise associated with these vibrations, it is necessary to gain a thorough understanding of the flow in this particular hydraulic component and of the dynamic characteristic of the hydraulic system. The three-dimensional and turbulent flow in the valve is studied and characterized applying Computational Fluid Dynamics (CFD). Subsequently, dynamic phenomena, such as pressure pulsation, flow rate fluctuation, and body fixed acceleration, are analyzed from measurements on a hydraulic test assembly.

The combined method using CFD and direct measurements suggests itself as a very suitable approach to gain more understanding not only of the flow processes in the valve but also of the causes for the instability of the valve closing body.

INTRODUCTION

Seat valves are widely used basic components of hydraulic systems, especially employed in passenger cars and trucks. Some of these are antilock braking systems (ABS), traction control systems (TCS), continuously variable transmission (CVT), steering systems, dynamic level control systems, and fuel injection systems.

The current trends in development of these systems are smaller size, higher power density, better efficiency, and cost reduction. Therefore, each system component must contribute to satisfy these requirements.

In this paper, an example for a typical seat valve used in a plunger pump is presented. To study the flow in the valve and its influence on system dynamics, there are several methods as follows: flow visualization via enlarged models, numerical investigations (Computational Fluid Dynamics [1], [2] and one-dimensional parametrized behavioural simulation for hydraulic networks [3]), and direct measurements on a hydraulic test assembly.

Enlarged models are usually based on the principle of the same Reynolds number of both flow regimes (original and similar flows). Methods, such as Laser Light Sheet (LLS) and Particle Image Velocimetry (PIV), are generally used to visualize the flow in the enlarged models. They are quite useful for complicated geometries and extremely difficult flows involving strong curvature, swirl, and body force effects but rather expensive. However, it is very difficult either to visualize the three-dimensional velocity field experimentally or to measure pressure and temperature distributions on any sectional surface.

Alternatively, Computational Fluid Dynamics (CFD) is often applied to simulations of flows in the considered component of the system. In recent times, both computer capacities and simulation programs have quickly been developed so that we can calculate the flow for complicated geometries with a very fine grid and complex turbulence models. Unfortunately, CFD is presently not suitable to compute dynamic phenomema like pressure pulsation, flow rate fluctuation, etc. Thus, the one-dimensional behavioural simulation for hydraulic networks has to be carried out. However, some effects, such as cavitation and air release from the liquid that usually take place in the hydraulic components, have not correctly been modeled in CFD and in 1-D behavioural simulations for hydraulic networks.

Hence, direct measurements by means of pressure transducers, hot film anemometer, and acceleration pickup on the original component cannot be replaced yet. Analytical methods in the frequency domain, such as spectrum analysis via Fast Fourier Transformation (FFT), autocorrelation, and cross-correlation, have been used to analyze the measured results.

To study the internal flow in the valve, CFD and direct measurements are chosen in the product design and development phase. This combined method is very suitable to analyze and solve complex dynamic problems fast and economically.

DESCRIPTION OF THE SEAT VALVE IN THE PLUNGER PUMP

The cost-efficient plunger pump described in Figure 1 applies a very simple principle. The plunger is moved backwards and forwards by an eccentric wheel for the intake and the compression phase, respectively.

At the beginning of the intake stage, the pressure in the internal pump chamber decreases until it is less than the pressure at the pump inlet. Thereupon, the hydraulic liquid starts flowing from the inlet through the circular channel on the plunger into the internal chamber.

The pressure in the internal chamber continuously rises at the beginning of the compression stage until it is larger than the given pressure at the pump outlet. The closing body (further called „ball") is then opened and the pumping process continues as long as the plunger is moving forward.

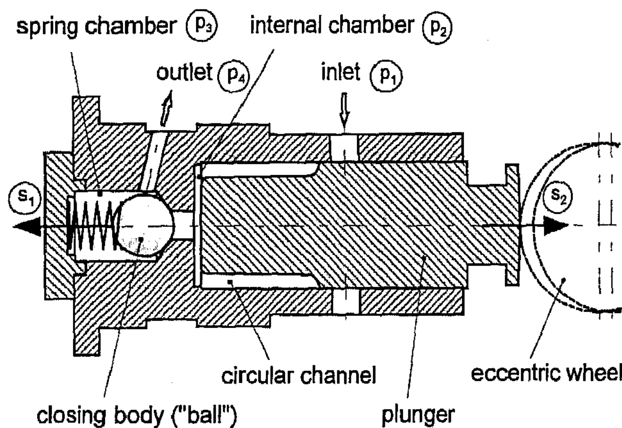


Figure 1: Schematical description of the plunger pump

NUMERICAL STUDY AND RESULTS

COMPUTATIONAL GRID - For the numerical study of the pumping process, it is necessary to model the flow in the outlet seat valve of the pump consisting of the internal chamber, the spring chamber behind the ball and the outlet bore (see Figure 1).

The simulated domain is one half of the seat valve because of the geometrical symmetry of the cases where the ball is axisymmetric or eccentric only in the direction towards the outlet bore. The computational grid is directly generated from the CAD geometry by using the code ICEM-CFD [4]. Figure 2 shows this grid for the seat valve.

Defining the degree of mesh refinement necessary to ensure a mesh independent numerical solution is quite difficult, particularly for three-dimensional problems. It depends on many factors, such as geometrical configurations, flow properties (viscous regions, separation, possible unsteadiness), mesh qualities (cell aspect ratio, cell skewness), numerical schemes, and turbulence models. Various types of structured or unstructured meshes can be applied, but in all cases, besides numerical schemes, mesh qualities will greatly influence accuracy and convergence of numerical solutions.

In general, finer meshes reduce discretization and truncation errors but cause rather high costs of the computation. Principally, the mesh must be fine enough to generate sufficiently accurate solutions but should not be finer than necessary. Much experience is required to determine the appropriate number of nodal points as well as their distribution for the computational problem.

A multiblock and structured mesh whose elements are trilinear and hexahedral is used for this numerical study. It includes 27 topological blocks and about 100,000 nodal points.

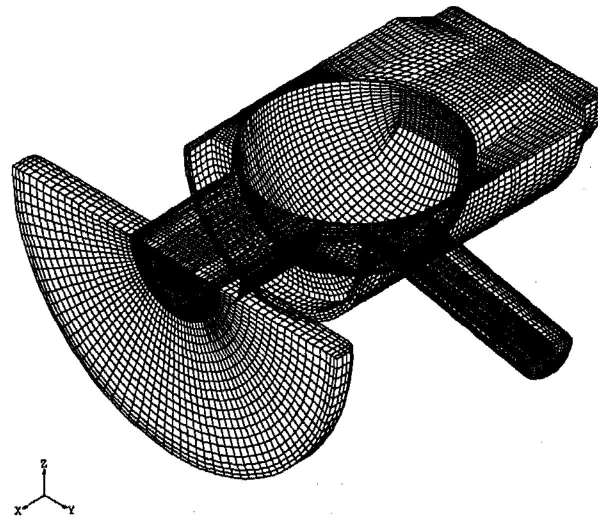


Figure 2: Computational grid for the seat valve

Grid generation nowadays takes most of the working time in CFD simulations, in particular for complicated geometrical configurations. Hence, there is a strong need for automatic mesh generation algorithms, especially for structured and unstructured meshes with hexahedral elements.

NUMERICAL SIMULATION - The flow in the valve is assumed as steady-state, isothermal, and incompressible. The simulation is based on the Reynolds Averaged Navier-Stokes equations (RANS) and the continuity equation [5].

At first, the continuous governing partial differential equations are discretized by the Galerkin approach, a weighted residual method. The principle of the Galerkin formulation, the finite element method, is that the integral of the weighted residual of the equation system over the computational domain is zero [6]. For the steady case, a system of algebraic equations is generated from the discretized equation system. This algebraic equation system connecting the solution nodal values is solved

either directly or iteratively [7]. In this paper, the commercial code FIDAP [8] has been used for the computation.

The considered numerical model is an elliptic boundary value problem for subsonic flows, i.e., Ma < 1. For this reason, boundary conditions for the computational domain have to be supplied.

The relationship between the lift of the ball and the plunger displacement can be determined from the one-dimensional behavioural simulation for hydraulic networks. The lift is defined as zero, when the valve is closed. Therefore, in the three-dimensional computation the flow in the valve is simulated for each given constant lift.

The velocity of the fluid used for the boundary condition at the valve inlet is assumed to be the speed of the plunger, which is easily calculated from the rotational velocity of the eccentric wheel and its eccentricity. The pressure boundary condition at the valve outlet is set to be zero as the reference pressure. On the wall of the valve body, all velocity components are equal to zero. Because of the symmetric boundary condition, the velocity component normal to the symmetric valve plane must be equal to zero as well.

The Reynolds number Re that defines the relation between the momentum force and the viscous force is generally used to determine the flow regime of the fluid. For the flow in typical valves in mobile hydraulics, the Reynolds number at the valve seats amounts to a value of the order of 10^3.

The valve geometry shown in Figure 1 induces an annular jet flow at the seat where the flow jet becomes mostly unstable even at lower Reynolds numbers, i.e., Re < 10^3.

In order to get accurate numerical solutions, either the total nodal points of the mesh must be in the order of $Re^{9/4}$ for 3-D problems, or the computation must be run with a turbulence model for the corresponding flow regime. Even though the flow is at relatively low Reynolds numbers, the first approach involving direct numerical simulation (DNS) cannot be carried out on workstations, because it would need immense capacities only available on fastest super-computers including optimized numerical algorithms. Hence, a turbulence model for low Reynolds numbers is usually applied to modeling the flow in the considered hydraulic valves.

In the standard k-ε turbulence model regarding the gradient-transport concept, the Reynolds stress is related to the mean-velocity gradient via a so-called turbulent eddy viscosity that is assumed to depend on the turbulent kinetic energy k and the turbulent dissipation ε. This assumption is quite suitable for high Re numbers but rather inaccurate for low Re number flows involving viscous boundary layers, transition, and separation.

Therefore, we apply an alternative approach to conventional closure schemes using the renormalization group theory (RNG), which has been developed by Yakhot and Orszag [9]. The RNG method uses dynamic scaling and invariance together with iterated perturbation techniques to evaluate the transport equations for the large-scale modes and their transport coefficients. Computations with the RNG turbulence model have shown a better agreement with experiments for complex flows (separation and swirling flows) where the standard k-ε turbulence model often fails [9]. Additionally, the main advantage of the RNG analysis is its independence of any experimentally adjustable coefficients.

The RNG approach is effective in both low and high Re number flow regions. Hence, it is generally available for near-wall areas as well as for highly turbulent flow regions. However, computation only with the RNG model requires a very fine mesh in the near-wall region. In order to save computation costs, the combination of the RNG turbulence model and the Reichardt's universal near-wall model has been used for the numerical calculation. With this approach, the so-called local Reynolds number y^+ on nodes next to the wall can be less than 30 instead of between 30 and 100 in case of the standard k-ε turbulence model with a near-wall model; i.e., the RNG model working with a high order of accuracy has directly been applied in viscous sublayers or in separation regions.

NUMERICAL SCHEME - In order to stabilize the numerical solutions, the streamline upwinding (SU) scheme developed by Hughes and Brooks [10] has been used in the computation. The high-order convection term resulting from the finite element Galerkin method can be unstable. Such instability may be alleviated by the SU scheme via explicitly adding numerical diffusion only along the flow direction. Practically, the combined scheme (Galerkin and SU) substantially stabilizes convection instabilities with only a minor loss of accuracy.

Due to the limited main memory of computers, the discretized system of equations, particularly for 3-D problems, is usually solved iteratively. The pressure projection scheme, a modified algorithm of the MAC (Marker and Cell) method [11], has been used where velocities and pressures are iteratively corrected in two steps [8].

The applied segregated algorithm requires the solution of both non-symmetric and symmetric linear equation systems at each iteration step. The conjugate gradient squared (CGS) and the conjugate residual (CR) iterative methods have been utilized to solve the non-symmetric and symmetric linear equation systems, respectively [7].

In order to substantially improve the convergence of conjugate gradient methods, the application of preconditioning is necessary. Gauss-Seidel and Jacobi preconditioning are available for symmetric and non-symmetric linear systems of equations, respectively.

In addition, for accelerating the convergence of the linear or nonlinear iterations, the residual projection on the subspace (RPS) scheme of Karamyshev et al. [12] is employed, where the solution vector is corrected by an adjustment vector consisting of a linear combination of the previous residual vectors.

NUMERICAL RESULTS - Some selected computational results are presented. The first calculation begins with a lift of the ball of 0.100 mm for the case of the axisymmetric position of the ball, i.e., the center of the ball lies on the valve axis. A second computation with a lift of 0.150 mm for the case of an eccentric ball position of 0.100 mm has been carried out. The velocity of the fluid at the valve inlet of 0.5 m/s and a zero reference pressure at the outlet have been chosen as boundary conditions.

Figure 3 shows the velocity field at the valve seat and on the valve plane of symmetry. At the edge of the entrance of the cylinder joining the internal chamber to the valve seat, flow separation takes place. Since the total pressure drop mainly occurs at the valve seat, the fluid reaches a maximum velocity in the valve seat region near the entrance of the outlet bore. Note that the fluid velocities in the spring chamber are very small due to its large dead volume. In the region located between the ball and the wall of the valve opposite to the outlet bore, recirculation flow (vortex) is observed.

In order to study the internal flow in the valve in more detail, the velocity vectors at several selected sectional areas perpendicular to the valve axis are displayed in Figure 4. The flow in the domain near the ball equator gradually changes from its axial direction at the side opposite to the outlet to the tangential direction along the ball equator. Therefore, most of the fluid flows to the outlet bore in front of the equator. However, a small part of the fluid first enters the spring chamber and then returns to the outlet.

At the inlet area, some fluid particles have been selected whose paths are shown from two different directions of view in Figures 5 and 6. The particles at the right side of the inlet area flow directly to the outlet bore. At the left side, the particles either change their directions behind the valve seat but in front of the ball equator and then move towards the outlet bore or recirculate from the spring chamber to the exit.

A pressure distribution in the valve is presented in Figure 7 where the dark and light shading values point out high and low isobaric domains, respectively. The maximum pressure at the valve inlet area is about 3 MPa, while the minimum occuring in the separation zone at the entrance of the outlet bore is approximately 0.4 MPa below the reference.

Since pressure energy is transformed into kinetic energy at the valve seat, most of the pressure drop takes place there, in particular near the outlet bore. The recirculation flow in the spring chamber accounts for its increased pressure being about 1 MPa above the reference.

Figure 7 shows that the pressure in the internal chamber is higher than the pressure in the spring chamber. Hence, a force on the ball is acting in the negative x direction. In addition, the pressure at the side opposite to the outlet exceeds the outlet pressure. Therefore, another force is acting on the ball in the y direction. Due to the symmetric boundary condition, the resulting force component on the ball in the z direction is of course equal to zero.

By way of integrating the normal (pressure) and shear stresses on the ball surface, a resulting force has been calculated. In this case, the force component in the negative x direction is 6 N, which is still greater than the return spring force, the component in the y direction is 2 N. This suggests that the ball will still be opened further resulting in a larger lift and simultaneously be drawn into the direction of the outlet bore, i.e., the position of the ball will be eccentric to the valve axis.

In order to determine whether the ball will be returned from the eccentric to the axisymmetric position, a second computation with a larger lift of the ball of 0.150 mm, an eccentricity of 0.100 mm from the valve axis to the outlet, and the same boundary conditions as in the first case has been carried out.

Figure 8 gives the pressure distribution in the valve for this case. Since the lift of the ball is larger than in the first case, the maximum pressure at the inlet area is reduced from 3 to 2.4 MPa. An additional throttling effect at the entrance of the outlet bore is created due to the ball eccentricity. Therefore, the pressure in the spring chamber increases from 1.0 (in the axisymmetric case) to 1.8 MPa.

The pressure distribution on the ball accounts for the change of direction of the x-force component from -6 to 1.5 N and the increase of the y-force component from 2 to 5 N. This implies that the ball has been stabilized at a lift between 0.100 and 0.150 mm leaning against the wall of the valve body near the entrance of the outlet bore.

The movement of the ball from the valve axis to the outlet bore causes the quasi-closing of the spring chamber. This effect is further called „ball turnover". Because of the additional throttling effect, the pressure not only in the spring chamber but also in the internal chamber rapidly rises. Due to the pressure increase in the spring chamber, the ball is pushed back into the valve seat. Thereupon, the internal pressure goes up further and the ball moves forward again. The movement of the ball is repeated until it stabilizes at the wall near the outlet entrance. Thus, pressure pulsations not only in the spring chamber but also in the internal chamber and the whole hydraulic system accordingly develop. The quasi-incompressibility of the liquid accounts for the high frequency and amplitude of the pulsations. That causes the vibration of the ball directly at the beginning of its lift before it stabilizes at the outlet bore. The instability of the ball will be discussed in the next section in more detail.

In order to get converged numerical solutions, well-estimated starting values are required. Hence, the solutions with the k-ε turbulence model have been used as starting values for the subsequent computation with the RNG turbulence model.

The computation for the valve with a grid consisting of about 100,000 nodal points requires a main memory of 75 Mbytes as well as 11 and 15 CPU minutes per iteration step on a workstation HP 9000-J200 for the k-ε and the RNG turbulence model, respectively.

For the applied segregated solver, a solution convergence tolerance of 10^{-3} has normally been employed for 3-D problems. For this convergence tolerance, 50 and 75 iteration steps for the k-ε and the RNG turbulence model, respectively, are required for the computation.

EXPERIMENTAL STUDY AND RESULTS

To supplement the computational investigations, the pressure characteristic at the inlet p_1, in the internal chamber p_2, in the spring chamber p_3, and at the outlet p_4 and the acceleration of the valve body in three directions have directly been measured on the hydraulic test assembly (s. Figure 1). The plunger movement s_2 has been determined by means of an inductive displacement pickup.

In this study the pressures at the inlet p_{in} and the outlet p_{out} are 1 and 10 MPa, respectively for all investigated cases.

At first, we investigate the vibrational characteristic of the valve at a low rotational speed of the eccentric wheel of 1500 rpm, then at a relatively high rotational speed of 3500 rpm.

The plunger movement s_2 and the vibrational characteristic of the pressures p_2, p_3, and p_4 at the rotational speed of 1500 rpm are presented in Figure 9. During the intake stage, the plunger moves backward from the top dead center (TDC) to the bottom dead center (BDC) and reversely for the compression stage.

During the intake stage, the pressure in the internal chamber p_2 is a little less than the inlet pressure p_1. However, the internal pressure p_2 is much larger than the vaporisation and the air saturation pressure at the operating temperature. Therefore, neither cavitation nor air release occurs in this stage. The internal pressure p_2 increases quasi-linearly with time at the beginning of the compression stage until the ball of the outlet valve is opened.

The resulting radial force on the ball determined in the computational study is relatively small at the low flow rate for this rotational speed. It is not strong enough to draw the ball to the outlet bore. For this reason, it is assumed that the ball vibrates mainly in the axial direction after the valve has opened. Thus, the pressures in the internal chamber p_2, the spring chamber p_3, and at the outlet p_4 are evanescently oscillating with a frequency of approximately 1.7 kHz until the beginning of the next intake stage. Note that the pressures in the spring chamber p_3 and at the outlet p_4 are nearly the same but the internal pressure p_2 is different from the outlet pressure p_4, i.e., the main pressure drop is at the valve seat. These results indicate that the ball stabilizes in the valve axis.

Figure 10 shows the vibrational characteristic of the pressures p_2, p_3, and p_4 and the plunger displacement s_2 at the rotational speed of 3500 rpm. Differing from the previous case, the internal pressure p_2 drops to nearly zero (vacuum) due to the high velocity of the plunger in the intake stage. Thus, air release and cavitation take place in the internal chamber. Even though the compression stage has begun, the internal pressure p_2 is still nearly vacuum for a while, before it suddenly shoots up to a pronounced spike. This characteristic can be explained as follows: the existing air and vapour bubbles in the fluid in the internal chamber must be compressed until they completely vanish. Thereupon, the pressure of the single-phase fluid (i.e., without air and vapour bubbles) in the internal chamber rapidly increases in the compression stage.

At this rotational speed corresponding to a high flow rate, the resulting radial force on the ball is strong enough to move it to the outlet bore. For the reason discussed in the previous section, a strong pressure pulsation occurs in the spring chamber and the internal chamber directly at the beginning of the ball lift. In this case, the pressures in the internal chamber p_2 as well as in the spring chamber p_3 are nearly the same because the additional throttling effect due to the ball turnover covering the outlet bore is stronger than the flow resistance at the valve seat. Hence, the pressure drop between the internal chamber and the outlet is substantially increased compared to the previous case. These results confirm that now the ball stabilizes leaning against the wall of the valve body near the entrance of the outlet instead of oscillating in the valve axis.

The instability of the ball at the high rotational speed of 3500 rpm has two consequences. First, the high pressure pulsations in the internal chamber p_2 and the spring chamber p_3 cause vibrations in the valve body associated with noise. Second, the additional throttling effect accounts for the low efficiency of the pump and therefore its high electric power input. This is particularly critical in winter when the increased fluid viscosity due to low temperatures raises the electric power requirements even more.

In order to characterize the vibrations induced by the flow in the valve, accelerations of the valve body have been measured in various directions. The Fourier analysis has been applied to transform the acceleration vs. time function into frequency function [13].

To suppress the pump frequency, accelerations of the valve body perpendicular to the plunger displacement are presented for two investigated cases with 1500 and 3500 rpm in Figures 11 and 12, respectively. At the rotational speed of 1500 rpm, the ball turnover effect has not happened yet; the pressure pulsation is very small. This results in a low excitation of the valve body that is visible through a peak of the acceleration signal at the frequency of nearly 1.7 kHz (s. Figure 11). Quite conversely, at the rotational speed of 3500 rpm, high pressure pulsation occurs because of the ball turnover. Thus, the maximum acceleration of the valve body is also seen at the frequency of about 1.7 kHz in Figure 12.

To analyze the correlation between the pressure pulsations and the accelerations of the valve body in more detail, the cross-correlation method has been used [13]. The resulting cross-correlation function of two signals contains only those frequency components common to both waveforms.

The cross-spectrum of the internal pressure p_2 and the acceleration of the valve body at the rotational speed of 3500 rpm is displayed in Figure 13. A good correlation is indicated at the pump frequencies of the first (0.058 kHz) and second order (0.116 kHz) and at the higher frequency of 1.7 kHz. This result indicates that the internal pressure pulsation is a major noise source in the valve with the high excitation frequency of 1.7 kHz.

CONCLUSIONS

The presented methods allow us to better understand the flow and the dynamic characteristic in the valve. Furthermore, the causes of the flow induced noise and of the low pump efficiency have been explained. This thorough understanding is necessary to develop, design, and optimize new valves for low noise, high pump efficiency, and low costs.

In order to verify the assumptions on the ball movement, we will try to develop a method to measure it. Further direct insight will be gained when the 3-D, turbulent, and unsteady-state flow computation is coupled with the 1-D behavioural simulation of the other mechanic and hydraulic system components.

REFERENCES

[1] H. Nguyen-Schaefer, T. Grauer, C. Jansson, K. Engelsdorf, and M. Mettner: Fluid Dynamics in Vehicle Valves, VDI, 5th International Congress, Würzburg, 1st-3rd Oct. 1990.

[2] T. Grauer, H. Nguyen-Schaefer, C. Jansson, and K. Engelsdorf: Numerische Strömungsdynamik für die Ventilentwicklung, Beitrag zum Fachgebiet Hydraulik des 9. Aachener Fluidtechnischen Kolloqiums, Aachen, 20.-22. März 1990.

[3] N. Mittwollen: Hydraulic Simulation of Cavitation Induced Pressure Fluctuations with Peculiar Periodicities in a Fluid Power Unit, 8th Bath International Fluid Power Workshop on Fluid Power Systems, Bath, Sept. 1995

[4] Control Data, ICEM-Technology: ICEM-CFD Version 3.1.2-BCU7-B, April 1996.

[5] C.A.J. Fletcher: Computational Techniques for Fluid Dynamics, Vol. I and II, Springer-Verlag, 1991.

[6] G.F. Carey and J.T. Oden: Finite Elements - Fluid Mechanics, Vol. VI, Prentice-Hall, 1986.

[7] L-A. Hageman and D.M. Young: Applied Iterative Methods, Academic Press, 1981.

[8] Fluid Dynamics International, Inc.: FIDAP 7.6, Update Manual, 1996.

[9] V. Yakhot and S. Orszag: Renormalization Group Analysis of Turbulence, I. Basic Theory, Journal of Scientific Computing, pp. 3-51, 1986.

[10] A.N. Brooks and T.J.R. Hughes: Streamline Upwind/ Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, Comp. Meth. Appl. Mech. Eng., 30, 1982.

[11] F.H. Harlow and J.E. Welch: Numerical Calculation of Time-Dependent Viscous Incompressible Flow of Fluid with Free Surface, Phys. Fluids, Vol. 8, 1965.

[12] V. Karamyshev, V. Kovenya, A. Sleptsov, and S. Cherny: Accelerating Convergence of CFD Iteration Methods, Proc. of the 6th ISCFD, Lake Tahoe, USA, 1995.

[13] R.W. Ramirez: The FFT Fundamentals and Concepts, Prentice-Hall, 1985.

Author's address:

Dr. H. Nguyen-Schaefer
c/o Robert Bosch Company
Corporate Research and Advanced Development
Dept. FV/FLM
P.O. Box 10 60 50
70049 Stuttgart
Germany
Phone: +49-711-811-7668
Fax: +49-711-811-7607
Email: nguyen@flm.sh.bosch.de

About the author:

Dr. Nguyen-Schaefer received a Dipl.-Ing. degree in 1985 and a Doctorate degree in 1989 in mechanical engineering, both at the University of Karlsruhe, Germany. Since 1988 he has worked at Robert Bosch Company, in Section Corporate Research and Advanced Development.

His main research interests are in finite element and finite volume methods and internal flow applications for hydraulic components and systems used in passenger cars and trucks. For over eight years, he has actively engaged in various research and development projects.

Working domains: Computational Fluid Dynamics, One-Dimensional Behavioural Simulations for Hydraulic Networks, Noise induced by Flows, Cavitation and Air Release Modeling, and Direct Measurements on the original hydraulic components and systems.

APPENDIX

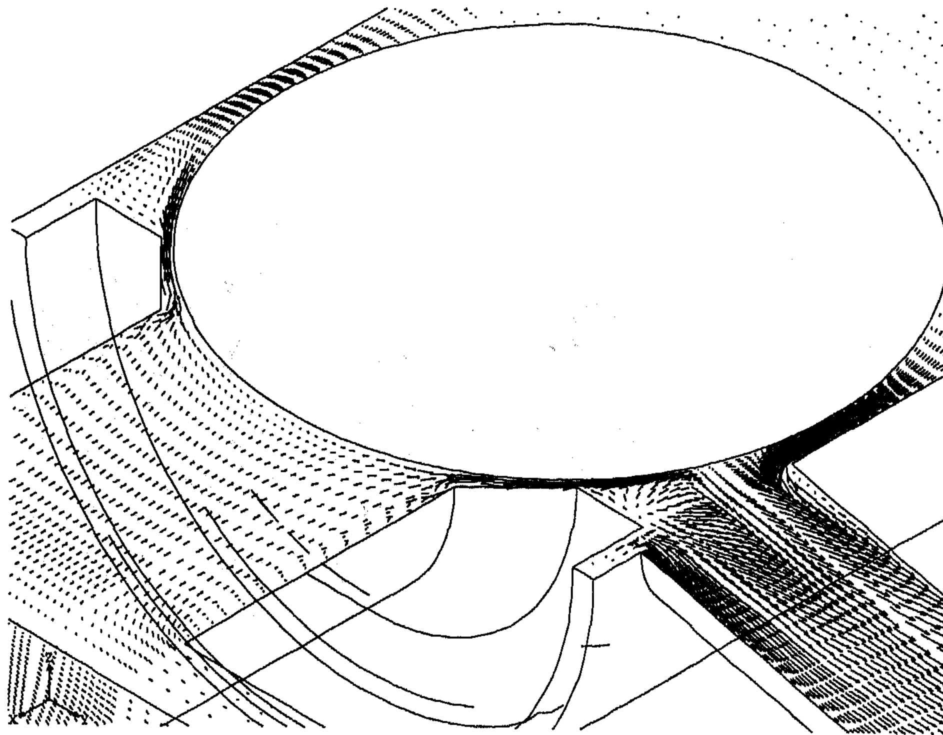


Figure 3: Velocity vectors at the valve seat ($v_{max} = 65$ m/s)

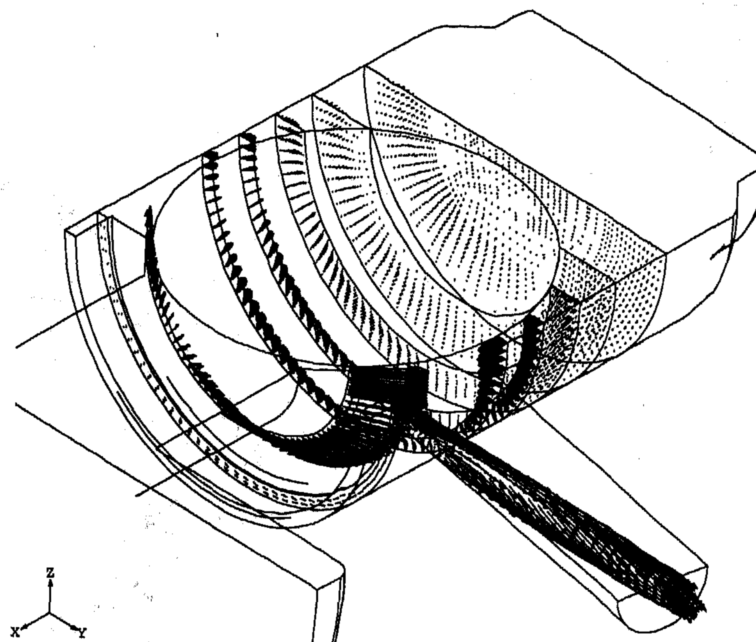


Figure 4: Velocity vectors at several sectional areas

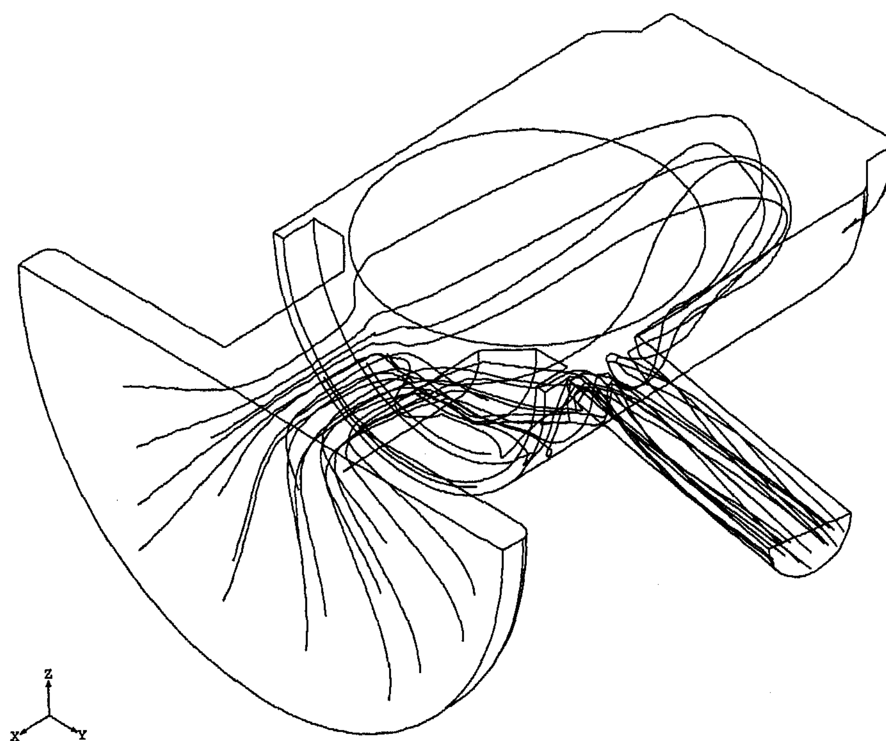


Figure 5: Several particle paths in the valve

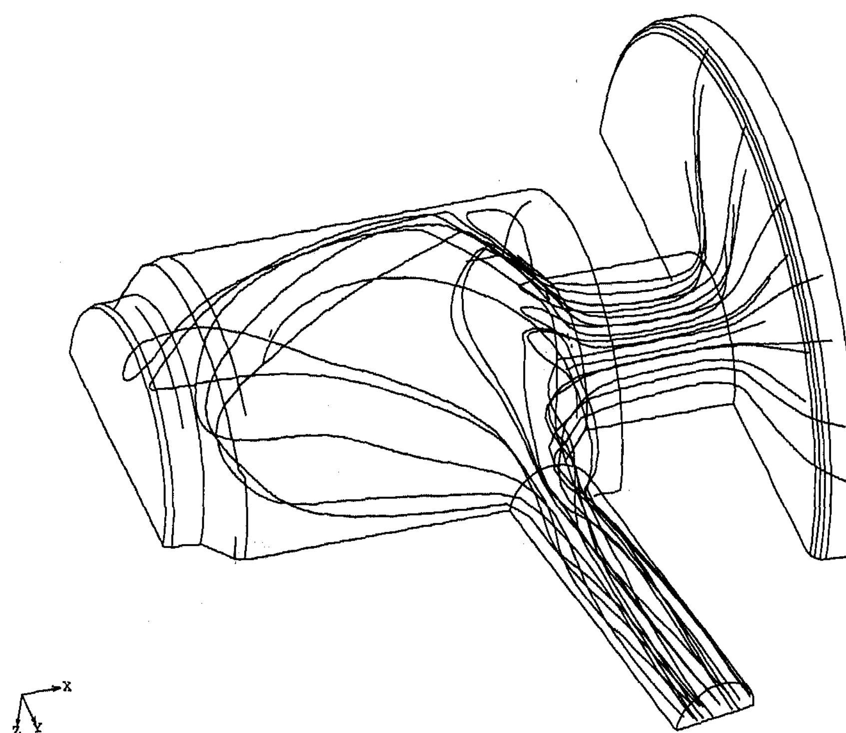


Figure 6: Several particle paths in the valve (continued)

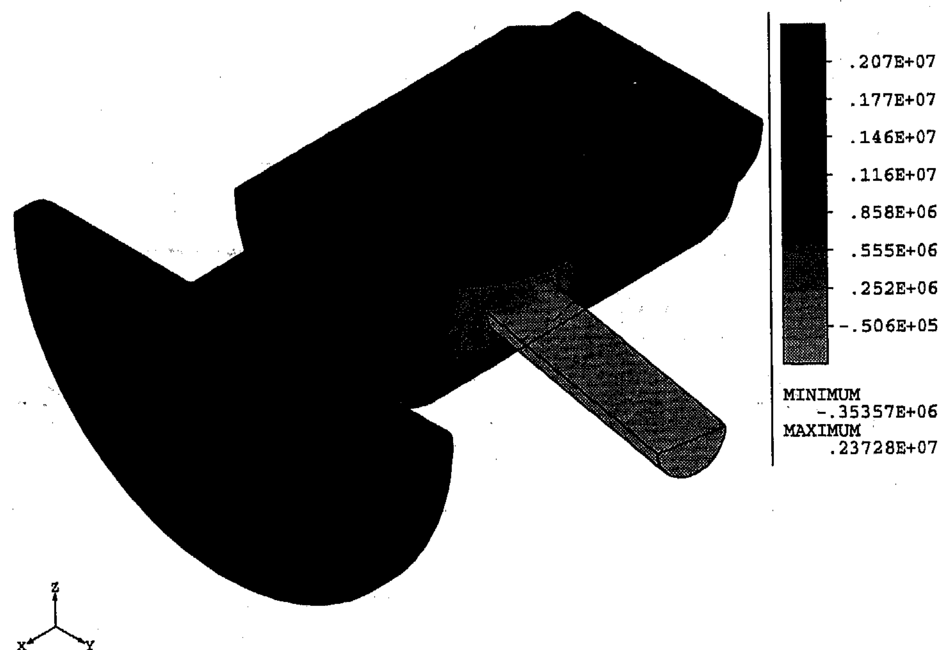


Figure 7: Pressure distribution in the valve in case of the axisymmetric ball position (pressure p in Pa)

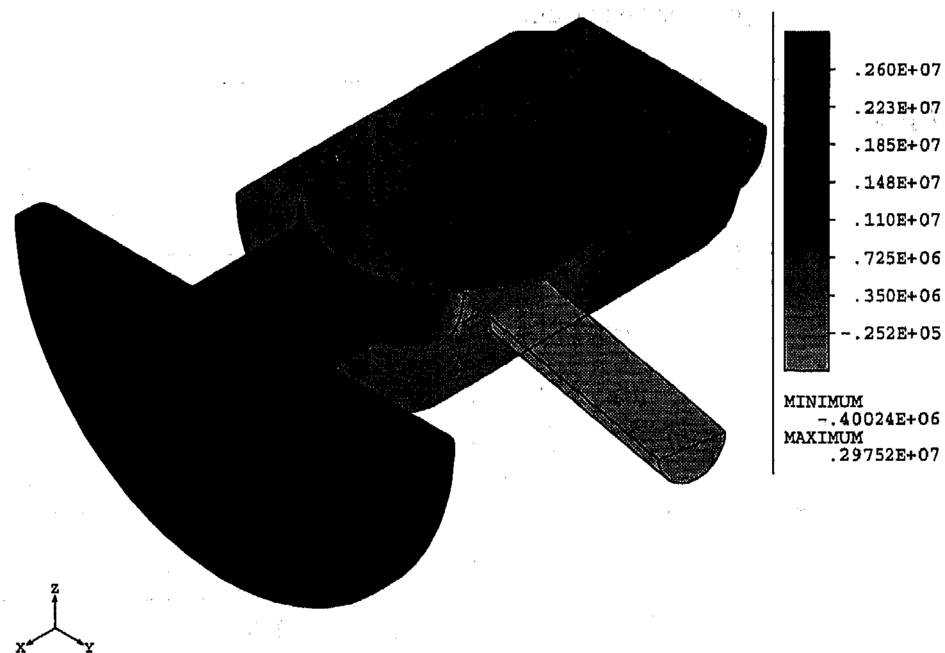


Figure 8: Pressure distribution in the valve in case of the eccentric ball position (pressure p in Pa)

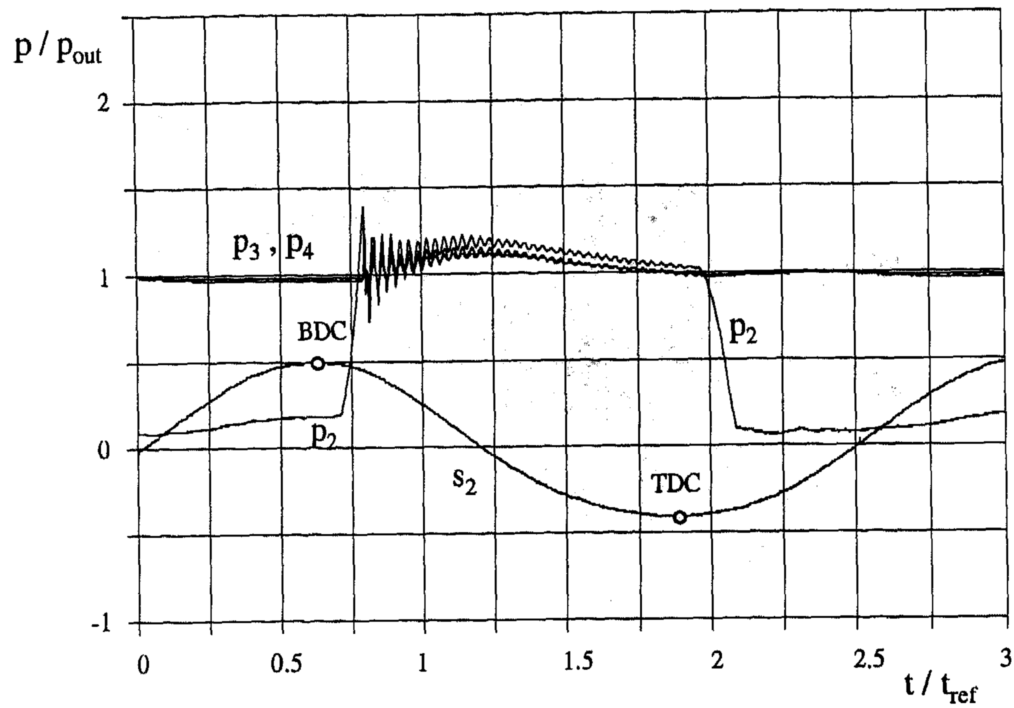


Figure 9: Pressure characteristic in the valve at a rotational speed of 1500 rpm (p_{out} = 10 MPa, t_{ref} = 20 ms)
p_2: internal chamber; p_3: spring chamber; p_4: outlet bore;
s_2: plunger displacement

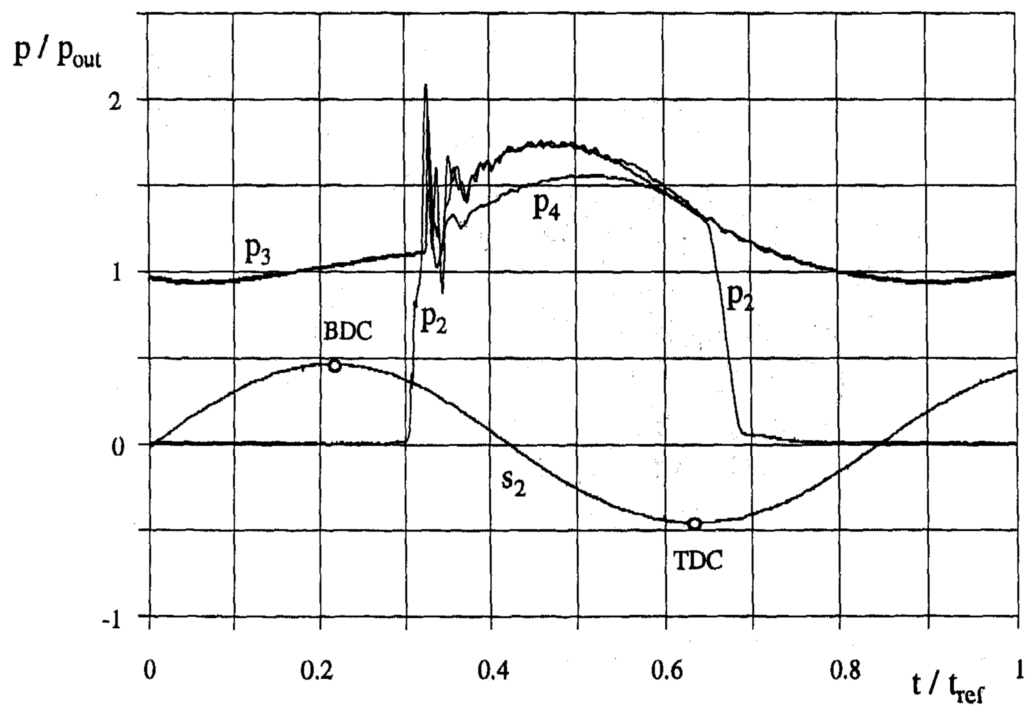


Figure 10: Pressure characteristic in the valve at a rotational speed of 3500 rpm (p_{out} = 10 MPa, t_{ref} = 20 ms)

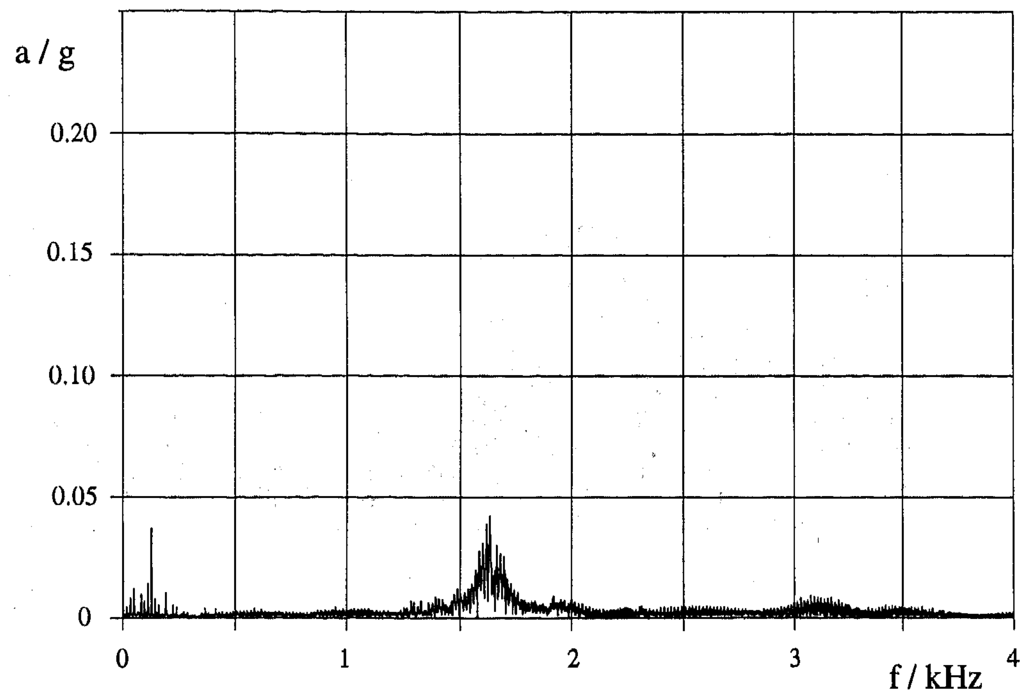


Figure 11: Acceleration spectrum of the valve body at a rotational speed of 1500 rpm ($g = 9.81\ m/s^2$)

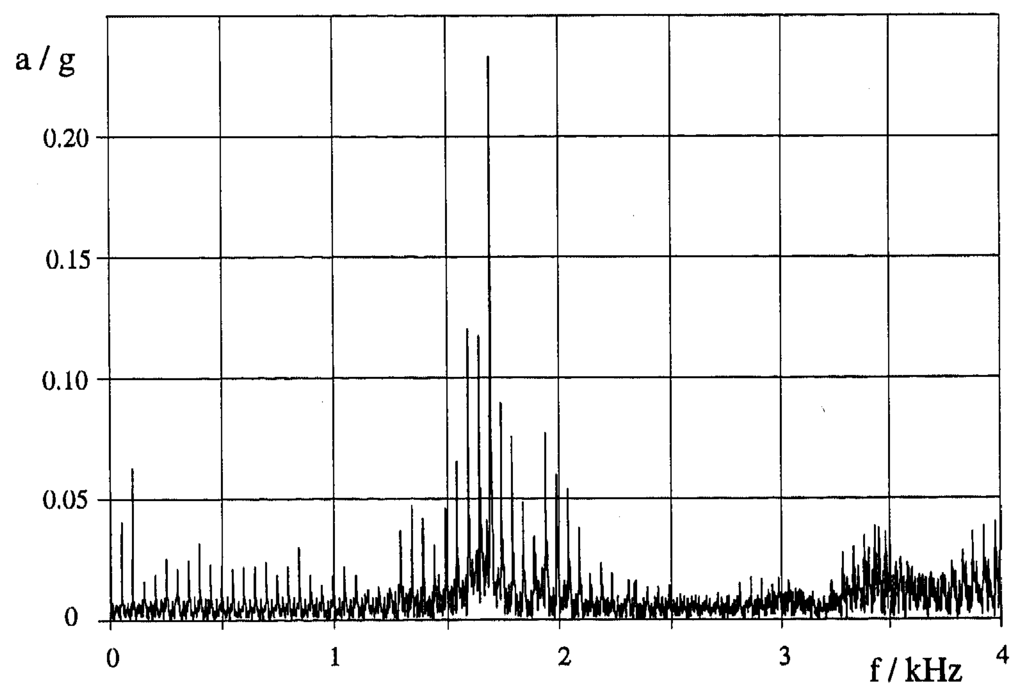


Figure 12: Acceleration spectrum of the valve body at a rotational speed of 3500 rpm ($g = 9.81\ m/s^2$)

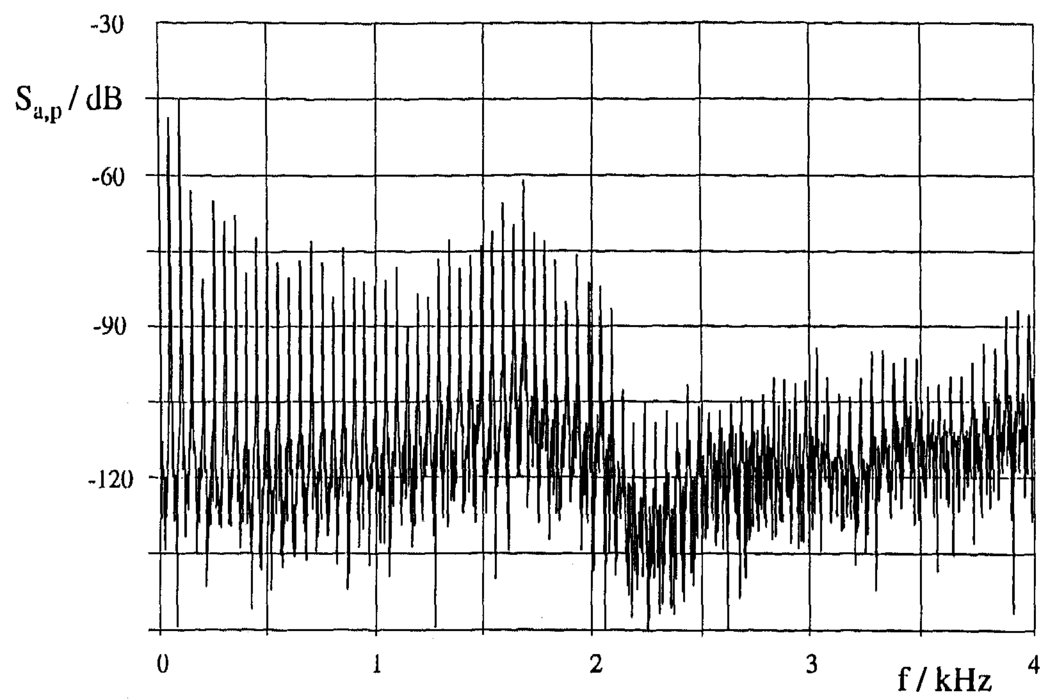


Figure 13: Cross-spectrum of the internal pressure p_2 and the acceleration a of the valve body at a rotational speed of 3500 rpm

970831

Measured and Predicted Effects of Air Flow Non-Uniformity on Thermal Performance of an R-134a Evaporator

Aidan M. Ryan and David J. Timoney
University College Dublin

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Tests have been carried out over a wide range of operating conditions on a single circuit plate-finned tube type of refrigerant evaporator, exchanging heat between ambient air and refrigerant R134a. Uniform and non-uniform air flow distributions were applied with the objective of studying the effects of non-uniformity on heat transfer.

The experimental results revealed that, for a given total volume flow of air, heat transfer performances could be up to 15% better with a non-uniform flow where turbulence levels were also high.

Extensive comparisons were also made between these measurements and predictions from a computer model. This was designed to simulate the thermal behavior of cross-flow heat exchangers on a tube segment-by-segment basis, allowing for flow non-uniformity. Predictions were found to match measurements satisfactorily for test cases involving uniform air flows and a relatively large temperature difference between fluids. In contrast to the measured trends, the simulation results predicted a degradation in heat transfer performance for non-uniform air flow cases.

INTRODUCTION

Refrigeration and air conditioning system designers often consider the air flow velocity distribution across compact heat exchangers to be uniform when estimating heat exchanger performance. In practice, the flow profile can be highly non uniform. Space limitations and ducting arrangements often lead to mal-distributed air flow profiles, particularly in domestic and automotive applications. Use of axial and centrifugal fans in forced draught mode may also give rise to non-uniformity in the air flow.

A program of experimental and analytical work is underway at University College Dublin, aimed at the development of a reliable performance simulation model for cross-flow evaporators. This paper presents some results from experimental and analytical studies of the performance of one particular evaporator under conditions of (i) low turbulence uniform air flow (provided by a wind tunnel), and (ii) high turbulence, non-uniform air flow from a centrifugal fan. These results give some insight into the mechanisms by which air flow velocity and turbulence profiles can affect the overall performance.

The overall objective of the work is to establish whether design changes could usefully be made to an evaporator so that benefits could be obtained from a non-uniform air velocity distribution. Once adequately validated, the simulation software is intended for use as a design optimization tool.

REVIEW OF PREVIOUS WORK

Kitto & Robertson [1] suggest that the causes of mal-distribution in heat transfer equipment fall into one of the following categories;

1. Non-uniformities induced by mechanical design.
2. Flow self induced non-uniformity.
3. Two-phase flow effects.
4. Fouling.

In practice, the effects of these non-uniformities manifest themselves as degradation of thermal

performance or interruption of mechanical operation (i.e. failure).

For the type of evaporator being studied here, mal-distribution could be expected to affect thermal performance because;

(1) Tube side fluid property variations occur continuously throughout the heat exchange coil. Transformation from two-phase to single phase flow is unavoidable, thereby introducing significant variations in local tube-side heat transfer coefficients.

(2) Choice of fan and ducting arrangements will introduce non-uniform air flow, with consequent variations in local gas-side heat transfer coefficients.

A number of the published studies of the effects of non-uniform air flow have been based entirely on computer simulation. For example, Fagan [2] investigates the effects of parabolic, linear and step air flow distributions on the performance of a plate finned tube heat exchanger, assuming the refrigerant pressure drop to be negligible and the tube-side heat transfer coefficient to be constant. He suggests that mal-distributions of up to 75% of the average face velocity can produce evaporator capacity losses of 3 to 16%.

Chiou [3] proposed a numerical method to determine deterioration in heat exchanger effectiveness due to non-uniformity. He suggests that a generalized relationship between "flow non-uniformity factor" and performance deterioration could be developed. Two subsequent publications [4,5] are concerned with application of this method to studies of the performance of condensers and evaporators.

The work of Berryman & Russell [6] looks at the effects of forced and induced-draught air flow on air cooled heat exchangers. They suggest that mal-distribution will lead to loss in thermal performance of only a few percent. Their results also suggest that induced air flow can be as non-uniform as forced flow.

The effects of non-uniformity of inlet air temperature and velocity profile were investigated by Rabas [7] who used an analysis procedure similar to that of Fagan [2]. The results show that degradation of effectiveness increases with increasing non-uniformity of the air velocity profile. Temperature variations can have positive or negative effects but are less important than the effects of two dimensional flow mal-distribution.

Kondo et al [8] suggest that the *rms* value of the air flow distribution and thermal performance deterioration rate can be simply related. The space-averaged *rms* value of the entire velocity profile or degree of non-uniformity *K* is defined by;

$$K = \sqrt{\frac{1}{A}\int_A \left(\frac{V(A) - \overline{V}}{\overline{V}}\right)^2 dA} \qquad (1)$$

A is the total frontal core area, *V(A)* is the local velocity and $\overline{V}$ is mean frontal core velocity. This definition was also used in the present work.

Simulation studies were carried out as part of the present work. The results were in general agreement with those presented in these published works. However, the present authors have also undertaken extensive experiments to provide validation data and to clarify the experimental effects of air flow non-uniformity.

Simulation of Heat Exchanger Thermal Performance

General Description

The simulation modeling approach used for the present work was based on that used in the "EVSIM" program published by Domanski [9, 10]. This is based on a tube-by-tube analysis of a compact evaporator and accounts for local variations in tube-side heat transfer coefficients. On the air side, two-dimensional flow distributions can be specified at entry to the evaporator, providing a basis for evaluation of local heat transfer coefficients. The original EVSIM model has been altered significantly. Its workings are now in S.I. units and subroutines have been written to calculate the thermodynamic and transport properties of R-134a. The modified program has been named "EVSIM-2D" and all predicted results presented in this paper are based on this version.

The tube-side convective heat transfer model considers three regimes; (a) annular, (b) dispersed two-phase flow or (c) superheated vapor flow. The refrigerant-side heat transfer correlation of Wattelet et. al. [11] was used for the present work. The air side heat transfer coefficient is computed using equations from Gray & Webb [12] and the fin-tube contact resistance uses the correlation of Sheffield et. al. [13]. The formation of liquid on the tubes as a result of

condensation of ambient water vapour is represented with a continuous film model which takes account of the increased thermal resistance of the condensate film and of the associated latent heat transfer.

The pressure drop through the tubes is estimated using the Pierre [14] correlation for two phase flow and that of Fanning for single phase. Note that the refrigerant pressure drop is caused by (i) the frictional effects and (ii) the momentum change of the refrigerant it changes from liquid to vapor.

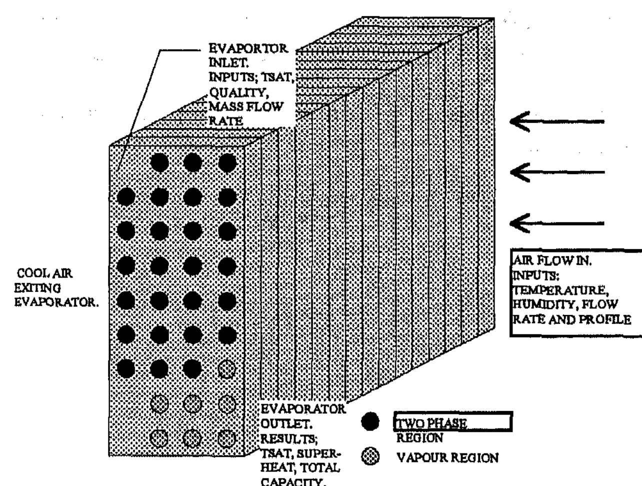


Figure 1: Geometric layout of single circuit plate-finned-tube evaporator - showing computer simulation input and output parameters. Note that the refrigerant flows in series through 33 tubes.

Properties of R-134a

Correlations for thermodynamic properties of R-134a have been published by a number of research groups, including [15-20]. The correlations of Kabelac [21] have been used for the present work.

Transport properties of R-134a are less widely publicized than the thermophysical properties and the data from different sources show less agreement. For the present work, liquid viscosity is calculated using an equation from Jung & Radermacher [22] whilst the vapor phase viscosities are found using data from I.C.I. Klea [20]. The liquid thermal conductivity is determined using a correlation from Jung & Radermacher [22] and that of the vapor is based on the correlation of Tanaka et. al. [23], the latter having been developed and validated with experimental data using R-134a.

Program Input / Output

Input and output data formats have also been revised from the original EVSIM format. *Refrigerant mass flow rate is now specified as input* rather than computed and the refrigerant exit pressure and temperature are computed rather than required as inputs. The air flow analysis has been changed so that less orthodox tube arrangements can be specified (e.g. vacant tube slots) and the program now provides for 2-dimensional air flow non-uniformity. Each tube is considered as a series of short segments where the heat transfer and condensation for each segment is influenced by the local variations in velocity and consequent variations in temperature, humidity and local heat transfer coefficient.

The program input and output parameters are illustrated in Figure 1.

Experimental Work

Refrigeration circuit

The evaporator under test formed part of a specially constructed refrigeration circuit, as shown in Figure 2. This was designed to provide maximum flexibility in the range of operating conditions achievable and was comprehensively instrumented.

The single circuit plate-finned tube type of heat exchanger tested (Figures 3, 4 & 5) had 33 tubes arranged horizontally in a 4 x 9 matrix (with 3 vacant slots). These were configured in a 25 mm in-line layout with a fin block length of 300 mm, a fin spacing of 4 mm and an external tube diameter of 9.5 mm. The total air side heat transfer surface area was 3.194 m^2. The refrigerant flows in series through 33 tubes, as illustrated in Figure 4.

The air temperature and humidity were determined in most cases by the ambient conditions in the laboratory. Air Inlet temperatures varied between 16 and 23°C and relative humidities ranged from 25 to 60% during the tests reported here.

An over-sized refrigerant compressor was used in conjunction with a hot gas by-pass loop. This provided evaporator mass fluxes in the range 40 to 400 kg/m^2s. Measurements of instantaneous refrigerant mass flow rate were made using a Pelton wheel type flow meter, installed in the liquid line. The condenser and hot gas by-pass flows were water cooled using plate heat exchangers. Flow through the evaporator was initially controlled with an internally equalized thermostatic expansion (TX) valve (Test Group A) but this was found to hunt unless the level of superheat was kept unreasonably high. Note that testing was carried out only when steady state refrigerant mass flow conditions prevailed. High superheat levels give rise to inefficient use of the available heat transfer area.

This is because a large number of tubes are involved in superheating. Overall heat transfer coefficients are low for these tubes and refrigerant pressure drop can be high. The situation was improved markedly when the an externally equalized (TEX) valve was fitted (Test Groups B, C & D).

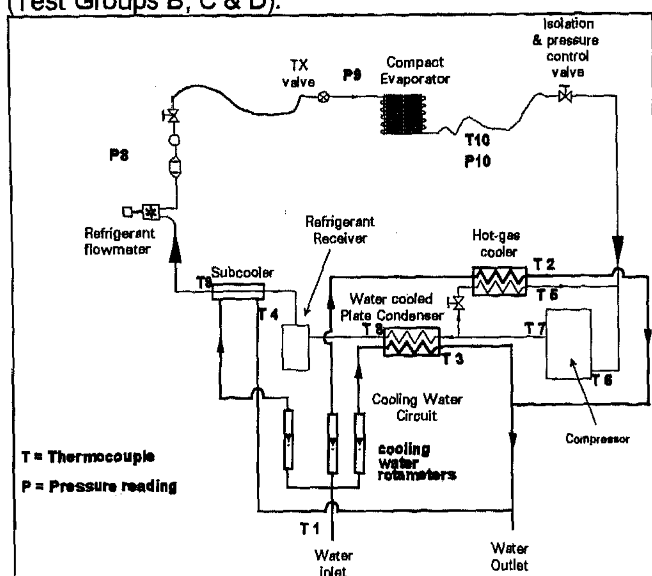


Figure 2: Schematic diagram of refrigeration circuit incorporating test evaporator.

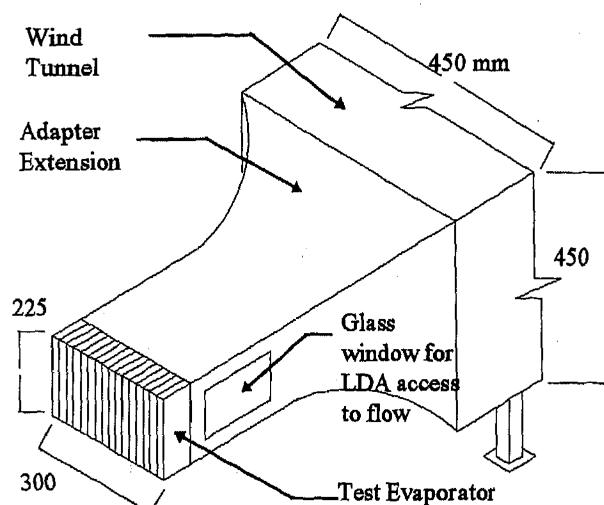


Figure 3: Evaporator mounting on open circuit wind tunnel.

Uniform Flow Test Rig

Air flow for the uniform case was provided by a blower wind tunnel (Figure 3). A converging duct was used to reduce the tunnel area to that of the evaporator and Plexiglas panes were inserted into this to provide optical access for LDA measurements. Also, holes were drilled at regular intervals, 50 mm upstream, to allow for insertion of a hot wire anemometer probe.

The hot wire was used to measure the flow velocity and turbulence characteristics whilst the LDA was used for calibration only. With this arrangement, the non-uniformity was 3.3% with average turbulence levels of 0.7%.

Non-Uniform Flows

The work discussed here has been directed at study of non-uniform air flow cases. Arbitrary profiles were produced by a dual inlet, forward bladed centrifugal fan (Figure 5). The fan is powered by a 3-phase 0.75 kW induction motor, fitted with a continuous speed control system. The duct connecting the fan outlet to the heat exchanger was fitted with Plexiglas side plates through which LDA and hot wire measurements were taken.

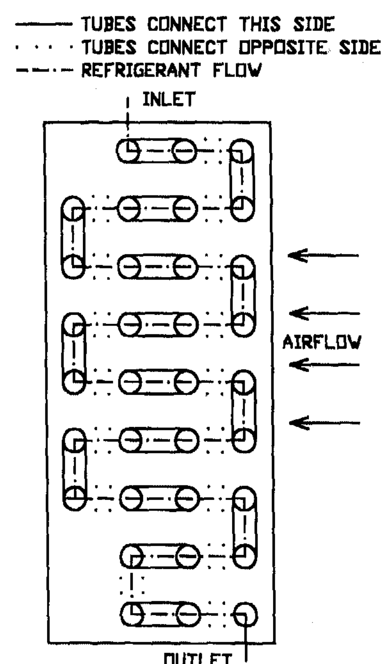


Figure 4: Refrigerant flow path through evaporator.

Air Flow Surveys (Non-Uniform Flow Tests)

Hot wire anemometry was used to measure both uniform and non-uniform velocity profiles. Measurements were made using a Dantec constant temperature anemometer system with a 5 micron x 1.25mm probe wire, supported on a long tube of 6 mm OD which was inserted into the flow stream in a plane, 50 mm upstream of the evaporator inlet.

The hot wire measurement system was calibrated before every test batch using a laser Doppler Anemometer (LDA) system. This was done in a wind tunnel under conditions of low turbulence. A best fit to Kings Law was effected by regression analysis.

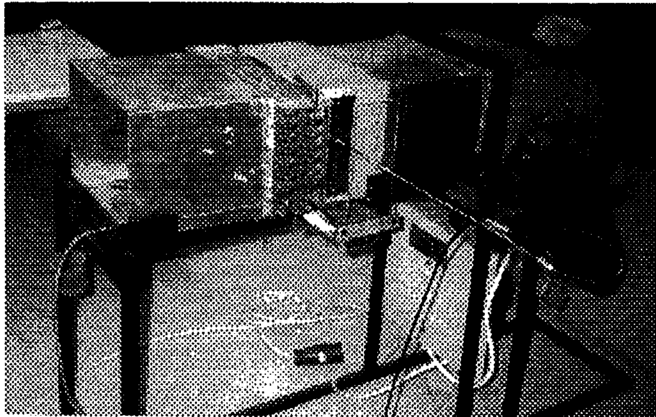


Figure 5: Photograph of centrifugal fan rig.

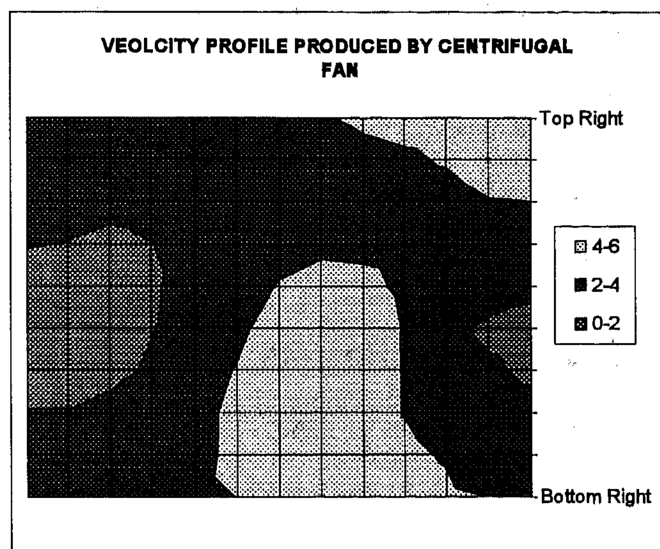


Figure 6: Two dimensional velocity distribution across the evaporator inlet - produced by forward bladed centrifugal (velocities in m/s). Mean velocity = 3.28 m/s. Non-uniformity = 31.5%.

The velocity profile produced by the fan is highly non-uniform (Figure 6), with a non-uniformity factor, *K*, of 31.5%, for this case where the mean velocity was 3.28 m/s. Air is induced from the two inlets at either side which meets centrally in the duct creating high flow velocities at the center of the duct. This is slightly asymmetrical due to the presence of a fan belt and guard on one side.

Figure 7 illustrates the turbulence intensity profile 50 mm upstream of the evaporator. For this test, the average turbulence intensity (over 130 measurement positions) is 30.3 %, with a local maximum of 70% and minimum of 10%.

Figure 8 summarizes the non-uniformity of the air flow velocity profile produced by the centrifugal fan over the full range of average velocities tested. Non-uniformity is calculated using equation (1).

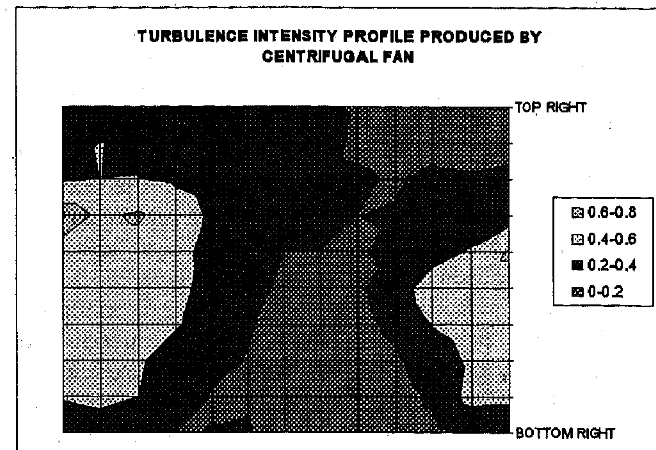


Figure 7 Profile of turbulence intensity created in centrifugal fan air flow profile for mean velocity of 3.28 m/s.

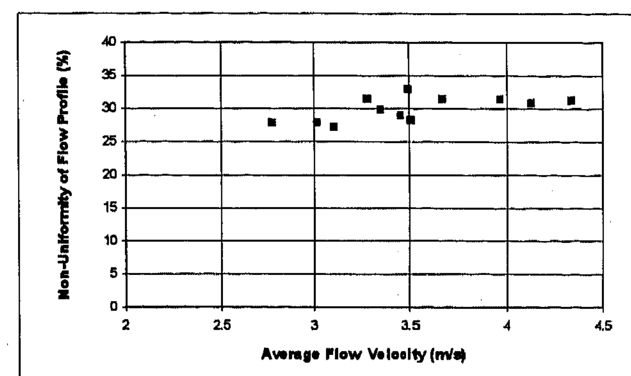


Figure 8: Non-uniformity (*K*) of velocity profiles produced by centrifugal fan at different flow rates.

Oil Concentration in Refrigerant

The presence of lubricant oil from the compressor flowing in the refrigeration circuit can alter the heat transfer process occurring in the evaporator. The mechanisms that cause this include alteration of the thermodynamic and transport properties, changes in the saturation profile due to pressure, temperature & concentration relations and to the retention of liquid refrigerant/oil mixture at outlet of the evaporator.

Measurements were made of the oil concentration in the refrigerant liquid line. The method used was similar to that described by McGovern [24] The results showed oil concentrations to be small, typically of the order of 0.2-0.4%. The effects of these quantities of lubricant on evaporation should be negligible.

Tube-by-Tube Refrigerant Temperature Profiles

Sixteen type K thermocouples were attached to the ends of various tubes in the coil. These facilitate detection of the onset and rate of superheating and study of the pressure profile (via saturation temperatures) along the coil. The thermocouples were secured mechanically and heat conductive paste was used to ensure good thermal contact with the tube surface. The temperature profiles along the coil were recorded using a sixteen channel 12 bit data logger, linked to a PC.

Since the quoted accuracy of type K thermocouples is relatively poor, the thermocouples used were individually calibrated over the temperature range of interest (-10° to +20°C). The measurement system errors were estimated to be ±0.2°C over the required range.

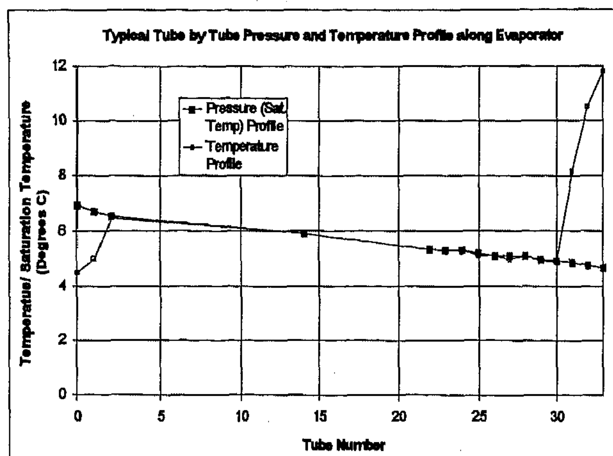


Figure 9: Tube-by-tube saturation pressure and refrigerant temperature profile along heat exchanger.

Results of the temperature profile from a typical test are shown in Figure 9. The temperature profile in the first few tubes is unexpected. The trend shown suggests the existence of a non-equilibrium sub-cooling effect, the refrigerant temperature being below the saturation temperature for the measured pressure. It is postulated that this is due to the under-pressure effect that has been observed for short tube and capillary tube refrigerant flow restrictors [25,26,27]. An important consequence is that local temperature differences are greater than might be anticipated, presumably increasing local rates of heat transfer.

Data Reduction

For the purposes of simulation verification the performance of the evaporator was evaluated in terms of the cooling capacity (kW) and the refrigerant pressure drop. The latter was determined by direct measurement of inlet and exit pressures. The cooling capacity $\dot{Q}$ was computed from the measured refrigerant mass flow rate $\dot{m}$ and change in specific enthalpy, i;

$$\dot{Q} = \dot{m}.(i_o - i_i) \qquad (2)$$

The specific enthalpy at inlet i_i is assumed equal to that of the liquid line and is evaluated at the prevailing liquid temperature. The inlet quality is calculated, assuming constant enthalpy across the expansion valve.

All other refrigerant properties were determined from measured data with the aid of correlations by Kabelac [21]. It has been assumed that the evaporator contains pure R-134a. The presence of oil or other impurities which may affect the properties of the working fluid are thereby ignored.

Experimental Results

Results are presented here from four distinct test groups, designated A, B, C & D;

Test Group A (Uniform Air Flow, TX Valve)

Measurements were made at 95 test conditions under conditions of uniform air flow, using the wind tunnel. These were based on one particular experimental rig configuration, using (i) a thermostatic expansion (TX) valve without external equalization, (ii) a relatively small compressor hot gas by-pass valve, and (iii) heating of ambient air before intake to the wind tunnel. All tests were confined to conditions of stable refrigerant mass flow, i.e. without 'hunting'. Refrigerant flow ranged from 1.7 to 14.5 g/s representing mass fluxes of 35-300 kg/m^{2}s. Typical superheat levels were 10-12 °C.

Test Group B (Uniform Air Flow, TEX Valve);

A total of 124 tests were carried out on a later rig configuration which used (i) an externally equalized expansion valve (TEX), (ii) a larger hot gas by-pass valve and (iii) no intake air heating. This arrangement allowed greater control of the

operating conditions and much lower levels of superheat (e.g. 5-7 °C) at outlet whilst retaining stable refrigerant flow conditions. As a consequence, overall temperature differences were generally lower and ambient water vapour condensation amounts were generally higher. Also, the nozzle hole size used in the TEX valve was smaller than that of the TX valve. This seemed to yield a more pronounced non-equilibrium refrigerant temperature profile in the first few tubes. Uniform air flow profiles were provided by the wind tunnel for this test group.

Test Group C (Non-Uniform Air Flow, TEX Valve);

97 additional tests were performed on the same refrigeration circuit and heat exchanger as Group B but with a highly non-uniform air flow, supplied from the centrifugal fan. The test data cover the range of average air flow velocities from 2.6-4.4 m/s, refrigerant mass fluxes of 180-320 kg/m^{2}s, air inlet temperatures of 17 to 23°C and relative humidity of 25-55 %.

Test Group D (Uniform Air Flow, TEX Valve);

40 tests were performed with the same rig configuration and similar conditions as for Group C but with uniform flow. In particular, care was taken to achieve similar total volumetric air flows or average velocities. The objective was to facilitate straight-forward comparison of results between uniform and non-uniform situations.

Experimental Comparison of Uniform vs. Non-Uniform Test Conditions.

Comparison of experimental results obtained from uniform and non-uniform test cases (Groups D & C respectively) are presented in Figures 10 and 11. These show overall heat transfer coefficients as a function of refrigerant mass flow rate at two different mean air velocities, 2.71 and 4.34 m/s respectively. These tests were all conducted with similar air inlet temperature and humidity conditions and equal volumetric flow rate (or average air flow velocity). The overall heat transfer coefficients are based on the following definition of temperature difference;

$$\Delta T = T_{air,in} - T_{ref,in} \qquad (3)$$

$$U_A = \frac{Q}{\Delta T . A_o} \qquad (4)$$

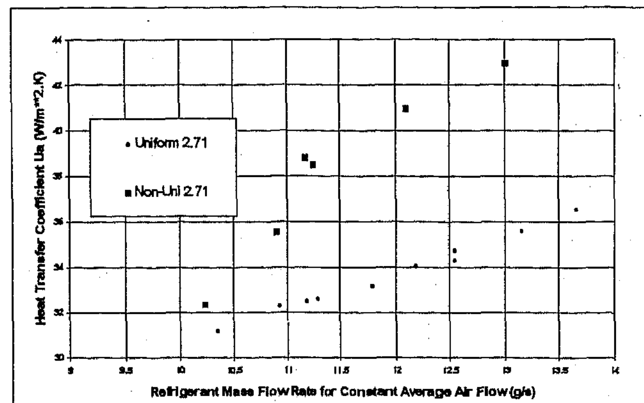


Figure 10: Experimental results showing overall heat transfer coefficients with uniform and non-uniform air flow at an average velocity of 2.71m/s. (Test Groups D & C). Air inlet temperatures were 18.9°C and 16.7° C for the uniform and non-uniform cases respectively. Relative humidity was 47% for all cases.

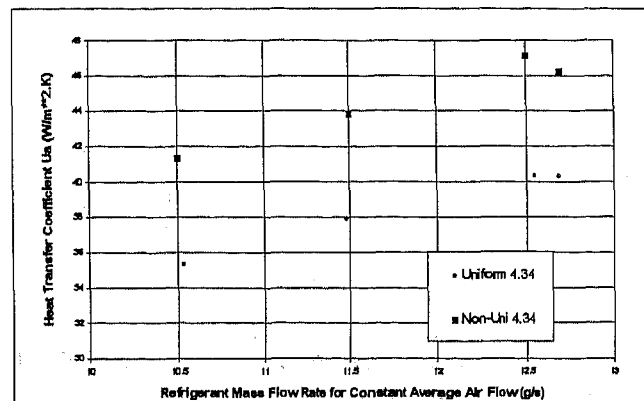


Figure 11: Experimental results showing overall heat transfer coefficients with uniform and non-uniform air flow at an average velocity of 4.34m/s. (Test Groups D & C). Air inlet temperatures and relative humidities were 21.0°C, 50% and 20.8°C, 44% for the uniform and non-uniform cases respectively.

These results show that the performance of the evaporator is superior by about 15% for the case of non-uniform air flow at a given volumetric flow rate. Note that the degree of outlet superheat was approximately constant for each of these test cases and that, for a given refrigerant mass flow, very similar amounts of heat transfer (kW) occurred. This was achieved with a lower temperature difference between fluids for the non-uniform flow conditions. The air inlet temperatures were determined by ambient conditions in the laboratory and the refrigerant evaporating temperatures were therefore proportionately higher for the non-uniform air flow cases.

The superior heat transfer characteristics observed with non-uniform flow seems to be primarily due to the enhancement of the air side heat transfer coefficients. These seem to be largely the result of the high levels of turbulence that are associated with the air flow pattern generated by the centrifugal fan.

Simulation results comparing uniform and non-uniform cases suggested that the performance of the evaporator would be *degraded* by the same distribution of air flow, if the effects of turbulence are ignored. With levels of non-uniformity in the region of 30%, the simulated performance degradation was of the order of 2 to 3.5%.

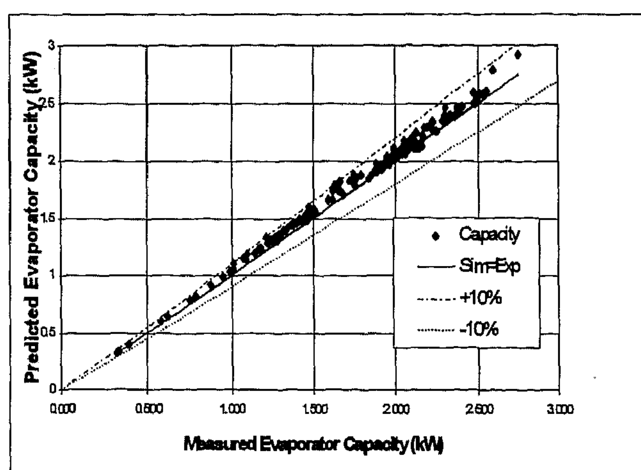


Figure 12: Simulated v measured evaporator capacity for uniform flow tests with TX Valve. (Test Group A, 95 test points).

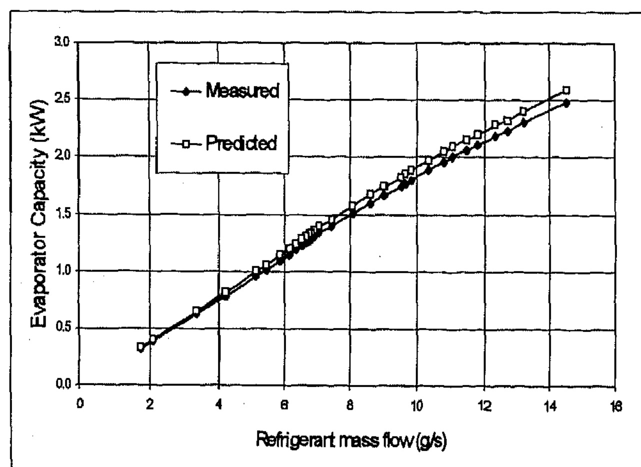


Figure 13: Simulated v measured evaporator capacity as a function of refrigerant mass flow rate at a constant uniform air flow velocity of 2.9 m/s, TX valve (33 data points, sub-set of Group A).

Comparison of Measured and Predicted Results

Test Group A (TX Valve, Uniform Air Flow)

Figure 12 shows a comparison of measurements and predictions of thermal capacity for the 95 test points of Group A. These cover a very wide range of operating conditions. All of the predicted data lies within -10% and +10% of the measured values, with an average over-prediction of 5%.

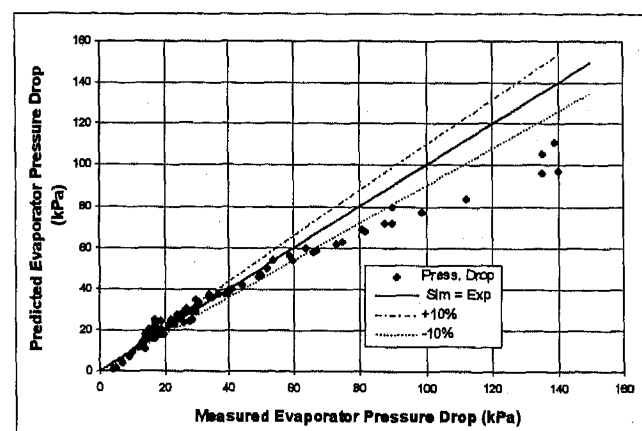


Figure 14: Simulated v measured refrigerant pressure. (Test Group A, 95 points, TX valve).

Results from a 33 point sub-set of the Group A tests are shown in Figure 13, where measured and predicted data are plotted as a function of refrigerant mass flow rate, for a constant mean air velocity of 2.9 m/s (uniform), with inlet air temperature of 22.5°C and relative humidity of about 29%.

The mean ratio of simulated / measured capacity is 1.044 for the test batch shown in Figure 13. Maximum and minimum ratios are 1.049 and 1.025 respectively and the standard deviation is .0005 about the mean. The agreement between modeled and experimental trends is clearly satisfactory. As discussed below, study of the refrigerant exit conditions yields some further insights into the reasons for the over-prediction.

Figure 14 shows the trends in both measured and predicted refrigerant pressure drop between evaporator inlet and exit for Test Group A. Absolute values of predictions are very close to the measured values for refrigerant mass flows up to about 12 g/s (250 kg/m^2s) and measured pressure drops below 70 kPa. Above this, the simulation underestimates the pressure drop significantly. The ratio of simulated to measured pressure drop has a mean value of 0.97.

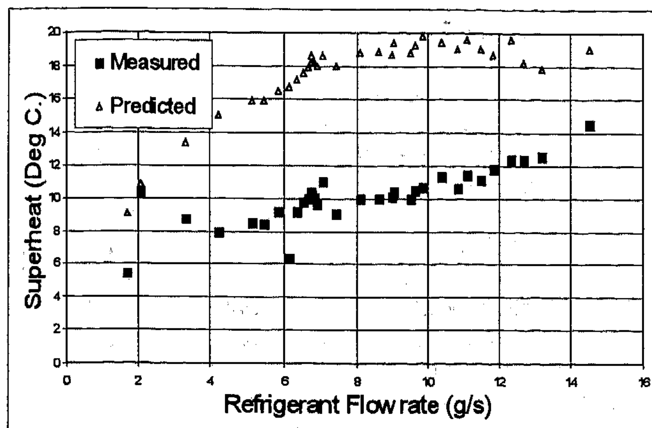


Figure 15: Simulated v measured refrigerant superheat at evaporator exit, as a function of refrigerant mass flow rate with constant uniform air flow velocity of 2.9 m/s, TX valve (33 data points, sub-set of Group A).

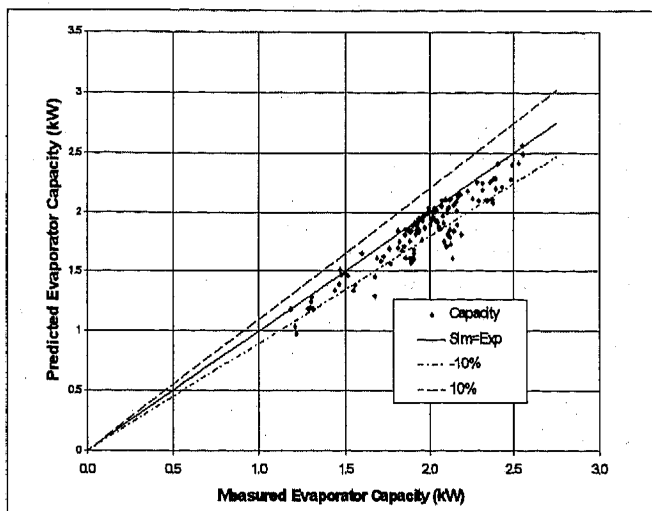


Figure 16: Simulated v measured evaporator thermal capacity for Group B Tests (124 points, uniform air flow, externally equalized TEX valve).

Figure 15 gives additional insight into over-prediction of thermal capacity for test Group A. Steady operation of the heat exchanger with the TX valve could only be achieved under conditions of relatively high superheat (typically 10-12°C). The 'minimum stable signal' for the TX valve is higher than that of the externally equalized TEX valve used in later tests (superheat 5-7 °C). As a consequence, a large number of tubes are dedicated to superheating in Test Group A and the heat exchanger is being used in an unrealistic and inefficient manner. Overall heat transfer coefficients in the later tubes are relatively low and the thermal capacity of the air side is far greater than required.

Test Group B (TEX Valve, Uniform Air Flow)

The results of the thermal performance predictions for test Group B (Figure 16) show a contrasting trend with those presented for Group A. The simulation under-predicts the data for this rig configuration, with an average predicted / measured ratio of 0.94 for the 124 Group B test points. The minimum ratio was 0.76 whilst the maximum was 1.04.

The refrigerant pressure drop was over-predicted by an average of 8% for Test Group B. Significant scatter is evident in Figure 17 but the more severely over-predicted points tend to occur at lower refrigerant flows. Note that significantly fewer tubes are involved with superheating in Test Group B than with Group A. The difference between air inlet and refrigerant evaporation temperatures is also lower for the Group B tests (typically 17-24°C for Group A and 14-19°C for Group B) and the ratio of air mass flow to refrigerant mass flow is lower, thus reducing the ratio of air-side to tube-side thermal capacities.

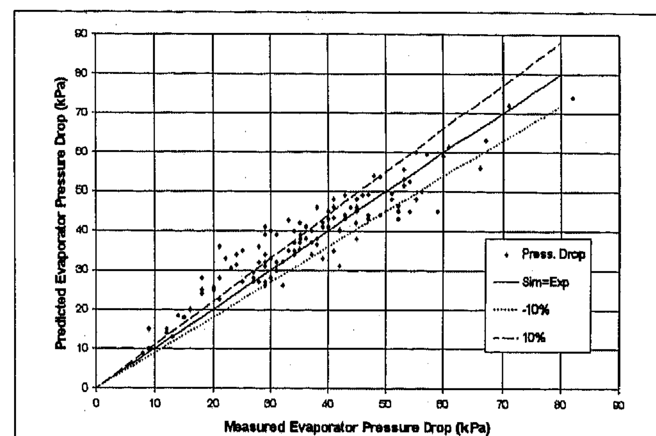


Figure 17: Simulated v measured refrigerant pressure drop across for Group B Tests (124 points, uniform air flow, externally equalized TEX valve).

The humidity levels hence the amount of condensation occurring are higher test Group B. Also, test cases involving larger temperature differences (and hence lower saturation temperatures) also tend to yield predictions which are in closer agreement with the measurements. The precise definition of log mean temperature difference used in the simulation for individual tube segments has a significant influence on the estimated heat flux. Uncertainties here, combined with possible inaccuracy in the air-side heat transfer correlation may be responsible for the poorer prediction accuracy. Also, the non-equilibrium inlet sub-cooling effect, illustrated in Figure 9, was observed to be more pronounced for tests in Group B.

This may not have been accounted for adequately in the simulation.

Test Group C (TEX Valve, Non- Uniform Air Flow)

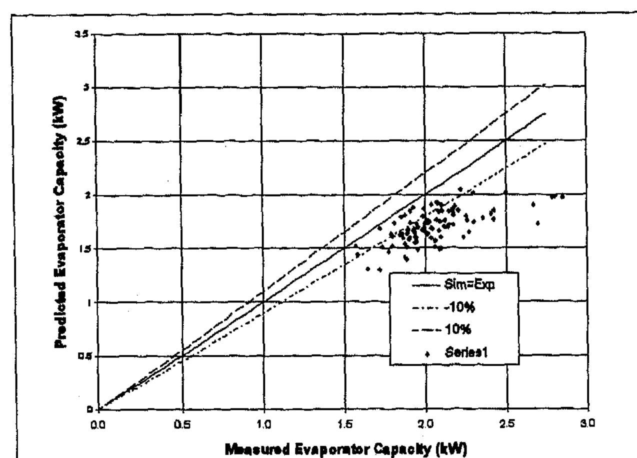


Figure 18: Simulated v measured evaporator thermal capacity for Group C Tests (97 points, non-uniform air flow, externally equalized TEX valve).

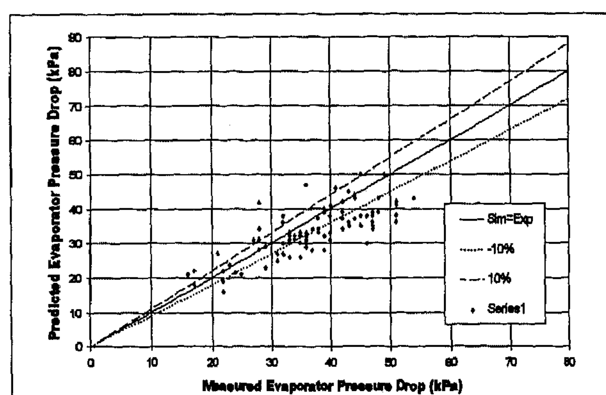


Figure 19: Simulated v measured refrigerant pressure drop across for Group C Tests (97 points, uniform air flow, externally equalized TEX valve).

Prediction of thermal capacity and pressure drop (Figures 18 & 19) for cases involving the non-uniform air flow profile created by the centrifugal fan is less satisfactory than that observed with the uniform flow cases. The simulation under-predicts consistently and the average ratio of predicted to measured capacity is 0.84, with minimum and maximum ratios of 0.64 and 0.98 respectively. The best matches between predictions and measurements were obtained for situations involving relatively high ambient air humidity. In these cases the latent capacity predicted is noticeably high despite experimentally observed levels of condensation being low.

CONCLUSIONS

An extensive series of tests have been carried out over a very wide range of operating conditions on a single circuit plate-finned tube type of refrigerant evaporator, exchanging heat between ambient air and R134a. Tests were confined to situations where the refrigerant flow was stable with time. Uniform and non-uniform air flow distributions were applied with the objective of studying the heat transfer effects of non-uniformity. Uniform flows were supplied from a wind tunnel whilst a centrifugal fan and duct were configured to provide forced flow with complex velocity and turbulence distributions. The latter were characterized experimentally using hot wire anemometry.

The experimental results revealed that heat transfer performances were 15% better with a turbulent non-uniform flow than those observed at the same volume flow of air with a uniform velocity distribution.

A computer simulation program is also being developed for ultimate use as a tool capable of assisting with design optimization in cases involving non-uniform flow. Extensive comparisons were made between predictions and measurements for a number of different rig builds. Predictions matched measurements satisfactorily for a group of tests (Group A) where the air flow was uniform and under conditions of relatively high average temperature difference between the two fluids. These experimental conditions were achieved with a thermostatic expansion valve which used internal pressure equalization.

A change to a different expansion valve configuration, which involved use of external pressure equalization, permitted experiments to be conducted at lower overall temperature differences and with higher ratios of air-side to tube-side thermal capacities (Groups B,C,D). Predictions matched measurements to a less satisfactory degree for these cases where a uniform air flow was retained.

The experimentally observed enhancement in heat transfer performance for non-uniform flow cases was not reflected in the simulation results (Group C). In fact, a slight degradation in performance was predicted. The reasons for these discrepancies are under investigation. The theoretical treatment of condensation on the air side requires many simplifying assumptions. However, the latent heat transfer has been kept to a minimum and the under-prediction observed with Group C cannot be adequately explained by this phenomenon alone. It

should be noted that the air-side heat transfer correlation used did not take account of the turbulence characteristics of the incoming air stream. This is thought to be the most significant reason for the observed enhancement under non-uniform conditions.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support provided by Thermo King (Europe) Limited and The Irish American Partnership.

REFERENCES

1. **Kitto, J.B. & J.M. Robertson.,** Effects of Mal-Distribution of flow on Heat Transfer Equipment Performance., *Heat Transfer Engineering*, Vol. 10, No. 1., 1989.
2. **Fagan, T.J.,** The effects of Air-Flow Maldistributions on Air-to-Refrigerant Heat Exchanger performance, *ASHRAE Transactions*, DV-80-7 No. 4. 1980.
3. **Chiou, J.P.,** Thermal performance Deterioration in Crossflow Heat Exchanger due to the Flow Nonuniformity, Transactions of the ASME, *Journal of Heat Transfer*, Vol. 100, pp. 580, 1978
4. **Chiou, J.P.,** The effect of Air Flow Nonuniformity on the thermal performance of Automobile Air Conditioning Condenser, *SAE Transactions*, (830542), 1984
5. **Chiou, J.P.,** The effect of Air Flow Nonuniformity on the thermal performance of Evaporator of Automobile Air Conditioning System, *SAE Transactions*, (840381),1985
6. **Berryman, R.J. & C.M.B. Russell.,** The effect of maldistribution of air flow on Air-Cooled heat exchanger performance., *Transactions of the ASME, Journal of Heat Transfer*, 1987.
7. **Rabas, T.J**, The Effect of Nonuniform Inlet Flow and Temperature Distributions on the thermal performance of Air-Cooled condensers., *Transactions of the ASME, Journal of Heat Transfer*, 1987.
8. **Kondo, F. & Y. Aoki.,** Prediction method on effect of Thermal Performance of Heat Exchanger due to Non-uniform Air Flow Distribution., *SAE Transactions*, (850041), 1985.
9. **Domanski, P.A.,** EVSIM- An evaporator model accounting for refrigerant and one-dimensional air distribution, NISTIR Rept. 89-4133. Gaithersburg, MD: National Institute of Standards and Technology, 1989.
10. **Domanski, P.A.,** Simulation of an Evaporator with Nonuniform One-Dimensional Air Distribution, ASHRAE Transactions, NY-91-13-1, 1991
11. **Wattelet, J.P., J.C. Chato, A.L. Souza & B.R. Christoffersen.,** Evaporative Characteristics of R-12, R-134a, and a Mixture at low Mass Fluxes., *ASHRAE Transactions:* Symposia, vol. 100(1), pp. 603-615, 1994.
12. **Gray, D.L. & R.L. Webb,** Heat Transfer and Friction Correlations for Plate Finned-Tube Heat Exchangers having Plain Fins, Proc. of the eighth international Heat Transfer Conference on Heat Transfer, Hemisphere Pub. Corp., vol. 6, pp. 2778-2750, 1986.
13. **Sheffield, J.W., R.A. Wood & H.J. Sauer, Jr,** Experimental Investigation of Thermal Conductance of Finned Tube Contacts., Experimental Thermal and Fluid Science, vol. 2, pp. 107-121, 1989.
14. **Pierre, Bo.,** Flow Resistance with Boiling Refrigerants- Part I, ASHRAE Journal, September 1964.
15. **McLinden,M.O., J.S. Gallagher, L.A. Weber, G. Morrison, D.Ward, A.R.H. Goodwin, M.R. Moldover, J.W. Schmidt, H.B. Chae, T.J. Bruno, J.F. Ely, M.L. Huber,** Measurement and formulation of the thermodynamic properties of refrigerants 134a (1,1,1,2-Tetraflouroethane) and 123 (1,1-Dichloro-2,2,2-Triflouroethane)., *ASHRAE Transactions*, 3282 (RP-588) 1989.
16. **Piao,C., H. Sato & K. Watanabe,** Thermodynamic charts, tables and equations for refrigerant HFC-134a., *ASHRAE Transactions*, 3518., 1991.
17. **Piao, C., M. Noguchi, H. Sato & K. Watanabe.,** An improved equation of state for R-134a., *ASHRAE Transactions*, vol. 100(1), pp. 358-366, 1994.
18. **Basu, R.S. & D.P. Wilson**, Thermophysical properties of 1,1,1,2-Tetraflouroethane (R-134a)., *International Journal of Thermophysics*, vol. 10, No. 3, pp. 591-603., 1989.
19. **Zhu, Ming-Shan., Jiang Wu & Yi-Dong Fu.,** New experimental vapour pressure data and a new vapour pressure equation for HFC-134a., *Fluid Phase Equilibria*, vol. 80, pp. 99-105, 1992.
20. **I.C.I. Klea,** Chemicals for a Better Future. Klea 134a Physical Property Data Sheet, S.I. Units, I.C.I. Klea, Runcorn, Cheshire, 1994.
21. **Kabelac,S.,** A simple set of equations of state for process calculations and its applications to R-134a and R-152a . *International Journal of refrigeration*, vol. 14, pp. 217-222, 1991.

22. **Jung, D. & R. Radermacher.,** Transport properties and surface tension of pure and mixed refrigerants. *ASHRAE Transactions: Research*, 3445, 1991.
23. **Tanaka, Y., M. Nakata & T. Makita**, Thermal conductivity of gaseous HFC-134a, HFC-143a, HCFC-141b & HCFC-142b., *International Journal of Thermophysics*, vol. 12, No. 6, pp. 949-963, 1991.
24. **McGovern, J.A.,** Low loss Measurement of oil concentration in a refrigerant-oil mixture in a Liquid Line, International Journal of Refrigeration, Vol. 12., 1989.
25. **Pasqua, P.F.,** Metastable Flow of Freon 12, Refrig. Eng., Vol. 61, 1084A-1088, 1953.
26. **Krakow, K.I. & S. Lin,** Refrigerant Flow Through Orifices, ASHRAE Transactions, Vol. 94, Part 1, PP 484-506 (3145), 1988.
27. **Aaron, D.A. & P.A. Domanski,** Experimentation, Analysis and Correlation of Refrigerant-22 Flow Through Short Tube Restrictors. ASHRAE Transactions, Vol. 96(1), pp. 729-742(AT-90-1-2), 1990.

970832

Optimization of Fin Louver Design Based on CFD

S. Bouzida and C. Mignot
Valeo Engine Cooling

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper presents an optimization methodology for the design of fin louvers of compact heat exchangers based on 2-D CFD calculations. The mesh generation and its automatization is described. The validation of the methodology and the results of this validation are presented. Finally, an example of application is given.

INTRODUCTION

Compact heat exchangers, like the ones commonly used in the car industry (radiators, condensers, ...), are based on the following technology: two fluids (air and water, air and oil or air and air) exchange energy by conduction and convection through surfaces (usually aluminum or copper). The first fluid, usually liquid, flows through pipes, which represent the direct surface of the exchanger and can be modelized by its internal thermal resistance (Y_{int}). The second fluid, usually air, flows externally around those pipes to exchange heat via convection which is modelized by the external thermal resistance (Y_{ext}). Looking at the order of magnitude of both resistances, it turns out that

$$Y_{ext} >> Y_{int} \quad \text{and} \quad Y_{ext} \cong 10\, Y_{int}$$

For that reason, the external surface is increased by a factor 10. This is done by adding a row of fins between the tubes of this exchanger (*fig.*1).

In most cases, this indirect surface represents between 70 to 90% of the total exchange area. More over, this surface is louvered to increase the flow circulation of the fluid. Again, depending on which particular heat exchanger we have, the louvered surface represents up to 90% of the indirect surface (*fig.*2). In other words, almost 80% of the total exchange surface is louvered.

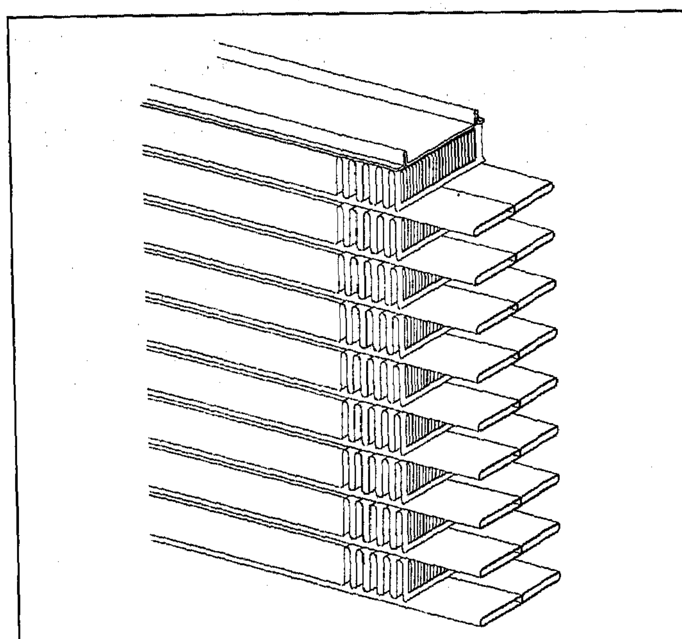


FIG. 1 - Fins in a heat exchanger

The thermal resistance is mainly external, so improving the performances of an exchanger implies improving its external surface and in particular, its louvers geometry (*fig.*3). In order to do so, we need to have a better understanding of the flow structure in between these louvers and on the phenomenon that occur at these small scales. Typically, the length of a louver is between 1 and 2 mm long and local experimental measurements are hard to set up. A lot of questions concerning the flow pattern still remain unanswered. For example, is the flow steady or unsteady (unsteadiness comparable to a Karman vortex street in the case of a flow past a cylinder)? Is there a boundary layer separation on each louver or only on a few of them? Is the recirculation of the flow good for the heat exchange or is it something that should be avoided?

Clearly, the lack of information on what is happening between the louvers makes predictions very difficult.

Unfortunately, too many parameters are involved in a proper study and the systematic design of a new prototype becomes too costly and time consuming.

We selected a C.F.D. approach in order to optimize the geometry and lower the cost and time of development. If it's true that CFD can turn out to be cheaper and faster and at the same time give satisfying results, it's also true that it can sometimes be the opposite: if it takes 3 persons on a full time basis to built a 3-D model, in a period of 6 months, with an expensive code and computer and, in the end, get a couple of results that never get validated, we can be very septic concerning the added value of CFD in a project development and the saving of time and money previously mentioned. To avoid this, we have limited ourselves to the study of a 2-D case. Of course, a 2-D modelization is not always possible due to a geometry which is purely three dimensional but with some reasonable assumptions, we show that in most cases of our applications, these assumptions hold. Also, the 2-D approximation might be less accurate than a full 3-D model but the gain in time is tremendous.

A full validation of the methodology was conducted using experimental results of two types of condenser fins. This methodology does not pretend giving the absolute performances of an exchanger that has never been tried but does give the rate of improvement of a new geometry as compared to the old one.

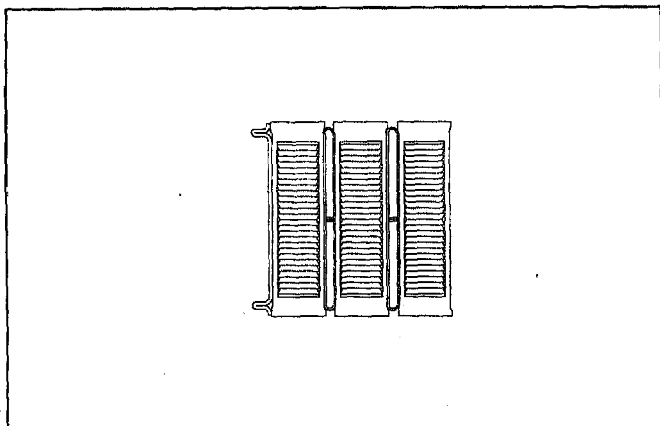


FIG.2 - louvered surface of fins

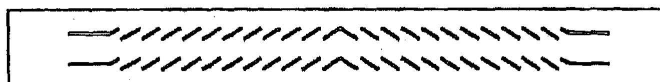


FIG. 3 - cross section normal to the fin surface

MESHING / PRE-PROCESSING

A REALISTIC MODELIZATION - In a C.F.D. study, the most time consuming part is with no doubt the pre-processing: creating the geometry and meshing the domain of interest is usually, depending of course on the geometry we are dealing with, a task that can take up to several weeks. So the idea of the study that we wanted to make was to reduce this time by as much as possible via an "automatic model generator" so that not only we could spend more time on the post processing which is the only part of interest really for the engineer but also extend the study to a bigger amount of cases and applications. At the same time, it would allow us a higher reactivity so that we can meet the development team needs in terms of time and schedule. Also the validation stage could be made in a shorter time by comparing the numerical results with the experimental results. Our main concerns and constrains were as follows:

-Create a model that would make a minimum of assumptions so that we simulate a more realistic geometry (too many modelizations are idealistic and therefore suffer the confrontation with the experiment).

-Adjust the meshing so that the boundary effects are well predicted without ending up with huge models that would take weeks in terms of CPU time.

-Reduce the pre-processing time by at least a factor 10.

To fulfill all these requirements, the idea of a parametric meshing program came. A first program written in FORTRAN was done. This program generates a serie of commands (stored in a file) which gradually builds up the model. At the beginning, we had to choose the parameters that we wanted to leave free for the user to change knowing that at this point the only difficulty was to integrate these parameters in the program. After having gathered all the potential parameters that were of interest for us in our applications (there are 18 of them), we were ready to create the meshing itself built on the geometry defined by the vertices. Those 18 parameters or "degrees of freedom" that are completely defining the geometry are: (*fig.* 4, 5 and 6)

- The flat plate length : that's the part of the domain which is not louvered.

- The number of louvers

- The stretching coefficient of the first louver: for the case where we want the first louver to be longer or shorter. Indeed, in almost all cases, the very first and very last louver of the geometry is half the size of a louver. This is due to the fact that the first and last louvers are cut on only one side. Nevertheless, it is also possible to design the louvers in such a way that these louvers are actually longer or shorter than the standard half louver length. (figure 5 shows the case where this louver is equal to the louver length). This parameter allows the user to control this particular length.

- The stretching coefficient for the central louver: for the case where we want the central louver to be longer or

shorter. This is exactly the same thing than seen previously. This parameter allows the user to have control of the central louver length.

- The core thickness

- The fin pitch: distance between 2 fins

- The gage: thickness of the metal sheet

- The mean angle of the first row of louvers

- The mean angle of the second row of louvers

- The half-louver angle for the first row

- The half-louver angle for the second row

- The central-louver angle for the first row

- The central-louver angle for the second row

- The central flat length

3 more parameters were defined for the conditions of the problem itself:

- Upstream mean velocity

- Wall temperature

- Upstream temperature

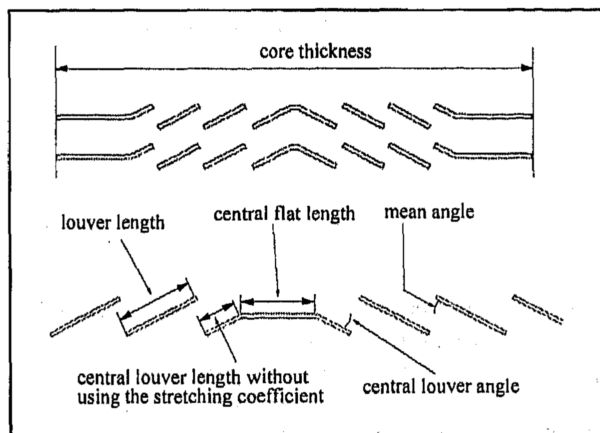


FIG. 4 - geometrical parameters

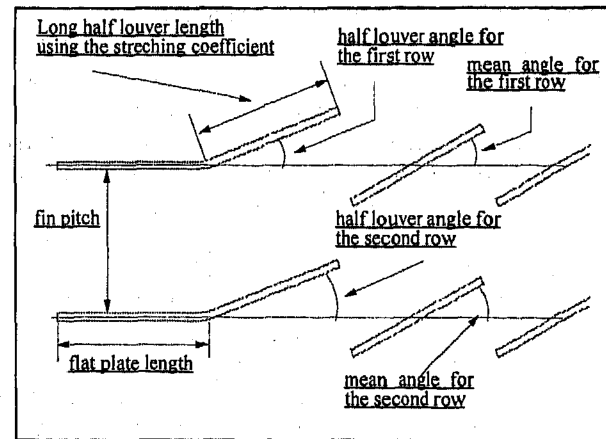


FIG. 5 - geometrical parameters

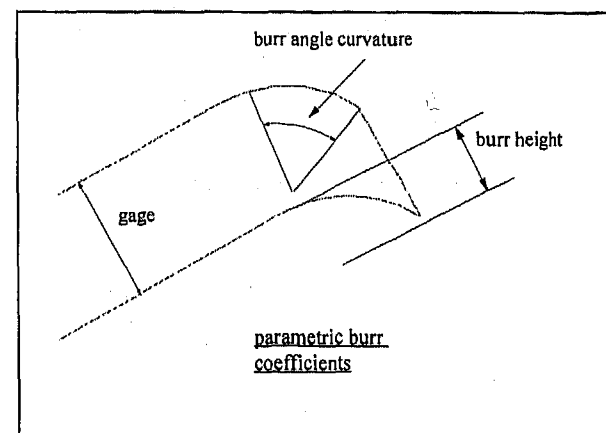


FIG. 6 - parameters related to the burrs

Now, other parameters were introduced. Their purpose was, as we mentioned previously, to fit reality as much as possible. A close observation of the production helped us identify the potential degradation modes. Taking into account these degradation modes would quantify their effects and give an estimation on the consequences they have on the performances of the exchanger. Those parameters and what they represent can be described as follow:

- The burr angle curvature: when we represent the burrs, they are represented by a radius (which is the burr height itself) and the burr angle curvature which gives the rate of "distortion" of the burr.

- The angle dispersion by default (ADD) and angle dispersion by excess (ADE): We have noticed that during the fin manufacturing, a small dispersion of the angle distribution could occur. Indeed, if on "paper" the angle was, say α, a close observation of what the angle really was after manufacturing showed that the angle was never exactly α. In fact, the value of the angle was $(\alpha + \Delta\alpha)$ where $\Delta\alpha$ could be positive or negative. A general law of this dispersion couldn't be drawn, in other words, if in the first row of louvers the first 3 louvers had a negative $\Delta\alpha$ and the next 4 a positive $\Delta\alpha$, the second row of louvers would show a totally opposite distribution. We have also noticed that the magnitude of this dispersion

was closely related to the age of the industrial tools, the older the tools, the bigger it was. The conclusion was that the angle dispersion was random and the question that arose from this observation was the influence of this angle dispersion. Therefore, we have introduced those two parameters which represents precisely this random angle dispersion. The user can, after having selected all the angles of his geometry, decide on the magnitude of the angle dispersion but has no control on which louver is going to be affected by the dispersion and by how much within the specified range. As an example, let's consider a heat exchanger whose fins have 2 blocs of 6 louvers. "On paper", all the angles were given equal to 25°. Looking closely at how the prototype really turned out to be, the angle measurements gave respectively:

24.0° 25.0° 23.0° 24.0° 27.0° 25.5°

and from the same prototype another measurement on different louvers gave:

23.5° 26.5° 24.0° 23.0° 25.5° 25.0°

now, the program that we present takes this random distribution into account with the two parameters mentioned. The user has to set a value for the mean angle and then set the values of the angle dispersion by default and the angle dispersion by excess. In this particular example, those values were:

mean angle = 25.0 °
ADD = 2.0 °
ADE = 2.0 °

Once the model was built, the angle distribution was:

24.2° 26.2° 25.3° 23.7° 24.8° 25.9°

At this stage, the user knows that the model represents the manufacturing dispersion, knows its order of magnitude but does not know exactly beforehand which louver is affected and by how much exactly. Of course, it is also possible to modelize a geometry with no dispersion at all. But the idea of modelizing reality by introducing a random angle distribution according to the dispersion specified by the user is, although it sounds like a paradox, a way to stick more to reality to a certain extent. The manufacturing dispersion (even small) exists? Let's modelize it.

- The louver burr: Another imperfection of manufacturing is the burr on the edge of the louver created when the tools cut the metal sheet (*fig.*7). We have noticed here again that the older the tool, the bigger the burr and the question was to find out what consequences they had. Were they improving the efficiency of the exchanger or were they making things worse? In other words, is it really necessary to spend money on maintenance or it's not worth it after all. We tested them and we found out that some burrs were "good ones" and others "bad ones". Without really understanding why, we thought that taking them into account in our modelization was necessary.

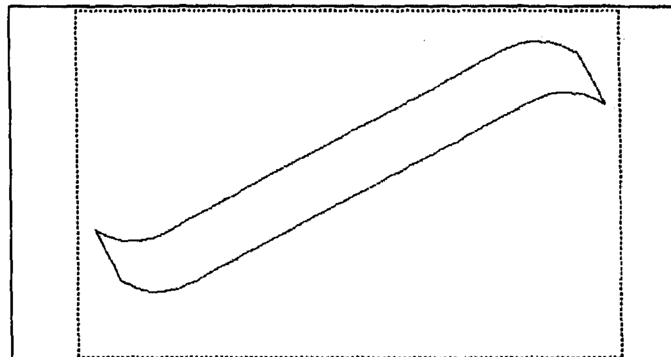


FIG. 7 - modelized louver with burr

- The burr height by default (BHD) and burr height by excess (BHE): Here, we have exactly the same approach than before. The burr height is random and so we have introduced its random distribution in the modelization. The user can specify the maximum height that the burr can have above its reference value and the maximum height that it can have below its reference value. But the actual value of the burr height is not directly none by the user. Again we have modelized the unknown of manufacturing imperfections.

- Two rows of louvers: Another aspect of the degradation of the manufacturing tools is the variation of the angle dispersion from one row to another. Indeed, we have observed that in a few cases, the fins would have one row of louvers at a constant angle α and a second row at a constant angle β with $\alpha > \beta$ (*fig.*8). This is due to the fact that one roll is a driving one and the other roll is a driven one, and so the stresses in the metal are different from one row to the next one.

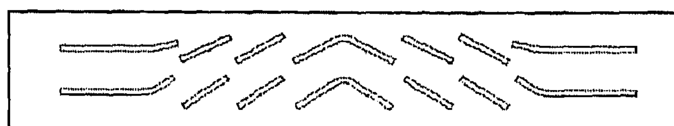


FIG. 8 - Two rows of louvers with angle dispersion

Therefore the modelization of two rows of louvers would allow us to anticipate the behavior of an exchanger in this particular degradation mode. This is the reason why all the simulations that we present in this paper are based on two rows of louvers instead of one

THE MESHING - Five more parameters were introduced and are related to the meshing itself. The idea was to allow the user to modify certain aspects of the meshing by just modifying the value of these parameters.

- The number of cells in the boundary layer (NCBL): To predict the heat transfer at the boundary with accuracy, it is necessary to get an accurate solution of the velocity field itself. Indeed, the heat transfer coefficient h, just like

the friction coefficient Cf, is proportional to the velocity gradient at the boundary. Therefore, the quality of the mesh in this region is of primary importance. This implies that close to the boundary, the mesh must be finer than in other parts of the model where the mesh can be left coarse. The program we present makes a first calculation of the boundary layer thickness based on a blasius solution for a flat plate. It is given by:

$$\delta = 5\sqrt{\frac{\nu L}{U}}$$

where L represents a characteristic length scale which for us could either be the pitch or the louver length and U the upstream velocity. Then we build a "skin" of cells, of thickness d, surrounding the louver in which we put the "appropriate" number of cells (*fig.*9&10). This "appropriate" number of cells was determined comparing the solution given by the code in the simple case of the flow over a flat plate with the exact solution. But, instead of fixing its value once for all, we decided to leave this new parameter free so we could adjust better to each new problem.

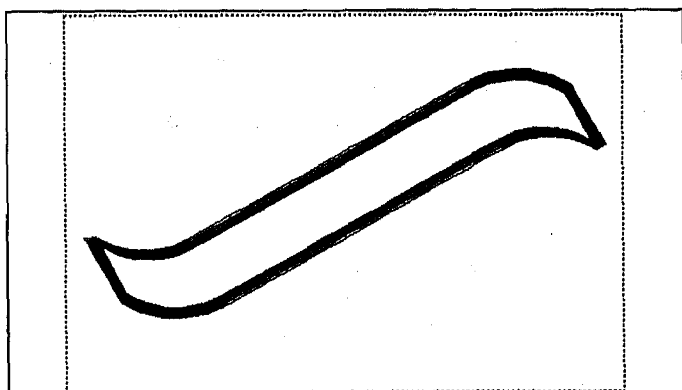


FIG.9 - layer of cells around the louver

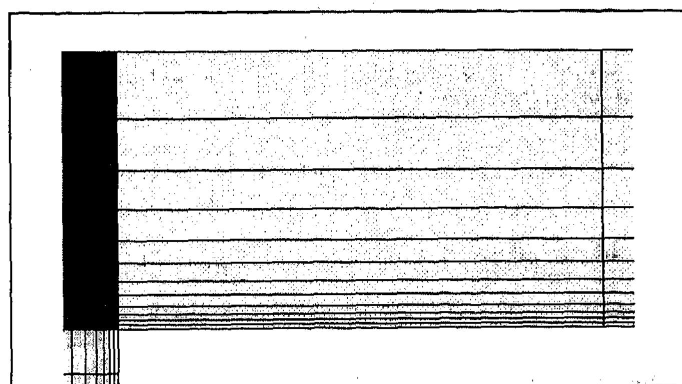


FIG. 10 - Non-linear repartition of cells within the boundary layer.

-The boundary layer thickness coefficient (BLTC): We saw that the last parameter could allow us to put a large number of cells in the boundary layer (B.L.) but the problem is that the numerical models usually do not accept transitions from large cells to small cells which induce numerical instability and convergence problems. Therefore the trick was to increase the value of the B.L. thickness by a factor chosen by the user (usually between 1 and 3) and at the same time keeping the number of cells within the B.L. constant. This way the size of the cells is increased and the transition from small cells to large cells is smoother. But again, this technic had the disadvantage of increasing the size of the first cell at the boundary, and therefore to lose accuracy. This problem was avoided by introducing the next parameter.

- Non-linear repartition of cells within the boundary layer (LRC): This parameter is simply a function defined in the code and allows a non linear repartition of cells along a direction. This is possible with practically all meshing code but here this parameter is fixed by the user in a file and the non-linear distribution of cells is applied throughout the model. In practice, it means that close to the boundary the cells are very small (which is good for the accuracy) and at the edge of the B.L the cells are almost as big as the external coarse mesh (good for the smooth transition).

- Number of cells along the curvature: The burr that we have is defined by its height and its radius of curvature. If its height is small its radius of curvature is small. This parameter simply defines the number of cells that we should put along the curvatures.

- The concentration of cells throughout the model: This last parameter allows the user to increase the total number of cells uniformly throughout the model. Increasing its value implies refining the mesh. Its purpose is to increase or reduce the number of cells and therefore the CPU time depending on the level of accuracy expected.

Some of these parameters may not sound of primary importance but they turned out to be very useful. With very little effort, we could determine things like the effect of the meshing on the solution, the effect of the number of cells within the boundary layer, avoid quite easily the problems due to the size of cells and so on...The important thing is that now that the program is written, very little time is spent on building the model. The user just has to give the values of the different parameters in a separate file, the program creates the files of the model in 0.8 second (CPU time). And then it takes the code another two hours (more or less depending on the number of nodes) to completely achieve the model.

EQUATIONS, BOUNDARY CONDITIONS AND FURTHER ASSUMPTIONS - We solve the Navier-stokes equations, with the viscous term, in steady state. Two types of flow are considered: Laminar flow and turbulent flow. In the case of the compact heat exchanger the flow is for low inlet velocity ($\cong$ 2 m/s) laminar with Re < 1000 and for higher flow rates where the inlet velocity is about 6 m/s the flow can be turbulent with Re > 1000. For this last case, the turbulent model used is a K-ε with a mixing length assumption. It is

argued by a lot of people that the K-ε is known to be wrong but we decided not to spend time on validating which model was best adapted because for one it was difficult to check with experiment since local measurements are almost impossible and for two we were not looking at absolute values but at the differences between two values given by the same turbulent model. Nevertheless, all the results presented in this paper were established using a laminar flow. The reason is that the Reynolds numbers are not "big" enough and even in the range 1000 - 2000 it is not obvious that turbulence plays an important role. In our applications, we believe that we are in the transition between a laminar flow and a turbulent flow and that the use of a turbulent model might not be well suited. Turbulence increases the heat transfer but using a turbulent model if the flow is laminar might lead to wrong results. The louvers are there to start a boundary layer at their leading edge where the heat transfer coefficient is the most important (*fig.*11-a). This is precisely this property of the heat exchange coefficient that we use when we increase the heat exchange by increasing the numbers of louvers. This is also clear when we represent the heat flux along the louvers as shown in figure 11-b where we see the areas and the parts of the louvers that are thermally speaking the most active : the louver leading edges.

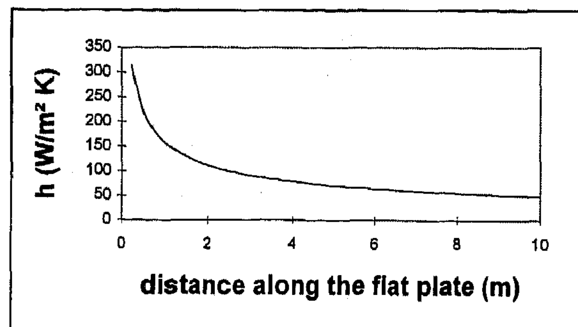


FIG. 11-a - heat exchange coefficient along a flat plate

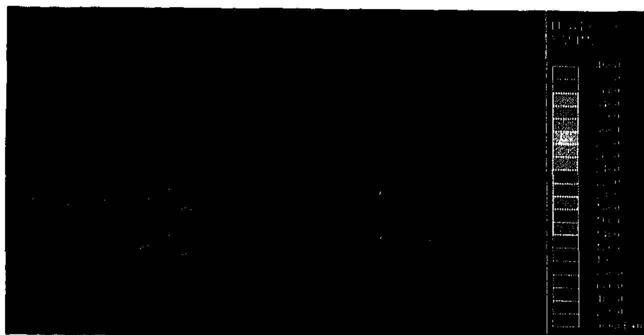


FIG. 11-b - heat flux along the louvers

We also solve the energy equation to get the temperature distribution in the flow field. Moreover, the flow is considered incompressible so that density is not a function of pressure but a function of temperature only:

$$\rho = f(T)$$

There are 5 types of boundary condition in the model:

- Inlet boundary condition: we prescribe the upstream velocity

- Outlet boundary condition: we insure the conservation of mass

- Cyclic boundary condition: traduces the periodicity of the model

- Symmetric boundary condition: condition for the modelization to be 2D

- Wall boundary condition: with no-slip + fixed constant temperature (*)

(*) the fixed temperature at the wall is an assumption. Unless we decide to mesh the metal, which would imply solving the conduction equation in the metal as well, we couldn't prescribe what the heat flux was at the boundary otherwise we would have been giving the solution in the B.C. since the heat exchange is an unknown that we are looking for. But in our case, the biot number (Bi) is small enough which implies that conduction is much greater than convection so that in first approximation we can consider the fin at a constant temperature.

$$Bi = \frac{L\,h}{k}$$

L = characteristic length scale
h = heat transfer coefficient
k = conductivity in the metal

(example: l=0.001 h=100 k=200 $\Rightarrow$ Bi= 0.0005 << 1)

Also, from experiment we know that the fins efficiency is close to 1 which again confirms this assumption. Another assumption is to claim that the flow is 2D. Indeed, it is true when the ratio

$$r = \frac{(\mathit{fin\ pitch})}{(\mathit{fin\ hight})} \ll 1$$

(example: pitch = 1.3 mm and fin height = 11

$\Rightarrow$ r = 0.118)

this approximation is no longer valid for r > 0.3 or 0.4 where the flow is purely 3D (*fig.* 12).

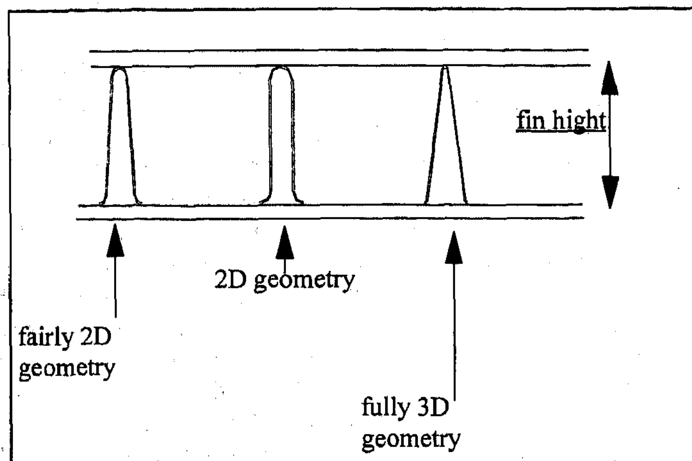


FIG. 12 - geometries qualifying for the 2D modelization

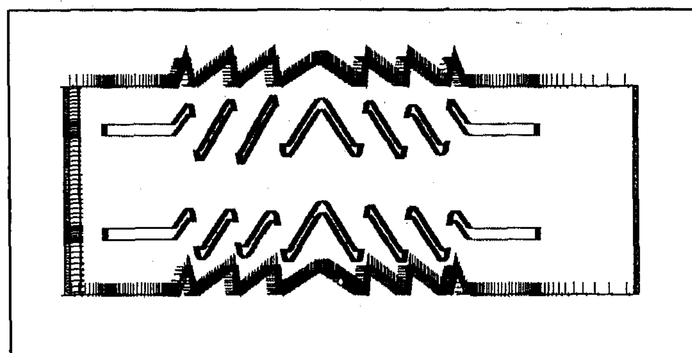
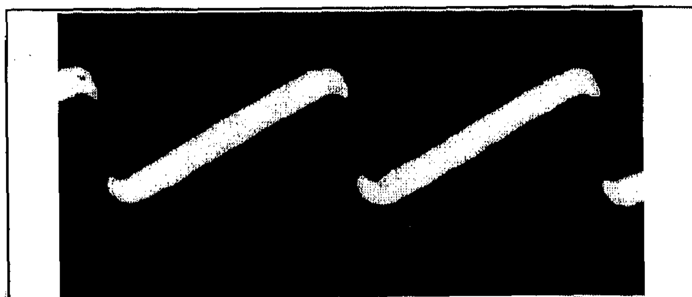
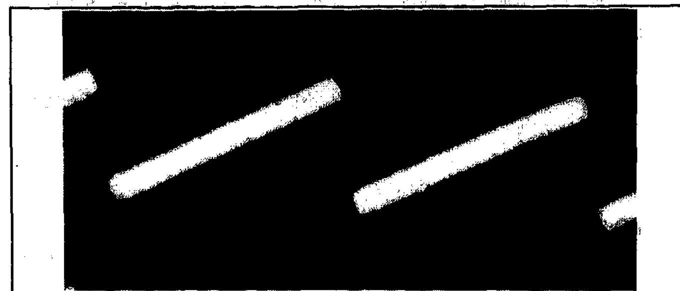

FIG.13 - boundary conditions

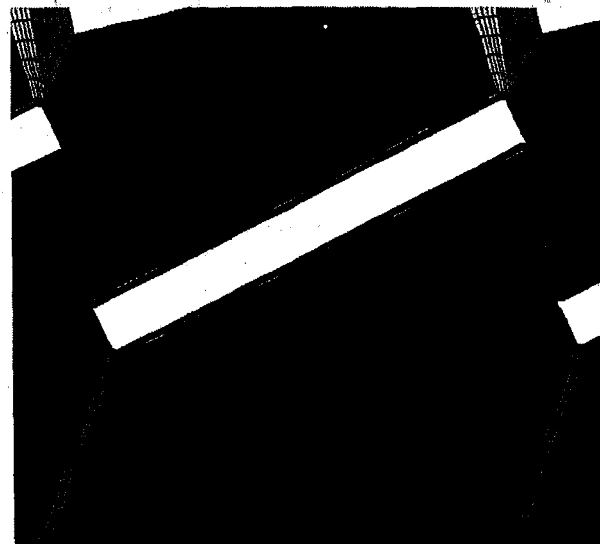
Figure 13 represents the boundary conditions. For convenience and clarity the symmetric boundary conditions are not represented

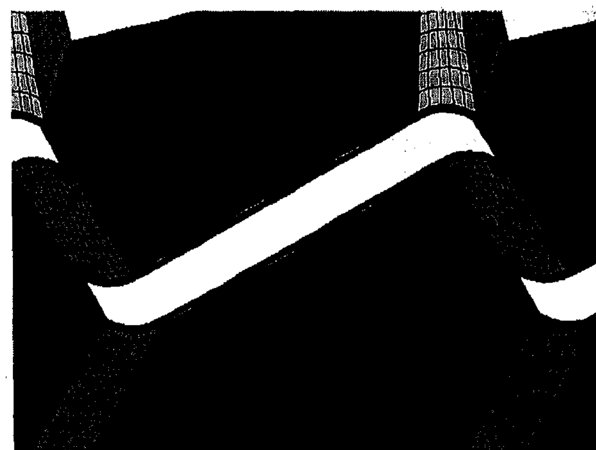
METHODOLOGY, VALIDATION AND RESULTS

FROM EXPERIMENT TO MODELIZATION-The validation was done on two different sets of condensers. We present here the results for condensers in the 22 mm range. We purposely took a "clean" condenser (meaning with no burrs) and a "burred" one (meaning with quite important burrs on the louvers that we did using very old tools). (figures 14 and 15 are photographs of these louvers in both cases, whereas Figures 16 and 17 represent the modelized geometry and the mesh around it). We cut them and reduced them to one pass having 12 tubes each.


FIG. 14 - louvers with burrs (real geometry)


FIG. 15 - louvers without burrs (real geometry)


FIG 16 - modelization of a louver with no burr


FIG.17 - modelization of a louver with burrs

We tested these condensers as if they were radiators and ran the analysis the same way we would have for a radiator. Results showed that they there was a non negligible difference in terms of performances between the two. Figure 18 shows a difference of 10 % on the friction coefficient and 7 % on the value of the heat exchange coefficient.

	Cf	h
clean	0.1223	221.5
burred	0.1348	205.5
Δ	+ 10 %	- 7 %

FIG. 18 - experimental results of the 22 mm condenser

note: all the results presented are for the same Re number (Re = 900)

The goal of the validation was to find this differences numerically.

METHODOLOGY - Now, it is important at this point to notice that since we have a 2-D modelization, we are not expecting results to match experiment in absolute values. Nevertheless, we do expect to find results going the right way in terms of general tendency. The methodology can be described as follow: We start with a geometry which is well-known, that's to say we know its dimensions, all its geometric characteristics, we have all its results from the test bench from which we have computed all its dimensionless coefficients, namely the friction coefficient (Cf) and the heat exchange coefficient (h). Those coefficients are determined using global measurements on the test bench such as the total pressure drop, the inlet and the outlet temperature of the liquid in the exchanger, the air velocity at inlet and so on. Then, we modelize the indirect surface of this exchanger and we run a first calculation which gives us a value for Cf and h. we call these values: Cf_{2D} and h_{2D} Then we modify the geometry, whatever this may be, and we modelize this new geometry. A new calculation gives us new values for Cf and h, we call them: Cf_{2D}^* and h_{2D}^* Comparing the last 2 sets of results gives us a variation for these coefficients that we write: ΔCf_{2D} and Δh_{2D} Finally, the last step is to find the variation of the performances of the real geometry in terms of the variation of the performances of the numerical variation (the whole procedure is described in figure 19). If we take into account the fact that only a percentage of the total exchange surface was modelized we can establish the following relations:

$$\Delta(Cf) = F1\,(\Delta Cf_{2D})$$
$$\Delta(h) = F2\,(\Delta h_{2D})$$

The 2 functions F1 and F2 are deducted from the direct confrontation of the experimental results with the numerical results during the validation of the method.

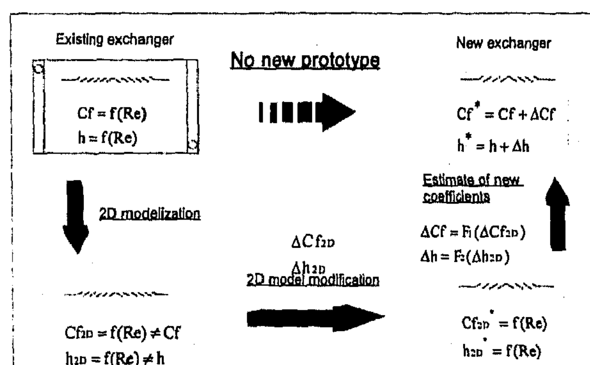


FIG. 19 - general methodology

POST-PROCESSING - That's one of the most important part but also one of the most difficult of the whole modelization process. The code will give us all the quantities that we want but the interpretation and manipulation of these datas is certainly not a trivial task. In particular, we must find a way to express the Re number, The hydraulic diameter D_H , the friction coefficient Cf and the heat exchange coefficient h so that it would be relevant with the experimental datas. As far as the hydraulic diameter is concerned, we calculate it the same way we do for the 3D geometry. It has to be homogeneous to a length. It is the ratio between the passing surface and the "wet" perimeter. The passing surface is a surface through which the fluid flows and the "wet" perimeter is the perimeter of that surface which is directly in contact with the fluid. More or less, it has the same order of magnitude than the louver length. Its calculation is straightforward but tedious. It includes complicated geometrical considerations that are not worth including in this paper.

$$D_H = \frac{4 \times passing\ surface}{wet\ perimeter}$$

The Reynolds number is built on that hydraulic parameter and is defined as follow:

$$Re = \frac{U\ D_H}{\nu}$$

The friction coefficient is a dimensionless expression of the pressure drop. The code does compute a value for Cf but the results were not satisfying. Instead, we used the pressure drop given by the code and we divided it by the relevant datas:

$$Cf = \frac{(\Delta P)\,(\frac{D_H}{4})}{(\frac{1}{2}\rho U_l^2)\,L}$$

where L represents the relevant length scale for the pressure drop: the core thickness. The factor 4 is just there to be consistent with the definition of the hydraulic diameter. U_l is not the upstream velocity but the average velocity between the louvers. Indeed, the velocity U_l is greater than the upstream velocity because the passing section for the fluid decreases as the flow enters the compact exchanger (conservation of mass) and the difference between the two is not negligible. Also, strictly speaking, the density is a function of temperature but since its variations are very small, we neglect its variations and we assume its value to be constant and we take it at the average temperature (T_a) which itself is defined as:

$$T_a = \frac{T_\infty - T_w}{2}$$

Finally, the h coefficient is a quantity directly given by the code. Just like the pre-processing, the post processing is completely done by a program written in FORTRAN

which computes all the quantities just mentioned, except the values given by the code itself.

VALIDATION AND RESULTS - The numerical results are presented in figure 20.

	Cf	h (W/m² K)
burred	0.206	168.0
clean	0.170	172.3
Δ (%)	+21	- 2.5

FIG. 20 - numerical results of the 22 mm condenser

As these Figures show, the absolute values are not directly comparable which is not surprising since the direct surfaces are not taken into account and because of the few assumptions that we made along the way. More interesting though is to compare the variations between the clean and the burred geometries for both the experimental results and for the numerical results (see figure 21).

	Δ(Cf)	Δ(h)
experimental	7.5 %	7.7 %
numerical	20 %	2.5 %

FIG. 21 - comparison of the experimental and numerical variations

We see that the numerical predictions overestimate the friction coefficient but underestimate the gain in heat exchange. But in any way results go the right direction. The visualization of the flow field helps understand why the burrs are degrading the performances. We see precisely where there is the boundary layer separation on the edge of the louvers and the surfaces that are not thermally active clearly appear. (figure 22 shows the iso-velocity lines for a "clean" exchanger i.e. with no burrs and figure 23 shows the iso-velocity lines for a "burred" exchanger i.e. with important burrs and 23-b shows the recirculation due to the burr).

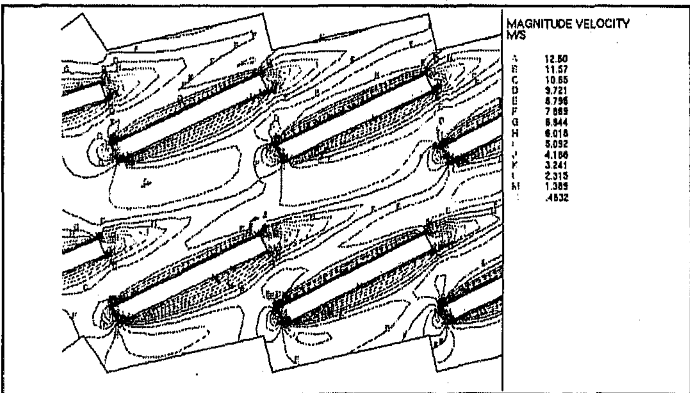


FIG. 22 - iso-velocity lines around louvers with no burrs

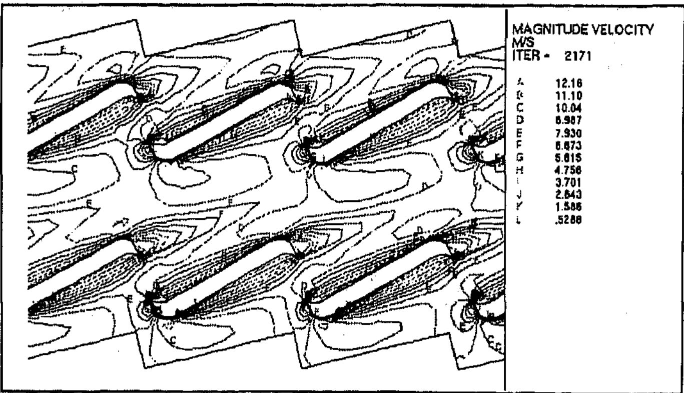


FIG. 23 - iso-velocity lines around louvers with burrs

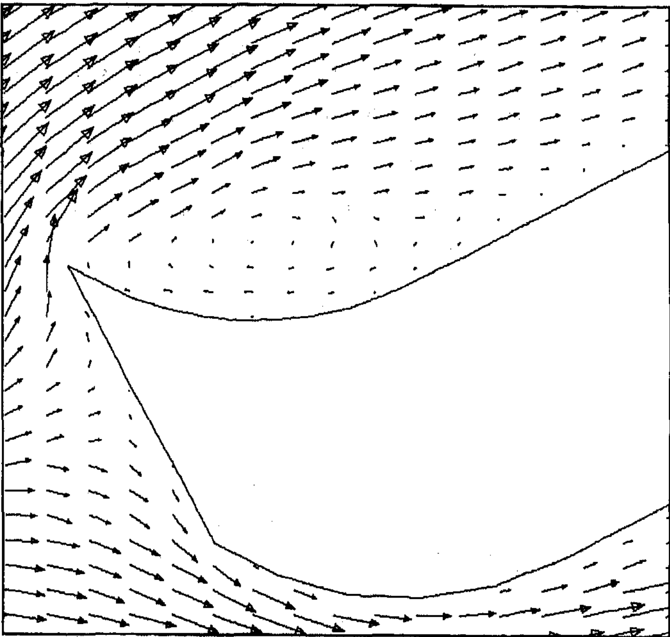


FIG. 23-b flow recirculation around the burr

The conclusion is that we can express the variations of the experimental results in terms of the variation of the numerical results in a general form:

$$\Delta(\text{experimental results}) = F\ (\ \Delta(\text{numerical results})\ )$$

where the function F takes into account the ratio of the indirect modelized surface over the total surface of the exchanger. Finding an analytical expression for F was possible only after having confronted a lot of numerical results with the experimental results. Using the same notation than before, we can write:

$$\begin{cases} \Delta(Cf) = F_1(\Delta Cf_{2D}) \\ \Delta(h) = F_2(\Delta h_{2D}) \end{cases}$$

or,

$$\begin{cases} Cf^* = Cf_{ref} + F_1(\Delta Cf_{2D}) \\ h^* = h_{ref} + F_2(\Delta h_{2D}) \end{cases}$$

where the upper script * denotes the value of the new geometry and the subscript "$_{ref}$ denotes the value of the known geometry.

$$\begin{cases} F_1 = \dfrac{(\ louvered\ surface)}{(total\ surface)} \times C_1 \times \Delta Cf_{2D} \\ and \\ F_2 = \dfrac{(total\ surface)}{(\ louvered\ surface)} \times C_2 \times \Delta h_{2D} \end{cases}$$

where C1 and C2 are constants which values are improved after each calculation we make. This form for the function F is explicitly independent on the Re number but implicitly depends on it since the two quantities ΔCf_{2D} and Δh_{2D} are themselves implicit functions of Re (they are the numerical solution of the Navier-stokes equations). The reason why F1 is proportional to the ratio between the indirect surface over the total surface is that most of the external pressure drop in a compact heat exchanger is due to the louvers and since what we modelize represents only the louvers, it is not surprising to find a higher prediction in pressure drop. The final form for the expression of the new geometry coefficients is:

$$\begin{cases} Cf^* = Cf_{ref} + \left[\dfrac{(louvered\ surface)}{(total\ surface)} \times C_1 \times \Delta Cf_{2D} \right] \\ and \\ h^* = h_{ref} + \left[\dfrac{(total\ surface)}{(louvered\ surface)} \times C_2 \times \Delta h_{2D} \right] \end{cases}$$

The next plot (*fig.24*) is the temperature distribution in between the louvers for a normalized velocity of 5 m/s (Re ≈ 900)

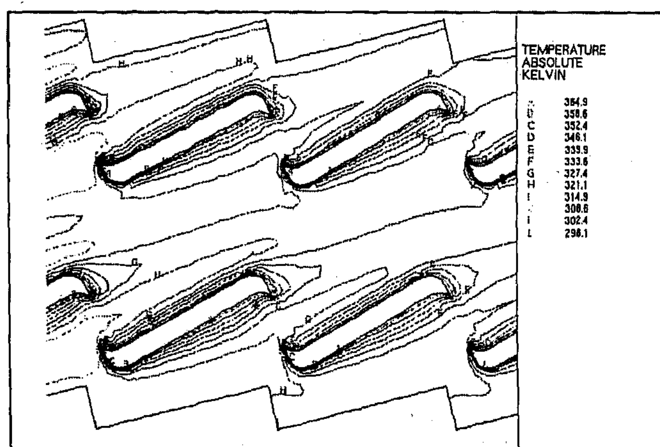


FIG. 24 - temperature distribution around louvers with burrs.

AN EXAMPLE OF APPLICATION

This was applied to the design of fins in the case of a mechanical radiator. Three different geometries were tried numerically and in the end one was chosen on the numerical results basis. One of the objective of the study was to determine the number of louvers required given a core thickness and whether it was necessary or not to leave unlouvered the central part of the surface. (The three different geometries are represented on figure 25). The first geometry has twelve louvers of 1 mm and no central flat length. The second geometry has the same number of louvers, each louver having the length of 1 mm but there is a central flat length of 0.9 mm. The third geometry has 14 louvers of 0.9 mm and a central flat length of 1.5 mm. The fin pitch, the core thickness, all the angles and the gage have the same value for the three models. So, this study allowed us to measure the influence of the central flat length and of the number of louvers in this particular example.

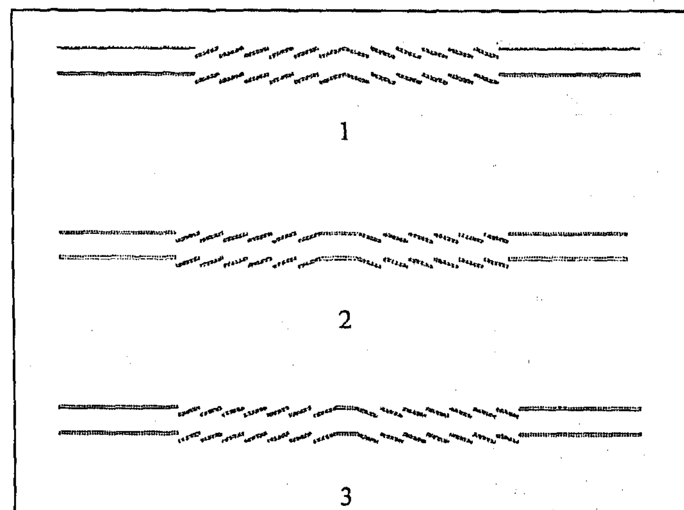


FIG. 25 - Three different modelized geometry.

Unlike the methodology previously described, we do not start from a known geometry but still we could determine which geometry was best adapted for this particular project. The numerical results were: (*fig.* 26)

	geometry #1	geometry #2	geometry #3
Cf	0.149	0.135	0.1490
h	294	295.1	294

FIG. 26 - numerical predictions for a mechanical radiator.

The best geometry turned out to be the second one. It's true that the variation in terms of the heat exchange coefficient is not important but they are two reasons for that: for one, the velocity is not very important (≅ 2.5 m/s) and therefore the variations are less obvious, we usually get more important variations when the inlet

velocities are higher (≅ 5 or 6 m/s) and for two, we saw during the validation of the procedure that the code underestimate the variation of h. Same remark applies to the friction coefficient Cf but we still get a variation of about 10 %. The conclusion was that a long central flat length was much better in terms of pressure drop and slightly better in terms of heat exchange. The visualization of the velocity field around the central flat length helps us understand why the central flat length brings better performances. (*fig.* 27, 28 and 29)

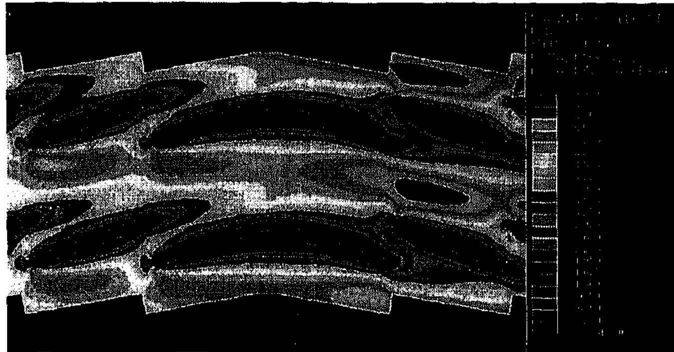


Fig. 27 - velocity field around a central louver without a flat length

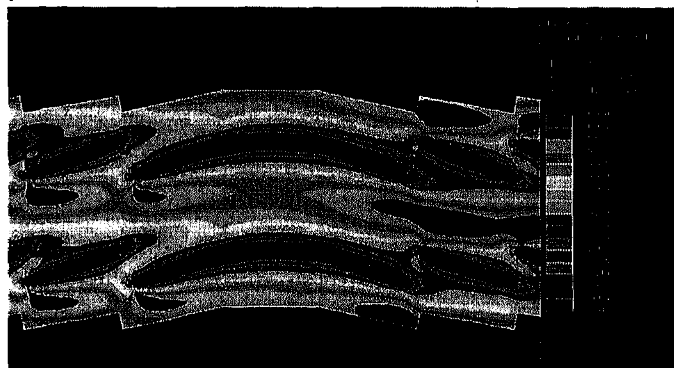


FIG. 28 - velocity field around a central louver with a long flat length.

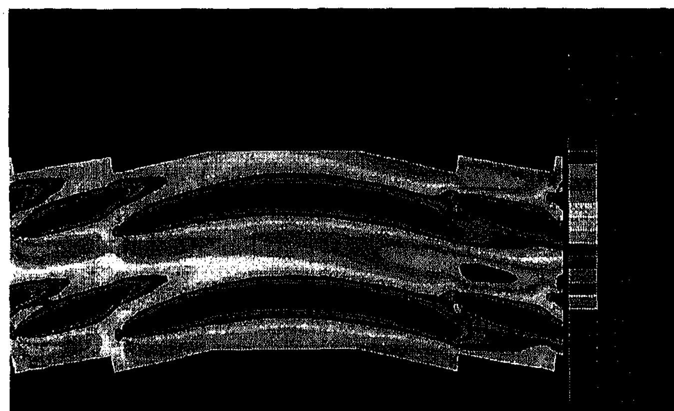


FIG. 29 - velocity field around a central louver with a short flat length

CONCLUSION

This method allows the optimization of the louvers geometry for a wide range of applications. An "ideal" optimized geometry can always be defined in a reasonable time. But more important is the possibility to quantify the effects of the degraded modes on the performances of an exchanger. We know that the aging of the tools is an unavoidable process. Taking into account these "imperfections" in the performance predictions is very important because it will define in a rigorous manner the dispersion that can be tolerated and the parameters of the geometry which variations have an important (or negligible) effect on the performances. It will enable us to tell the risky geometry (which cannot bear production dispersions) from the "safe" geometry (which can stand production variations). This way, we will be able to anticipate the behavior of the exchangers as the industrial tools get old and therefore to maintain a level of quality through a better control of dispersion.

NOMENCLATURE

Re: Reynolds number

D_H: hydraulic diameter

Y_{int}: internal thermal resistance

Y_{ext}: external thermal resistance

Cf: friction coefficient

h: heat exchange coefficient

T_∞: upstream temperature

T_w: wall temperature

ν : kinematic viscosity

ρ : density

970843

Development of Air Fuel Ratio Sensor for 1997 Model Year LEV Vehicle

Kazuya Mizusawa, Kazunori Katoh, and Shigeki Hamaguchi
Toyota Motor Corp.

Hidetaka Hayashi
Denso Corp.

Shinji Hocho
Nippon Soken Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

An exhaust air fuel ratio sensor (A/F sensor), which is applied to a 1997 model year LEV vehicle was developed. This sensor enables the detection of the exhaust gas air fuel ratio, both lean and rich of stoichiometric. This A/F sensor was developed from a lean mixture sensor, which has a proportional output to the exhaust gas air fuel ratio in the lean region only, by widening the detection range to rich air fuel ratios to 12:1. This sensor is comprised of a zirconia solid electrolyte and a platinum electrode with a ceramic coating used as a diffusion layer. As a result of improvements, it has a effective air fuel ratio range from 12 to 18 as required for LEV vehicles with model based air fuel control systems. It has a fast light off, -- within 20 seconds -- to minimize exhaust hydrocarbon content. Further, it has fast response times, less than 200 msec., to improve air fuel ratio controllability.

INTRODUCTION

Social demands for environmental improvement is increasing. For the automobile, exhaust gas criteria pollutants are required to be reduced under various regulations. Limits continue to decrease because air quality standards have not yet been met. The state of California has already introduced an LEV program and an EU-III program is under development in Europe.

Various exhaust gas control systems have been introduced to meet these emission regulations, and three-way-catalytic converter / oxygen sensor systems have been widely adopted. In the beginning, these systems had one catalytic converter and one oxygen sensor upstream of the catalytic converter. Toyota introduced the two-oxygen sensor system in 1987. This system allows more precise control of air fuel ratio. It consists of oxygen sensors both upstream and downstream of the catalytic converter. This system has

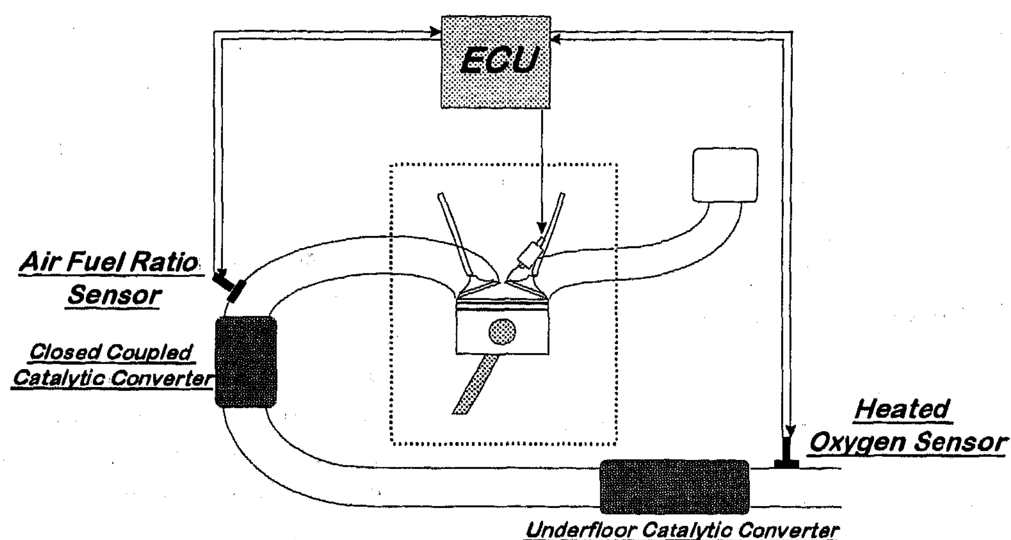


Fig. 1 : Exhaust System for 1997 Model Year 2.2 liter inline 4 engine for LEV vehicle

the following exhaust emission, durability, and reliability advantages:

(1) feed-back for variation of fuel composition
(2) compensation for deviation of the upstream sensor from the stoichiometric point
(3) tighter control within the window air fuel ratio range of the catalyst

The aim of the two oxygen sensor system is to maintain the air fuel ratio near the stoichiometric point, by detecting whether the oxygen content in the exhaust gas is only rich or lean. Since even deteriorated catalytic converters have reasonable efficiency in a narrow air fuel ratio window, system performance is improved.

Beginning in 1994, the California OBD-II regulation required detection of catalytic converter deterioration. To meet this requirement, most automobile manufacturers have designed their exhaust emission control system with an oxygen sensor downstream of the catalyst. By comparing the upstream and downstream oxygen sensor signals, catalytic converter efficiency can be inferred.

As emission regulations continue to tighten, it may seem easy to use a greater amounts of the noble metal (typically: platinum and/or palladium and rhodium) in catalytic converter to improve the system performance. However, availability, cost and deterioration require improved air fuel ratio control to better utilize the catalytic converter. In this context, new exhaust emission control systems have been under development.

SYSTEM CONFIGURATION

At steady running conditions, the air fuel ratio can be precisely controlled around the stoichiometric point using a conventional oxygen sensor. However, transient conditions such as acceleration or deceleration require more accurate air fuel ratio control to meet LEV level exhaust emission standards.

For the 1997 Model Year, Toyota has introduced vehicles which comply with the California LEV classification. In comparison with the 1996 Model Year TLEV vehicle, this system (Fig.1) consists of the items listed below:

(1) An exhaust A/F sensor is used instead of the oxygen sensor upstream of the catalyst at the exhaust manifold
(2) A close coupled catalytic converter with a low thermal mass metal foil substrate is used for the reduction of hydrocarbons during cold starts
(3) The under floor catalyst is enlarged to reduce emissions during hot engine conditions, and an oxygen sensor is set downstream

This system is designed to control air fuel ratio precisely around the stoichiometric point under steady conditions, and to quickly stabilize the air fuel ratio around the stoichiometric point during transient conditions.

Ohata et.al [1] and Nasu et.al [2,3] have developed a model based air fuel ratio feedback control system with counter control of fuel injected to the engine. This counter control requires quantitative measurement

of the exhaust gas air fuel ratio which an oxygen sensor cannot measure. An output proportional to the exhaust gas air fuel ratio must be used. Using this control, the integrated value of exhaust gas air fuel ratio into the catalytic converter is controlled to be stoichiometric. This results in maximization of the gas conversion in the catalytic converter.

(CONVENTIONAL) OXYGEN SENSOR

The oxygen sensor is widely used to control the air fuel ratio within a narrow range or window to optimize the efficiency of the three-way catalytic converter for both oxidation and reduction of exhaust pollutants. It consists of a zirconia solid electrolyte with platinum electrodes on both surfaces. One side is exposed to exhaust gas and the other to the atmosphere. This sensor is characterized by an output signal discriminating only rich or lean operation. This is a simple method which leaves much room for improvement and for optimizing air fuel ratio control.

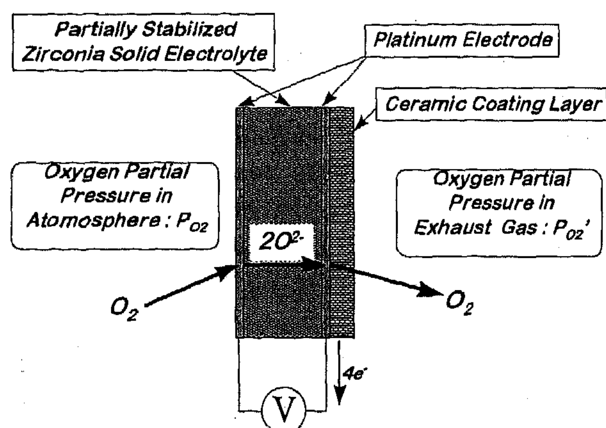


Fig. 2 : Detection Principle of (Conventional) Oxygen Sensor

Figure 2 shows the principle of a conventional oxygen sensor. The oxygen sensor is considered to be an "oxygen concentration cell." Depending on the difference in the oxygen partial pressures between the two surfaces of the platinum electrode, an electromotive force (E.M.F.) is produced according to the Nernst's Equation:

$$E = \frac{RT}{4F} ln\left(\frac{P_{O2}}{P_{O2}'}\right) \quad (1)$$

where

R: gas constant

T: absolute temperature

F: Faraday's constant

Po_2: oxygen partial pressure of the atmosphere

Po_2': oxygen partial pressure of the exhaust gas.

When the air fuel ratio is lean, E is about 0V, and when air fuel ratio is rich, E is about 1V. Near the stoichiometric point, small amounts of NOx, HC, CO and O_2 are present in the exhaust gas. Because of the catalytic ability of the platinum electrode of the sensor, these species are converted to stable compounds. This is why the difference of the stoichiometric point between the catalytic converter and the oxygen sensor is almost negligible.

Figure 3 is the output characteristic of this sensor. The output voltage changes abruptly at the stoichiometric point, so that the oxygen sensor can indicate only a rich or lean state. Using this sensor, fuel injection control is performed by reducing the fuel injection when the sensor output is above 0.45V and increasing the fuel volume when the output is below 0.45V.

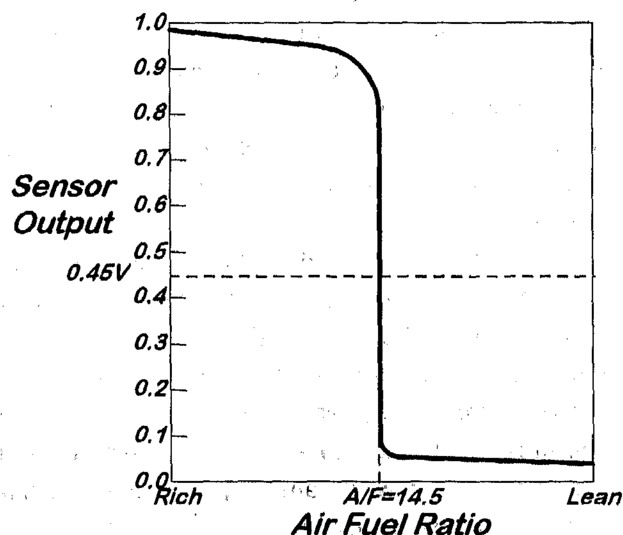


Fig. 3 : Output Characteristics of (Conventional) Oxygen Sensor

LEAN MIXTURE SENSOR

Lean mixture sensor was developed for lean combustion systems. It was the first production exhaust gas sensor to have a proportional output characteristic. This system allows the maintenance of an air fuel ratio of approximately 23, the optimal value for NOx emission reduction and fuel consumption. The electrochemical

reaction of the lean mixture sensor is shown in Fig.4 In this sensor, the ceramic diffusion layer is used to create a current output proportional to the air fuel ratio. This is represented as follows:

$$I = C\frac{D_{O2}}{T}\frac{PS}{\ell}\ln\left(\frac{1}{1-P_o/P}\right) \tag{2}$$

where

C: constant

D_{O_2}: diffusion coefficient at the diffusion layer

T: absolute temperature

S: electrode surface area

ℓ: length of diffusion layer

P_0: oxygen partial pressure of inner (atmosphere) electrode

P: absolute pressure

The diffusion layer consists of $MgAl_2O_4$, called "SPINEL", which has micro pores from 100Å to 1000Å radius. This diffusion layer restricts molecular diffusion according to the oxygen partial pressure.

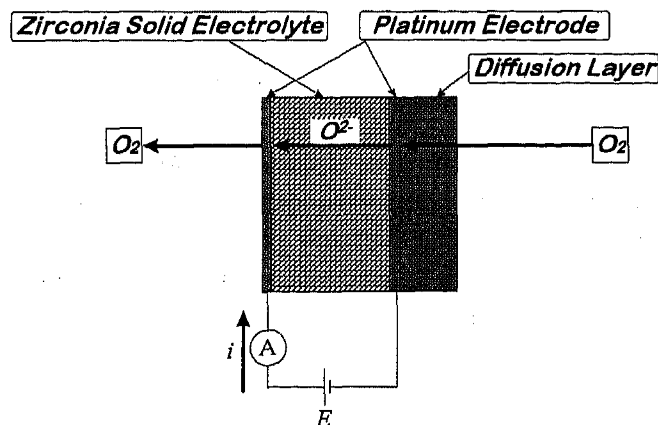


Fig. 4 : Detection Principle of Lean Mixture Sensor

According to Kamo et.al. [4], diffusion at the diffusion layer can be obtained with the character of "Temperature Independence" or "Pressure Independence" based on the radius of the micro pore. For the automotive application, the "Temperature Independent" type diffusion layer is used.

The output characteristic of the lean mixture sensor is shown in Fig.5. It varies based on the exhaust gas air fuel ratio and the applied voltage. This current curve shows a plateau region for an applied voltage of 0.2V to 0.9V. When a constant voltage (shown in figure by the dotted line) is applied to this sensor, e.g. 0.8V, the current flow is nearly proportional to the exhaust gas air fuel ratio. (Fig.6)

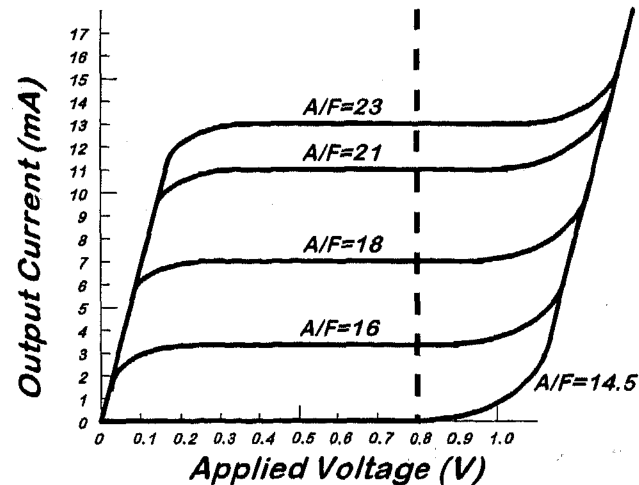


Fig. 5 : Output Characteristics of Lean Mixture Sensor

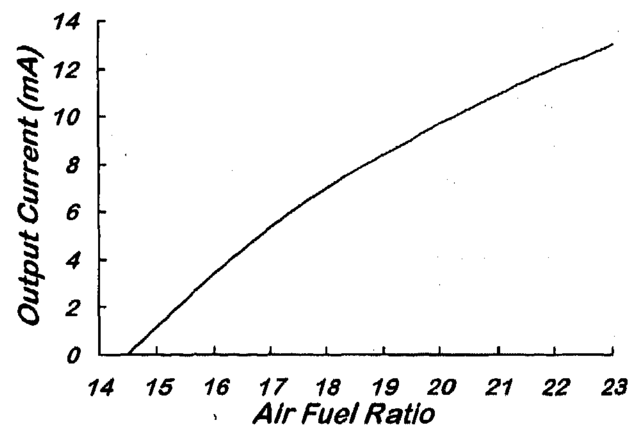


Fig. 6 : Output Current of Lean Mixture Sensor

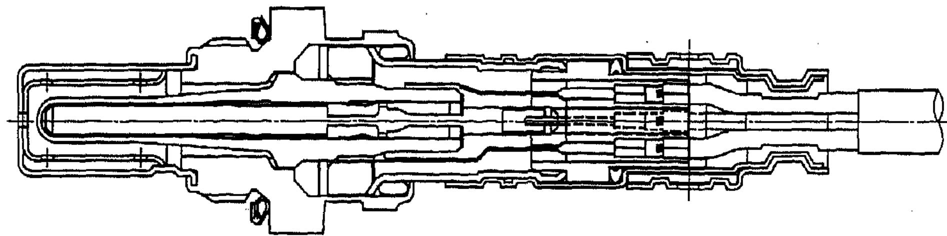


Fig. 7 : Schematic Illustration of Air Fuel Ratio Sensor

AIR FUEL RATIO SENSOR

Since the introduction of the oxygen sensor, most oxygen sensors and lean mixture sensors have had an element of cylindrical shape with one closed end. This configuration has high durability and reliability, because of its thermal strength against internal stress.

The A/F sensor's development was based on the lean mixture sensor. Its basic construction is shown in Fig.7. The required specification of the A/F sensor for LEV vehicles is as follows:

a) Functional air fuel ratio range from 12 to 18
b) Activation time from cold start of less than 20 seconds
c) Response times in the entire air fuel ratio range of less than 200msec

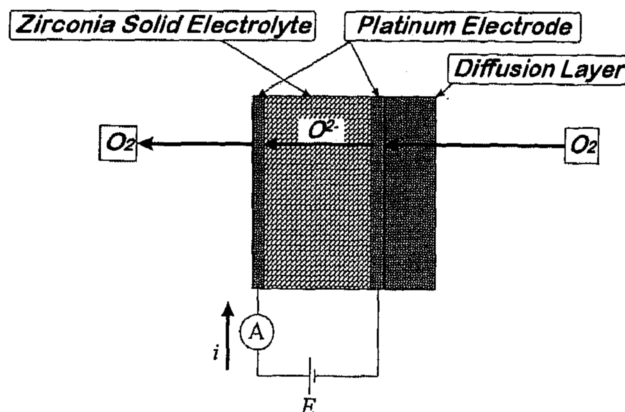


Fig. 8 : Detection Principle of Air Fuel Ratio Sensor in Lean Exhaust Gas

Since the electrochemical reaction at the sensing element is the same as that of the lean mixture sensor in the lean region, then, of course in the lean region, the output characteristic is also the same (Fig.8). This reaction is considered to be oxygen pumping from the exhaust gas (outer electrode) to the atmosphere (inner electrode). [5]

On the other hand, in the rich region, ion movement is in the opposite direction. The electrochemical reaction at the exhaust gas surface (outer electrode) is the combination of hydrocarbon and carbon monoxide with oxygen to form water and carbon-dioxide (Fig.9).

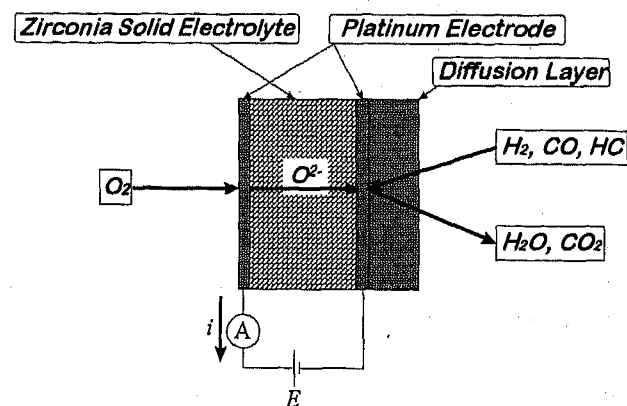


Fig. 9 : Detection Principle of Air Fuel Ratio Sensor in Rich Exhaust Gas

The diffusion layer is used to restrict molecule movement of both water and carbon-dioxide. From the mass and the partial pressure of the molecules, the velocity of molecule transportation in the diffusion layer is calculated from the following equation: [6]

$$\frac{|I_p(lim)|}{P_e} \propto \frac{n}{\sqrt{M}} \qquad (3)$$

where

- $I_p(lim)$: limit current obtained with A/F sensor
- P_e : partial pressure of each component in exhaust gas
- n : number of electrons for electrode reaction of each gas
- M : mass of molecule

To verify this relation with hydrocarbon, the case for propane is examined. For propane, n=20 and M=44, then we obtain $n/\sqrt{M}$ as 3.015. The observed

I_p(lim) and $n/\sqrt{M}$ is summarized in Table 1, and is shown in Fig.10.

	$n/\sqrt{M}$	I_p(lim) [mA]
C_3H_8	3.015	17.2
H_2	1.414	7.0
O_2	0.707	3.4
CO	0.378	1.8

Table 1 : Calculated $n/\sqrt{M}$ and observed output current of each gas

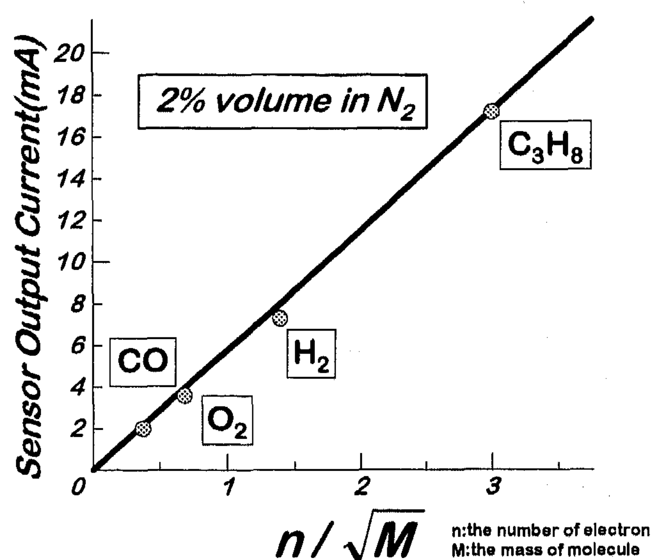


Fig. 10 : Air Fuel Ratio Sensor Output Current for Various Gas Molecule

DETECTION RANGE

The Lean mixture sensor is designed for detection of air fuel ratios around 23, based on the needs of the lean burn engine. Considering the electrochemical reaction at the lean region, oxygen molecules are pumped from the exhaust gas to the atmosphere. The pressure inside the sensor housing goes up, but there is no problem releasing oxygen to the atmosphere. Test results using ambient air (oxygen content about 21%) show no problem (Fig.11).

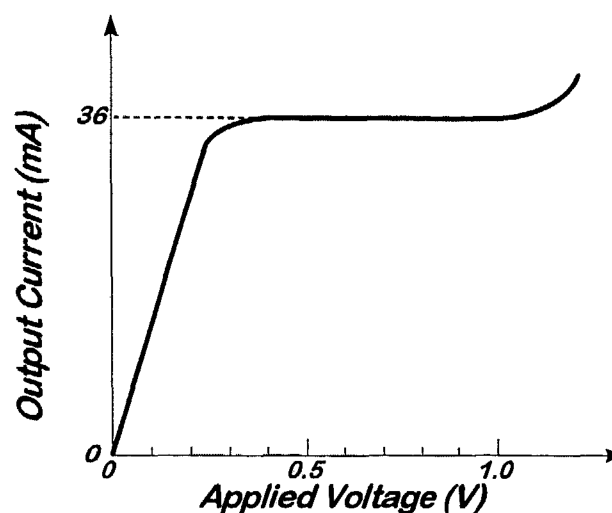


Fig. 11 : Output Characteristics of Lean Mixture Sensor in Atmosphere

In the rich region, however, the output current is saturated at an air fuel ratio of 13 (Fig.12 and 13). As described by Chujo et.al [7], oxygen flow determines the rich detection limit. During rich operation of the A/F sensor, oxygen is provided from the atmosphere to the exhaust gas through the path between the sensing element and the heating element. Assuming that the oxygen path between the sensing element and the heating element is a simple diffusion process and calculating the transport of oxygen molecules, there is insufficient path for detecting an air fuel ratio of 12. Additional detail is described in Appendix I. In order to widen the path between the sensing element and heating element, the inner radius of the sensing element was widened (Fig.14).

Experimental results with a 2.2 liter inline-4 engine, shown in Fig.15 and 16, has proven this improvement to be very effective in expanding the detection range to air fuel ratios of less than 12.

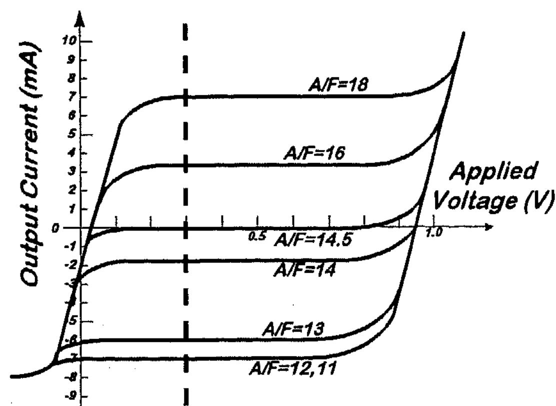


Fig. 12 : Output Characteristics of Lean Mixture Sensor in both Lean and Rich Region

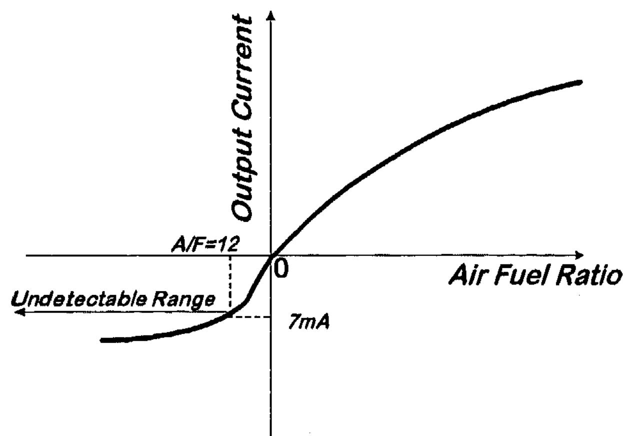


Fig. 13 : Undetectable Air Fuel Ratio Range with Lean Mixture Sensor

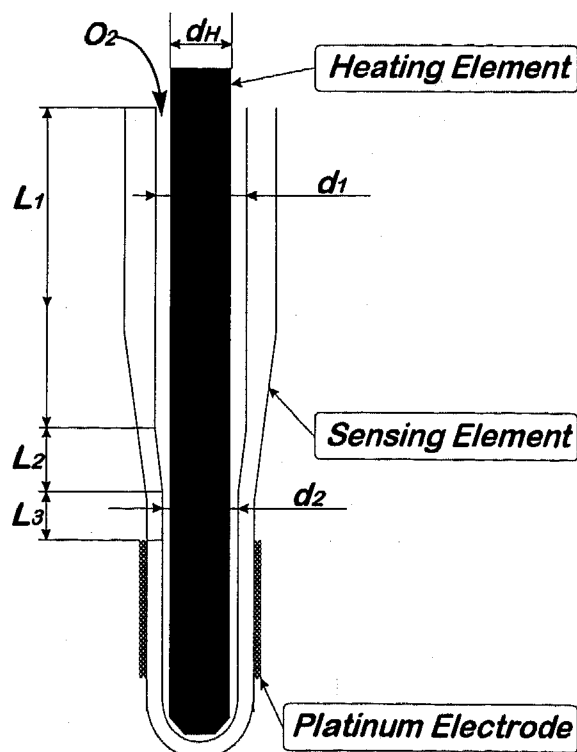


Fig. 14 : Oxygen Path between Sensing Element and Heating Element

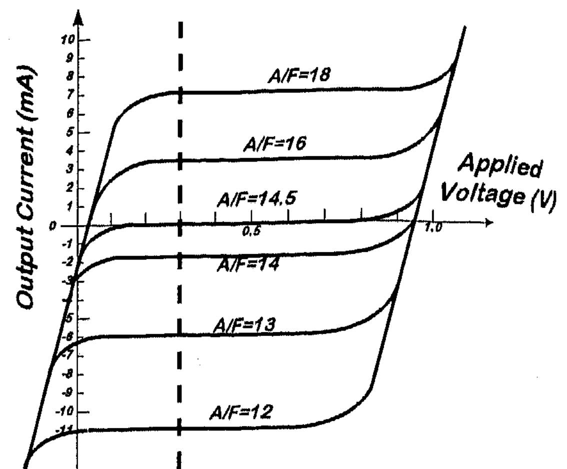


Fig. 15 : Output Characteristics of Air Fuel Ratio Sensor

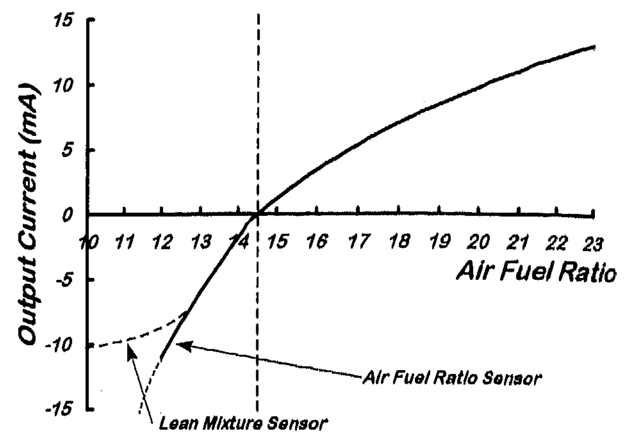


Fig. 16 : Improved Detection Range of Air Fuel Ratio Sensor

FAST LIGHT OFF

With regard to California's LEV Program, the hydrocarbon requirement is the most difficult to meet. HC emissions at cold start are nearly 70% of the total FTP cycle. Fast light off is required for both the catalytic converter and the A/F sensor.

Activation temperature for the A/F sensor is about 650 °C, instead of 400 °C for the conventional oxygen sensor. The optimization of fast light off without loss of durability is critical.

To improve the light off time, numerous methods to improve thermal conductivity from the heating element to the sensing element, improve activation temperature of the sensing element, and improve heating element capability have been considered. Because this A/F sensor's development was based on the lean mixture sensor, we first chose to improve the heating element.

Material

A heating element laminated on alumina was adopted for the lean mixture sensor. This heating element was designed only to maintain the sensing element in its active temperature region and is insufficient to meet the requirement of activation from a cold start in 20 seconds. The alumina heating element, requires more than 1200 °C to activate the A/F sensor within 20 seconds. This will cause problems such as cracking or oxidation of the heating material. Figure 17 shows the temperature rise of the A/F sensor with the alumina heating element. The A/F sensor is activated at a temperature greater than 700 °C, and the temperature above which the durability of heating element deteriorates is about 1200 °C. When sufficient power to activate the A/F sensor within 20 seconds is introduced to the heating element, its temperature exceeds 1200 °C. Therefore, it is not feasible to have both fast activation and durability using the alumina heating element.

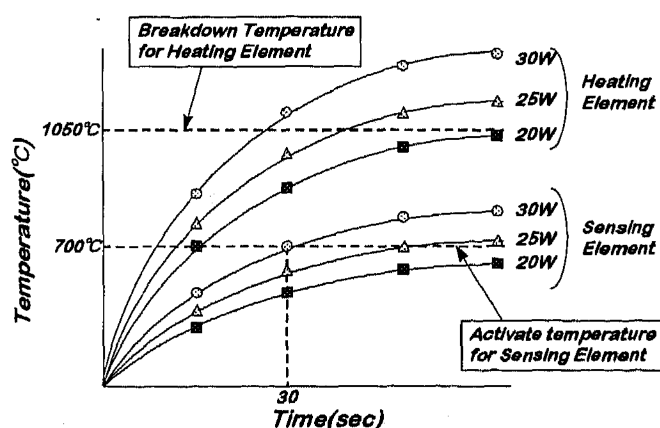


Fig. 17 : Heating Element and Sensing Element Temperature Rise after Cold Start

A silicon nitride substrate can be made using a hot press method to create a rigid structure. This material is widely used in oil stoves to quickly evaporate fuel after being switched on. It is also used in the turbine of turbo chargers which are exposed to high temperature.

Figure 18 shows the temperature rise of the silicon nitride heating element. Its breakdown temperature is greater than 1300 °C. This data shows that the A/F sensor can be activated using the silicon nitride heating element within 25 seconds. Farther, changing the input power control (100% duty cycle at cold start, 30W continuous after activation) for the heating element results in a faster light off. The final result is that the A/F sensor is activated within 20 seconds.

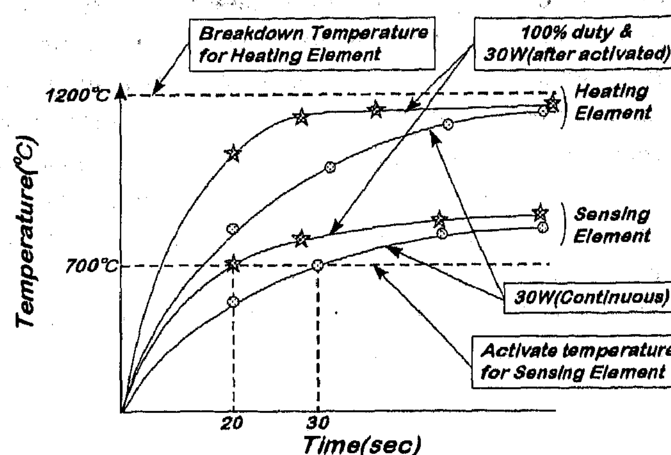


Fig. 18 : Cold Start Temperature Rise with Silicon Nitride Heating Element

Figure 19 shows the durability of the silicon nitride heating element at high temperatures. This graph shows that the silicon nitride heating element's durability can still be maintained to about 15,000 hours at 1300 °C, while the alumina heating element has reached the deterioration limit after only 9,000 hours at 1050 °C.

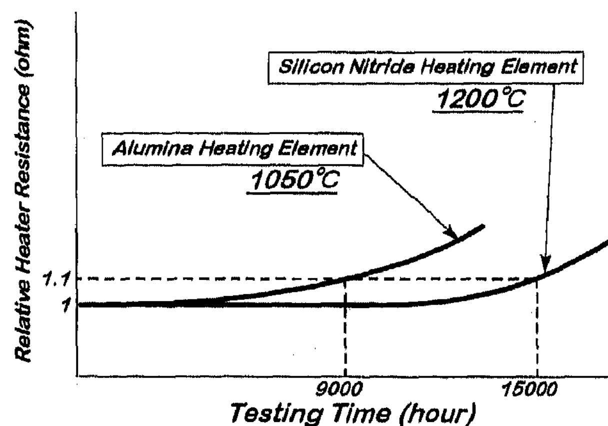


Fig. 19 : Durability of Silicon Nitride Heating Element with Continuous High Temperature Test

Geometrical Arrangement

The heating element is also required to have a geometrically constant heating ability to maintain the sensing electrode at a constant temperature. Also, for fast light off, the heater resistance should be around 1 ohm. Unfortunately, the heater material is a tungsten carbon composite, which has a relatively high electronic resistance. Another design requirement that must be considered is a uniform distribution of temperature around the heating element. To satisfy all the requirements, the geometric arrangement of the heater is designed as shown in Fig.20. Two heaters are placed in parallel to obtain a lower resistance.

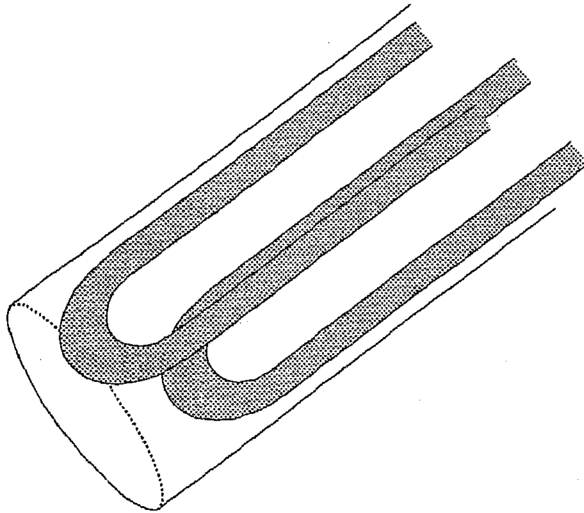


Fig. 20 : Heater Coil Arrangement for Silicon Nitride Heating Element

RESPONSE TIME

According to Nasu, et. al. [2], A/F sensor response time is critical to the model based control method. Response of the A/F sensor is defined as a 63% change in output current (Fig.21).

The response to various step changes in air fuel ratio was measured. Results are shown in Fig.22. The slowest responses were observed when there was a small air fuel ratio change across stoichiometric, either lean-to rich or rich-to-lean. We chose to concentrate on the response time of the sensor from air fuel ratios of 14 to 15 and 15 to 14.

Figure 23 shows the response of a prototype A/F sensor that has the same protection cover as the lean mixture sensor. As shown in Fig.24, this cover has 2 rows and 4 columns of gas apertures. The gas apertures are designed to avoid poisoning from oil contents. This sensor's response is not sufficient for the LEV application because it causes hunting in the air fuel ratio control.

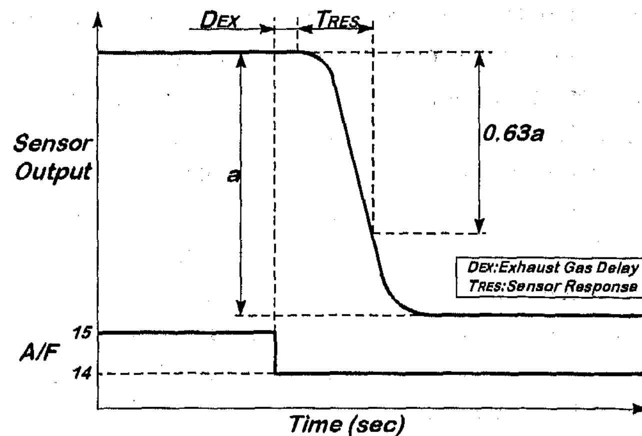


Fig. 21 : Response Definition of Air Fuel Ratio Sensor

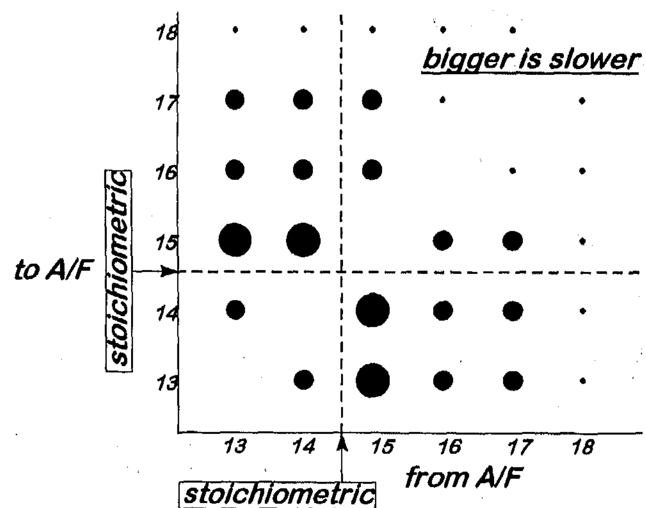


Fig. 22 : Response of Air Fuel Ratio Sensor at various Air Fuel ratio Stepwise Change

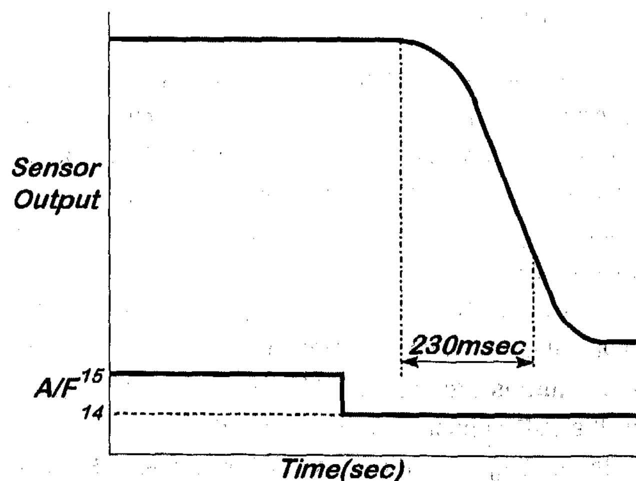


Fig. 23 : Response of Prototype Air Fuel Ratio Sensor

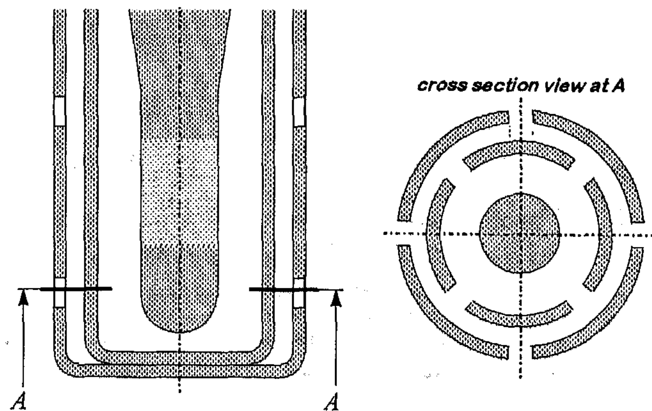


Fig. 24 : Aperture Arrangement of Protection Cover for Lean Mixture Sensor

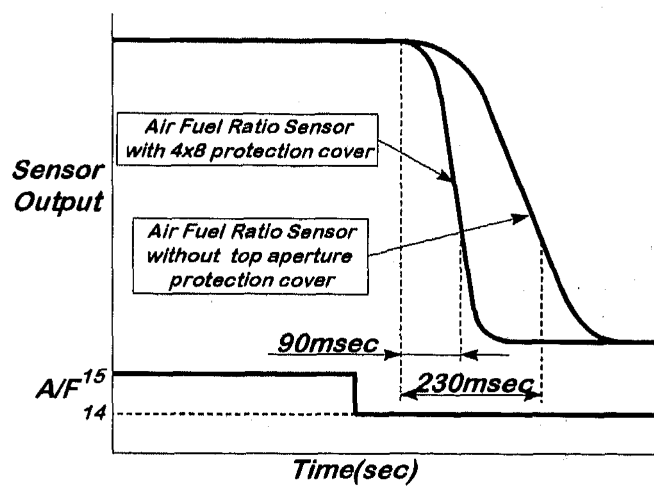


Fig. 25 : Aperture Arrangement of Protection Cover for Conventional Oxygen Sensor

Figure 25 shows the A/F sensor response using a conventional oxygen sensor protection cover, with 4 rows and 8 columns of gas apertures. This aperture arrangement is shown in Fig.26. The response using this cover is nearly the same as that of the sensing element itself, without a cover. This also shows that if gas flow is insufficient, there will be a delay in the response of the sensor.

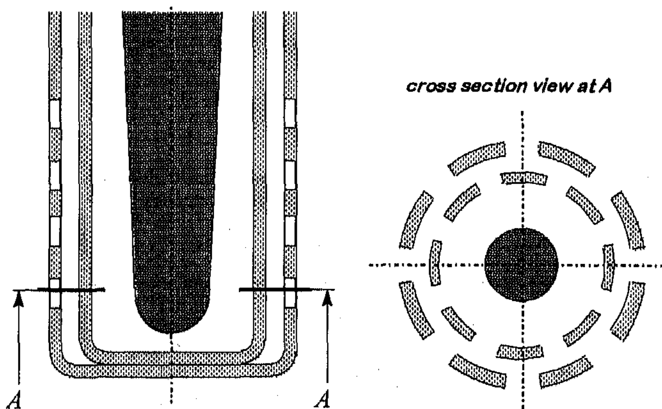


Fig. 26 : Aperture Arrangement of the Protection Cover for Conventional Oxygen Sensor

Gas flow analysis was carried out using an oil film method and a smoke tracing method [see Appendix II]. Figure 27 shows the result of these analyses. The main gas flow around the sensing element is directly across the sensing element resulting in insufficient gas exchange. To improve the gas exchange around the sensing electrode, an additional gas aperture was added at the top of the protection cover (Fig.28).

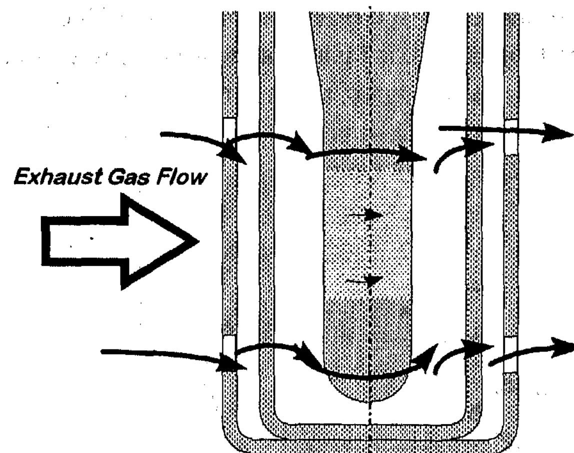


Fig. 27 : Gas Flow with Protection Cover (without Top Aperture)

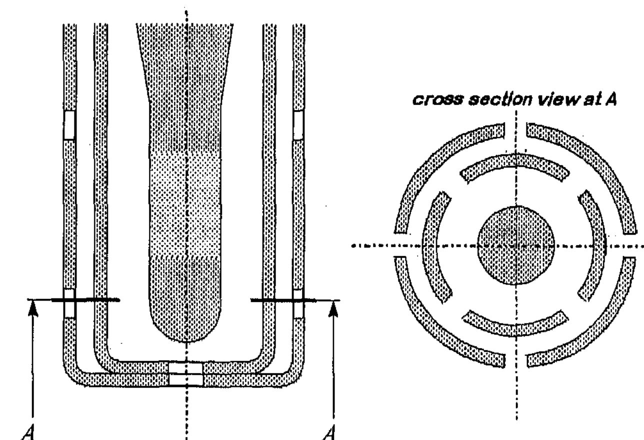


Fig. 28 : Aperture Arrangement of the Protection Cover with Top Aperture

Figure 29 shows the gas flow at the sensing element using the protection cover with the top aperture. The gas flow along the sensing element was obtained by adding the gas flow through the top aperture. The response of the sensor with the protection cover with the top aperture is shown in Fig.30 and 31. The worst case response time value is about 150 msec, which is sufficient for the required specification. These results are summarized in Table 2. The response of the cover with the top aperture is faster than that with 2 rows and 6 columns of gas apertures, and is almost the same as that with 4 rows and 8 columns.

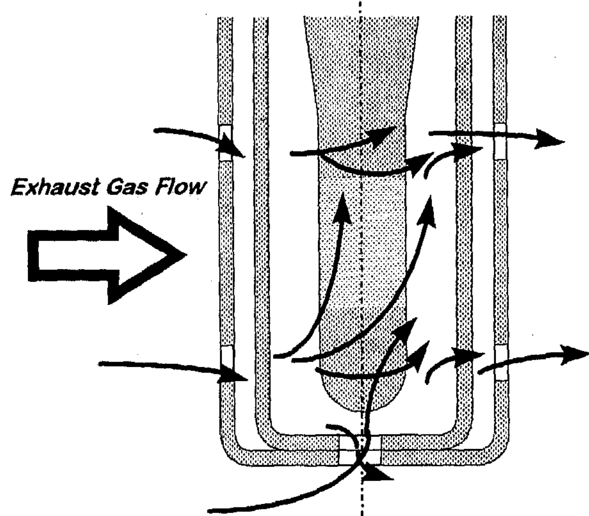


Fig. 29 : Gas Flow with Protection Cover (with Top Aperture)

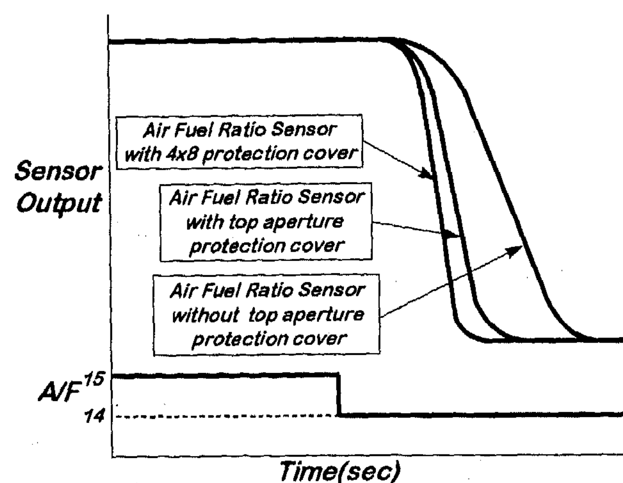


Fig. 30: Lean→Rich Response of Air Fuel Ratio Sensor

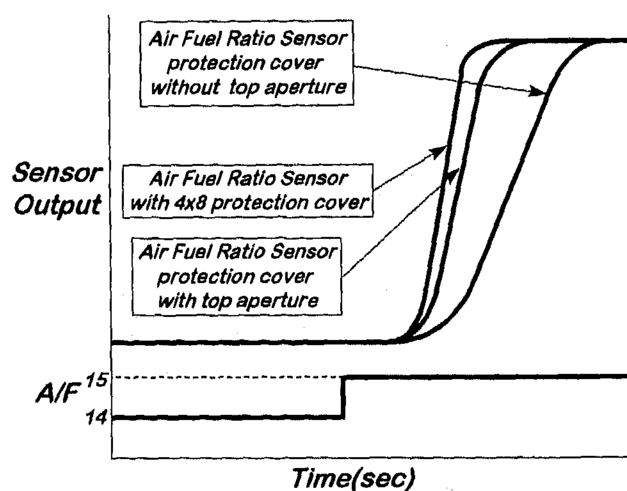


Fig. 31 : Rich→Lean Response of Air Fuel Ratio Sensor

Aperture Arrangement				Response (worst data) milliseconds	
Radius	Rows	Columns	Top Aperture	A/F: 14→15	A/F: 15→14
2 mm	4	8	No	130	140
1.5 mm	2	4	No	210	230
1.5 mm	2	6	No	150	180
1.5 mm	2	4	2 mm	140	150

Table 2 : Response of Air Fuel Ratio Sensors with Various Protection Covers

FUTURE DEVELOPMENT

Regulations requiring further reductions of exhaust emissions will continue into the future. The main obstacle in meeting these regulations is hydrocarbon emissions, which, as described in the introduction, hydrocarbon is emitted mainly during cold starting. In the future, much faster light off will be required of both the sensor and catalytic converter.

To improve fast light off, a new sensor that has a planar construction is being widely investigated because of its low thermal capacity. The cylindrical shaped sensing element is advantageous due to its low internal thermal stress. The challenge we face will be in developing a faster light off sensor while maintaining high durability and reliability.

Another theme for sensor development is high temperature durability to improve tolerance to little or no fuel enrichment at high load conditions.

We are trying to improve the exhaust gas sensor to meet these challenges. However, there are still many countries that allow leaded gasoline which deteriorates both exhaust gas sensors and catalytic converters. If global air pollution is to be reduced, all countries will need to adopt programs where leaded gasoline is outlawed and emission are strictly regulated.

CONCLUSION

The A/F sensor was developed with the following improvements: (1) expansion of the rich detection region using analysis of the gas flow between the sensing element and heating element; (2) improved light off using a silicon nitride heating element; (3) improved response time by using gas flow analysis to optimize the sensor protection cover aperture arrangement. This A/F sensor, to be used for air fuel ratio control in LEV and ULEV vehicles, represents a break through in environmental improvement.

ACKNOWLEDGMENTS

The authors would like to thank all the people who contributed to this development. Without their effort, this work could not have been completed.

REFERENCES

[1]A.Ohata, M.Ohashi, M.Nasu, T.Inoue, "Model Based Air Fuel Ratio Control for Reducing Exhaust Gas Emissions", SAE paper 950075, 1995

[2]M.Nasu, A.Ohata, S.Abe, "Model-Based Fuel Injection Control System for SI Engines", SAE paper 961188, 1996

[3]Y.Iwata, S.Abe, M.Nasu, K.Aoki, "Model Based Air Fuel Ratio Control with Exhaust Air Fuel Ratio Sensor & Heated Exhaust Oxygen Sensor", FISITA International Congress, P1650, 1996

[4]T.Kamo, Y.Chujo, T.Akatsuka, J.Nakano, M.Suzuki, "Lean Mixture Sensor", SAE paper 850380, 1985

[5]D.Yuan and F.A. Kröger, "Stabilized Zirconia as an Oxygen Pump", J.Electrochem. Soc., 594, May, 1969

[6]H.Tanaka, S.Nishikawa, S.Suzuki, M.Miki, T.Harada, M.Kanamaru, S.Ueno, N.Ichikawa, "Wide-Range Air-Fuel Ratio Sensor", SAE paper 890299, 1989

[7]Y.Chujo, K.Aoki, T.Kamo, M.Ogino, M.Suzuki, T.Saito, "Development of Onbourd Fast-Response Air-Fuel Ratio Meter Using Lean Mixture Sensor", ISATA Vol.I 89038, P459-475, 1989

APPENDIX I : CALCULATING OXYGEN PATH WITH SIMPLE DIFFUSION PROCESS

According to the diffusion equation, the output current of the A/F sensor is expressed as follows,

$$I\ell = C\frac{S}{\ell}D_n \ln\frac{1}{(1-P_{O_2})} \qquad (4)$$

where

$I\ell$: output current

C : constant (=4FP/RT)

S : area of diffusion path

ℓ: length of diffusion path

Dn : diffusion coefficient for oxygen molecule in the atmosphere (0.178)

PO2 : partial pressure of oxygen in atmosphere.

and, the s/ℓ is integrated with

$$\frac{S}{\ell} = \frac{1}{\frac{\ell_1}{s_1}+\frac{\ell_2}{s_2}+\frac{\ell_3}{s_3}+\frac{\ell_4}{s_4}+\frac{\ell_5}{s_5}\ldots\ldots} \qquad (5)$$

Calculating with the geometric value, we can obtain the output current limit value for both unimproved and improved element. The results are summarized in Table 3.

	Output Limit	A/F Limit
Unimproved	7mA	12
Improved	12mA	11

Table 3 : Rich Detection Improvement Result

APPENDIX II : VISUALIZATION OF GAS FLOW

Many methods have been reported for visualization of gas flow. We selected 2 methods which were, "Oil Film Method" and "Smoke Tracing Method".

With the oil film method we detected the gas flow around the surface of the outer cover and between the outer and the inner cover, and by the smoke tracer method, the gas flow around the sensor element itself.

OIL FILM METHOD

The oil film method is carried out as follows: First coat the surface of test objects with a thin layer of oil emulsified with TiO2 powder. Second take it into the gas flow. When the oil is blown off and drained away, the powder is stretched into lines along the gas flow line. Then, the gas flow can be seen as the lines on the objects.

SMOKE TRACING METHOD

The smoke tracer method is carried out as follows: All of the experimental apparatus is made using a clear material such as acrylonitrile plastic for components such as the cover and element. They are assembled and inserted into the gas flow. Then, smoke is added to the gas flow. The gas flow can be seen as the lines of smoke.

970845

Ignition System-Embedded Fiber-Optic Combustion Pressure Sensor for Engine Control and Monitoring

Thomas J. Poorman, Liangdao Xia, and Marek T. Wlodarczyk
Optrand Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

We report a low-cost fiber-optic combustion pressure sensor integrated with an ignition system of a passenger car. A miniature sensor is attached to a shaft extending from the spark plug boot. Simultaneously with the boot mounting on the spark plug the sensor shaft is pressed into a cylindrical hole of a modified production spark plug. The sensor cable is combined either with the boot, or boot and high voltage cable, depending whether a coil-at-plug or a Distributed Ignition System is used, respectively. The sensor optical and Application Specific Integrated Circuit (ASIC) components are packaged inside a miniature module, embedded either in a coil-at-plug package or a Distribute Ignition System module. Engine test results are reported for a single-cylinder, gasoline fueled stationary gen-set engine. When optimized for high linearity and reduced thermal shock, the spark plug-mounted sensor demonstrates approximately +/-1.5% pressure reading accuracy, comparable to that of a laboratory-grade piezoelectric transducer. With the life-time of 200 Million pressure cycles and targeted price of $7 to $10, the sensor meets the performance, durability, and cost requirements for production car applications.

1. INTRODUCTION

Cylinder pressure is the fundamental variable that determines a combustion engine's operating state. If available for all cylinders, combustion pressure information can be used in advanced engine applications ranging from diagnostics and testing through monitoring to closed-loop control systems. Based on cylinder-specific pressure information, engine control applications have been proposed for power balancing in natural gas burning large-bore engines [1], lean burn combustion in passenger cars [2], or stall control in aircraft engines. In the area of engine monitoring, combustion pressure, when detected in all cylinders, can provide the most deterministic information about engine performance and health [3]. In conjunction with other sensors cylinder pressure-based engine health monitoring systems can lead to increased engine reliability as well as reduced cost and frequency of preventive maintenance.

While the benefits of cylinder pressure information for engine monitoring or control have been known for decades, commercial implementations of practical systems have been hampered by one overriding factor: the lack of a cost-effective and reliable combustion pressure sensor. Conventional electronic pressure transducers, such as piezoelectric or piezoresistive, are not suited for continuous high temperature pressure measurements. Fundamentally, capacitance or strain gages used in pressure transducers exhibit large non-repeatable and unpredictable changes in gauge output at temperatures typically greater than 250°C or 125°C, respectively. These changes are caused by such effects as alloy segregation, phase changes, selective oxidation, and diffusion, and ultimately lead to premature failure of the gauge or lead wires.

In contrast to electronic devices, fiber-optic sensors are potentially very well suited for applications characterized by high temperatures and high levels of electromagnetic interference[4]. These benefits, combined with very small package sizes, make fiber optical sensors well suited for operation in automotive engine environment. In particular, these sensors can be combined with existing engine components, such as ignition spark plugs or fuel injectors; the ability to insert a pressure sensor into a combustion cylinder through an existing engine head opening offers significant cost advantage over techniques dependent on drilling a hole into an engine's cylinder head or other forms of engine modifications.

While special pressure measurement spark plugs have been used in engine research, they are not suitable for use in production engines. Measurement spark plugs typically have an off-axis, small diameter central electrode and compromise either spark plug performance or durability. In addition, piezoelectric sensors used with these spark plugs are very expensive and have a limited life time. A measurement spark plug suitable for engine control applications should preferably be derived from a production device for low cost and should have the same performance and durability characteristics of a conventional spark plug. In addition, for production engine applications, a spark plug-mounted pressure sensor should require no separate signal conditioner and a minimum number of cables and connectors. Finally, the sensor-spark plug module should be installed at minimum cost and allow low-cost maintenance.

2. SYSTEM DESCRIPTIONS

A measurement ignition system with embedded fiber optic combustion pressure sensor is schematically shown in **Fig. 1**. The system, termed the PSIcoil™, consists of four key components: (1) a miniature fiber optic pressure sensor extending from a spark plug boot, (2) a spark plug possessing a channel into which the sensor is inserted, (3) a high voltage boot/cable within which a fiber optic cable is embedded, and (4) an opto-electronic module integrated with an ignition coil, mounted either on or at spark plug or at the end of a high voltage cable.

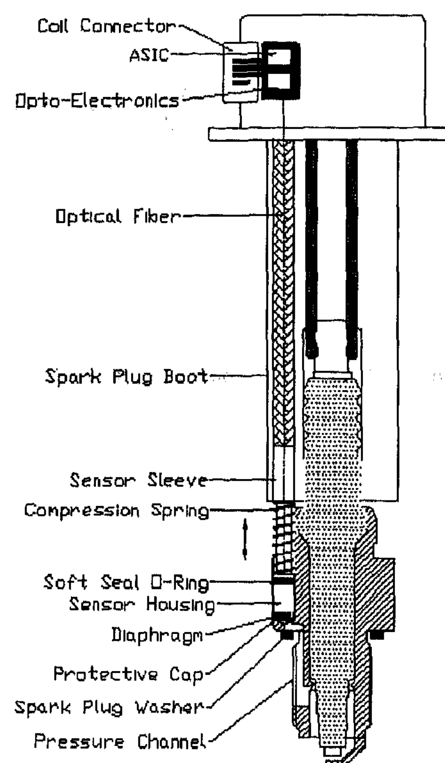


Fig. 1 PSIcoil™ schematic diagram

The fiber optic sensor is mounted at the end of a shaft extending from the spark plug boot. The sensor's diaphragm is protected by a metal cup which also provides an additional benefit of thermal shielding. As the spark plug connector latches onto the central electrode, the sensor is pushed into the measurement cavity and seals on the front of the protective cup. A spring, located between the spark plug boot and the shaft maintains a pressure tight seal. For added protection against pressure leaks an O-ring is mounted at the top of the shaft, away from high temperature area. The shaft section that is inserted into the spark plug is made of brass, or similar material with high thermal conductivity and good pressure sealing characteristics. Above the spark plug, the shaft is made of durable plastic so that the spark plug's electric field is not perturbed by the presence of any metal parts of the sensor.

The measurement spark plug is derived from a production-type device, as shown in **Fig. 1**, with an axially located central electrode. The spark plug key feature is a short, non-resonating passage connecting the inside spark plug area exposed to combustion pressure with the sensor located in the spark plug metal shell. The pressure passage consists of three sections: (1) a cylindrical hole parallel to spark plug axis extending to the bottom of the spark plug metal body, (2) a channel milled in the outside thread of the spark plug connected to the inside area by a transverse hole, and (3) a 1 mm diameter hole connecting the milled

channel to the sensor diaphragm. Note that by using the channel located in the spark plug thread the need to move the central electrode off axis, to create a pressure passage, is eliminated. The dimensions of the three sections of the pressure passage are optimized to minimize any distortions in dynamic pressure readings: (i) The total passage length between the spark plug top surface and the sensing diaphragm is kept to a minimum for minimum phase delays. (ii) The channel machined in the outside spark plug thread has a shape of letter V to eliminate any channel resonance. (iii) A hole connecting the outside and inside channels is optimized so combustion flame is quenched as it passes to the sensor chamber.

A number of options are available for this combustion pressure measurement ignition system. In one version, as schematically shown in **Fig. 1**, the sensor opto-electronics and Application Specific Integrated Circuit (ASIC) are mounted inside an ignition coil located on, or near, the spark plug. Through a channel formed in the measurement spark plug boot the sensor fiber optic strand enters the ignition coil package. There the strand is terminated with an opto-electronic module which in turn connects to the sensor's ASIC. The sensor is powered by the coil low-voltage supply cable and the combustion pressure output is available through an additional pin of the coil-on-plug electrical connector.

Alternatively, the fiber optical strand of the sensor can be broken into two sections, connected to each other via an optical connector. For ease of installation and low cost, the connector is combined with the coil electrical connector. Between the coil and Distributed Ignition System (DIS) the fiber optic strand is embedded right inside the high voltage cable, together with a carbon strand. Such an integration is only possible with a fiber optic cable, due to its non-electrical nature, so that neither the ignition field is perturbed nor the sensor cable is affected by the ignition system Electromagnetic Interference (EMI). Such a measurement ignition system is the most desirable for applications where an ignition coil may be exposed to temperatures over 125°C, such as for example in push rod engines.

Unlike in engine testing and R&D, a combustion pressure sensor intended for engine control applications does not need to be extremely accurate. The critical features are high reliability, low cost, and long life. Typically, no individual sensor calibration is needed in the interest to lower the total system cost. However, a high degree of linearity and low-hysteresis are required within a single combustion cycle so a normalization technique can be applied based on the ratio of cylinder pressure during compression and expansion stages of the cycle. Basic specifications for the PSIcoil sensor system are summarized in Table 1.

Pressure range:	0 to 1000 psi
Over-pressure range:	2x pressure range
Linearity & hysteresis:	+/-1.5%
Frequency response:	0.01 Hz to 15 kHz
Sensor housing continuous temperature:	-40 to 350 °C
Electronics and opto-electronics continuous temperature	-40 to 125 °C
Sensor Output:	0-5 V
Life time	200 Million cycles

3. PRESSURE SENSOR DESIGN

The fiber optic sensor described here consists of three basic components, a transducer, a fiber optic cable, and a "smart" connector/module located at the proximal end of the cable, as shown in **Fig. 2**

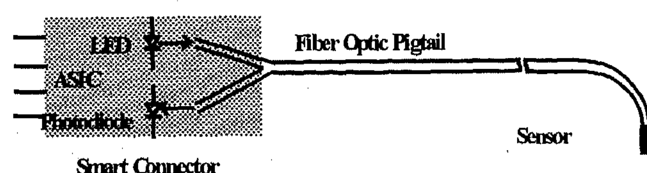


Figure 2. Sensor block diagram.

The "smart" connector consists of an opto-electronic module, containing a photodiode and a LED, and a dedicated ASIC. Two input electrical pins are for power supply and ground while two output pins are for sensor output and diagnostic voltages. The ASIC controls light intensity, amplifies and filters the photodiode signal, and provides the auto referencing function. The auto referencing function [4]-[5] regulates the sensor's input parameter, the LED light intensity, in response to any undesirable environmental conditions that may alter minimum detected light intensity. Based on the minimum detected pressure signal, the LED light intensity is adjusted in response to environmental conditions such that the minimum light intensity is always maintained at a constant level. This approach not only corrects for any possible offset drifts but sensor gain errors as well. In fiber optic sensors such errors may result from optical link transmission losses due to connectors' mechanical and thermal instabilities, fiber bending, or light source or detector temperature dependence. A side benefit of the auto-referencing technique, not available to other combustion pressure sensors, is the availability of sensor health monitoring output. By continuously monitoring the LED current level or its rate of change, one can identify potential sensor failure before it occurs. This ability is particularly important in control applications where sensor failure may cause malfunction or even failure of the controlled device.

The sensor head consists of a metal housing containing a sensing diaphragm laser welded to the sensor body, fiber

holding ferrule, and two fibers bonded inside the ferrule, as schematically shown in **Fig. 3**. Note that the dimensions are not show to scale.

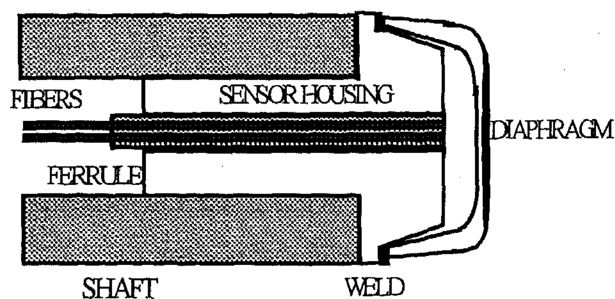


Figure 3. Sensor head construction.

The sensor response to pressure results from the displacement of a diaphragm that in turn results in changes in an optical signal collected by a receiving fiber. The light is emitted by a separate fiber. For a given diaphragm displacement, resulting from a pressure change, the sensor response can be adjusted by a suitable choice of optical fiber core diameters and numerical apertures, as well as relative position of the fibers in respect to the diaphragm. **Fig. 4** shows how the normalized amount of light collected by the receiving fiber changes as a function of the fiber to diaphragm distance.

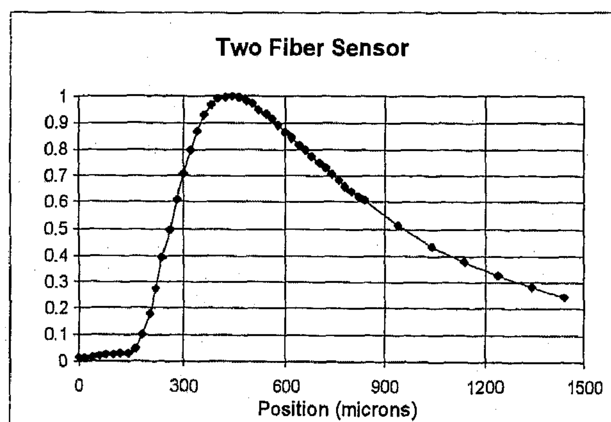


Figure 4. Normalized light deflection versus diaphragm fiber separation

Note that the response curve of **Fig. 4** is non-monotonic with a positive sensitivity (derivative of the curve) region for smaller fiber to diaphragm distances and a negative sensitivity region for larger separations. The positive response region is typically more sensitive to diaphragm displacement and is therefore preferred for pressure sensor applications. In a typical design the diaphragm is located approximately 200 microns away from the fiber tip when no pressure is applied. As this distance is optimized, the pressure response can be extremely linear, as shown in **Fig. 5**. The modulation parameter of **Fig. 5** is defined as a percentage of returned signal at full scale deflection to a signal at zero deflection.

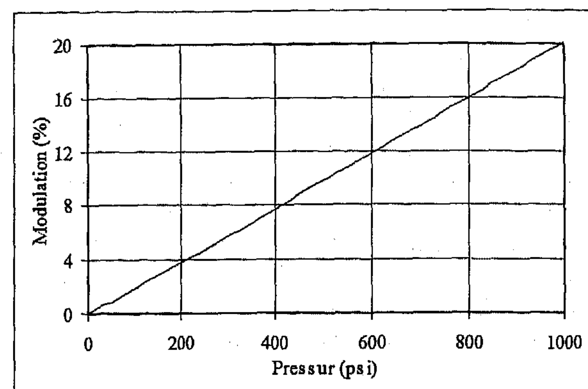


Figure 5. Calculated optical modulation versus applied pressure.

The primary design objective in designing the sensor diaphragm is to maximize its deflection while minimizing stress levels. Our proprietary hat-shaped diaphragm [6] made of Inconel offers an excellent high temperature strength, minimum stresses at nominal and over-pressure conditions, low creep, and good deflection characteristics. For a typical passenger car engine operation of 10 years, the sensor needs to meet the life-time requirement of approximately 200 million pressure cycles. Using an optimized diaphragm design the fatigue life time of 1 billion cycles is possible for the present sensor [4]. Other benefits of the present diaphragm construction include excellent linearity of pressure response and reduced sensitivity to direct flame effects. Finally, the diaphragm is welded to the sensor housing away from the flexing area improving long term weld stability and strength.

In annual production quantities of 1 Million the current sensor is expected to add between $7 and $10 to the cost of a conventional ignition coil. Unlike the competing electronic designs based on a piezoresistive diaphragm [7]-[8], our sensor has only a few components and its design lends itself to a low-cost manufacturing process. The ignition coil-integrated design of the fiber optic sensor package does not require a separate cable nor a connector, leading to minimum coil modifications. Finally, through the sensor mounting in a spark plug no addition installation costs are incurred associated with engine head modifications required for stand alone sensors.

4. LABORATORY AND ENGINE TEST RESULTS

Laboratory testing of the fiber optic sensor included linearity and hysteresis comparisons against commercially available piezoelectric sensors, temperature soaking and cycling, and sensitivity dependence on temperature. **Fig. 6** demonstrates a comparison of the typical room temperature dynamic pressure response, over 0 to 1000 psi pressure range, of OPTRAND's sensor and a reference piezoelectric transducer (PZT) (Kistler Model 6121).

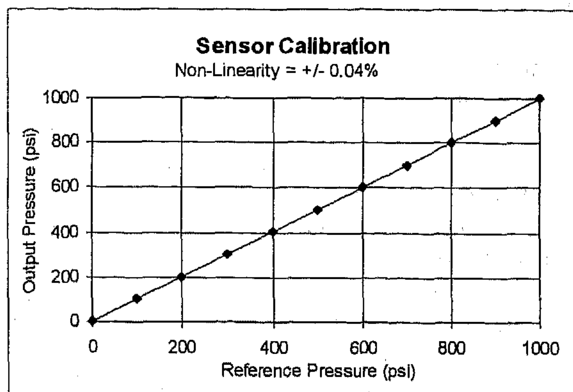


Figure 6. Dynamic pressure comparison between fiber optic and piezoelectric transducer at constant temperature

Note excellent non-linearity exceeding typical values of many laboratory devices. The signal to noise ratio (SNR) of approximately 2000 is comparable to that a reference transducer.

In a series of engine tests the performance of a spark plug-mounted fiber optic pressure sensor was compared to that of a piezoelectric reference transducer directly mounted in the engine head. For added accuracy the reference transducer was coated with a thin film layer of RTV and was installed behind a flame arresting hole previously identified as resulting in reducing thermal shock comparable to a water cooled sensor performance [9]. **Figs.** 7-9 demonstrate the results obtained on a single cylinder, gasoline-fueled motor generator set.

Fig. 7 shows the performance obtained with a sensor designed for nominal 1,000 psi pressure installed in OPTRAND's measurement spark plug, without a heat quenching feature in the spark plug. The left side vertical axis is for both optical and reference transducers while the right side axis shows the difference between the sensor readings. The measurement and reference traces are normalized so their peak to peak values are equalized.

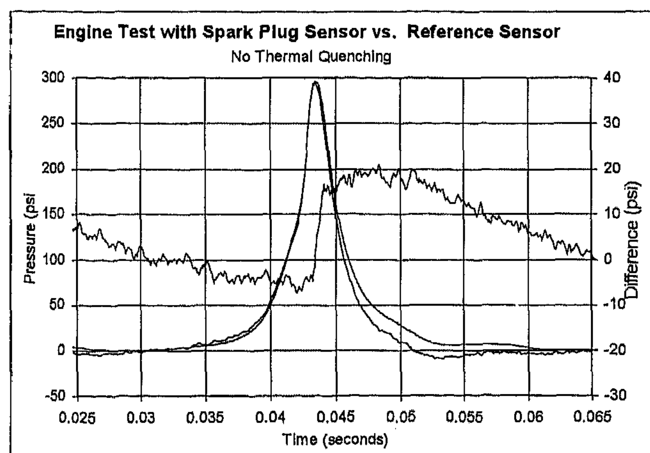


Figure 7. Engine test data for a spark plug-mounted sensor without heat shield

The maximum difference (error) between the sensors ranged from -7 psi to +20 psi. For a 1,000 psi sensor such an error corresponds to +2%, -0.7% full scale error, respectively. This error is due to the thermal shock effect [9] associated with intra-cycle changes in diaphragm temperature resulting from varying combustion gas temperature. Such an error is present in all types of diaphragm-based sensors including the widely used reference piezoelectric sensors. While the spark plug used in the present evaluation was equipped with a flame arresting hole connecting the sensor mounting cavity and the pressure channel, the hole alone does not provide the necessary cooling of combustion gasses before they impinge on the diaphragm.

To improve upon the thermal shock performance of the spark plug-mounted sensor, an additional heat quenching feature was incorporated in the spark plug design. This proprietary design reduces the direct effect of combustion gasses on the diaphragm without compromising the sensor frequency response. **Figs. 8-9** show the excellent performance of this new design spark plug sensor. A comparison of time-based response is shown in **Fig. 8** while **Fig. 9** illustrates a direct comparison between the two sensors demonstrating approximately +/-1.6% combined linearity-hysteresis performance of the enhanced spark-plug mounted fiber optic sensor.

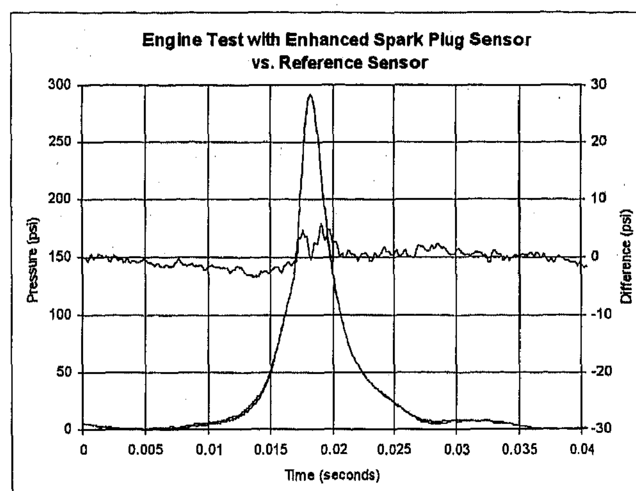


Figure 8. Engine test data for a spark plug-mounted sensor with heat shield

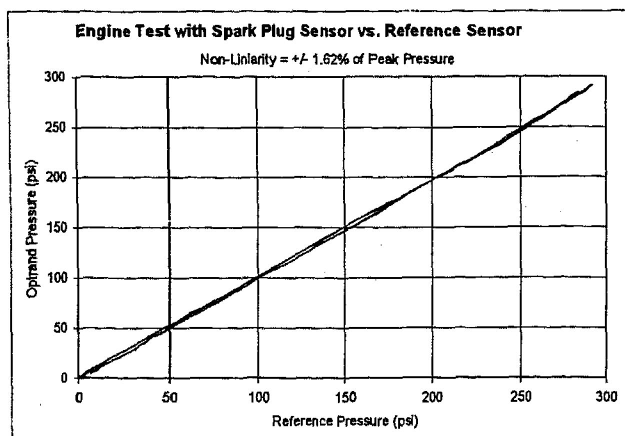


Figure 9. Linearity-hysteresis comparison between a spark plug-mounted fiber optic sensor and an engine head-mounted piezoelectric reference transducer

5. SUMMARY AND CONCLUSIONS

We have presented the design and performance of a fiber optic combustion pressure sensor integrated with an automotive ignition system. A miniature sensor is attached to a shaft extending from the spark plug. boot. The sensor is inserted into a modified spark plug during the process of boot pressing onto the spark plug connector. A smart auto-referencing technique compensates for sensor and fiber link environmental and temperature effects and guarantees drift-free operation. Using the optimized heat quenching design of the spark plug mounted sensor, the combined linearity and hysteresis of the present sensor stands at +/- 1.5%, demonstrating performance comparable to that of a laboratory grade piezoelectric transducers. With the targeted life-time of 200 Million pressure cycles and price between $7 and $10, the sensor meets the performance, durability, and cost requirements for production car applications.

6. ACKNOWLEDGMENTS

The work reported in this paper is funded in part by a Phase II SBIR research grant number DAAE07-96-C-X107 from the Defense Advance Research project Agency, monitored by the U.S. Army Tank and Armament Command in Warren Michigan.

7. REFERENCES

[1] "Cylinder Balancing System for Gas Engines", Diesel & Gas Turbine Worldwide, March 1995.

[2] N. Sugitani, and K. Tsukada, "A combustion pressure sensor for TOYOTA lean burn engine control system," SAE paper no. 930882, presented at the 1993 SAE International Congress and Exposition, Detroit, MI, March, 1993.

[3] G. M. Beshouri, "On the Development of Modern Analysis Techniques for Single Cylinder Testing of Large-Bore Engines," ASME Trans. Vol. 113, July 1991.

[4] T. Poorman, S. Kalashnikov, M.T. Wlodarczyk, "In-Cylinder Fiber-Optic Pressure Sensor for Engine Monitoring and Control," ASME Trans. ICE-Vol. 26-2, April 1996.

[5] Patent pending

[6] Patent pending

[7] W. Herden and M. Kusell, " A New Combustion Pressure Sensor for Advanced Engine Management", SAE Technical Paper Series 940379

[8]. T. Itoh, T. Adachi and H. Hashimoto, "One-Chip Integrated Pressure Sensor," 11th Sensor Symposium, 1992, pp. 241-244.

[9] A. Randolph, "Cylinder-Pressure-Transducer Mounting Techniques to Maximize Data Accuracy" SAE paper no. 900171, presented at the 1990 SAE International Congress and Exposition, Detroit, MI, February 26 - March 2, 1990.

970846

Dynamic Film Measurements in Journal Bearings Using an Optical Sensor

Tory M. Jaloszynski
Cummins Engine Co.

Lawrence W. Evers
Michigan Technological Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The primary goal of this research was to investigate the feasibility of measuring the oil film thickness in a journal bearing using an optical film thickness sensor earlier developed at Michigan Technological University. A bearing oil film thickness testing facility was designed to provide precise control of the sensor calibration and experimental measurements. The calibration of the sensor was achieved by the use of an in-bearing calibration setup that allowed manual changes of film thickness by moving the bearing sensor within the bearing. By varying the speed from 350 to 1100 RPM and the load from 0 to 55 kg in a journal bearing, the sensor was able to measure changes in oil film thickness from 0 to 25 μm. The trends caused by changing the load or speed were seen to be consistent with a mathematical model based on the two dimensional Reynolds equation for journal bearing design. The film thickness measurement technique developed was successful in evaluating the performance of a journal bearing under several operating conditions and should provide an excellent tool to further investigate bearing performance.

INTRODUCTION

The objectives of this research were the following:

1. Determine the feasibility of using the optical film sensor in the measurements of oil film thickness in a journal bearing.
2. Design a bearing oil film thickness test facility which can be used to perform in-bearing calibrations of the optical sensor.
3. Obtain accuracy and reproducibility in the calibration of the optical film thickness sensor in the testing facility.
4. Show dynamic film measurements obtained are consistent with expected bearing performance based on a mathematical model for speed and load effects.
5. Show transient response of the optical sensor during start-up.

These objectives were developed in response to the need for an unintrusive measurement system that allows the measurement of the actual oil film thickness. Accurate measurements of the oil film thickness in a journal bearing of an operating engine have been difficult to obtain because of the small, rapidly changing clearance seen in a journal bearing. Numerous different experimental techniques have been proposed to accomplish this objective, such as [1,2,3]. Many techniques have been developed that include the use of discrete probes or transducers located at various points around the circumference of the bearing. These probes have been designed to measure different physical quantities such as capacitance, magnetic fields, resistance, inductance, light waves, and eddy currents. The optical sensor for this research uses a single optical probe and reflected light to determine the changes in oil film thickness in a journal bearing.

The design of journal bearings is important to the development of rotating machinery such as compressors, pumps, turbines, internal combustion engines, motors, generators, etc. In its most basic form, a journal bearing consists of a rotatable shaft, called the journal, contained within a close-fitting cylindrical sleeve, called the bearing. The journal and bearing surfaces are separated by a film of lubricant that is supplied to the clearance space between the surfaces of the journal and bearing. The clearance is a critical factor in bearing design and has considerable effects on the assembly of the journal and bearing. The clearance provides space for the lubricant, allows for thermal expansion encountered in any harsh environments, and also provides tolerance for any shaft misalignment.

The internal combustion engine provides several excellent examples for uses of journal bearings. The journal bearings found on an engine's crankshaft main bearings, connecting rod bearings, and any accessory or balance shaft bearings provide sources of friction. The loads on these crankshaft journal bearings will be constantly changing, not only in magnitude, but also in direction. This is caused by the inertial loads of the piston/crank assembly and the pressures in the cylinder from combustion.

There is also considerable interest in reducing the fuel consumption in gasoline and diesel engines by using low viscosity lubricants [4]. As the viscosity is reduced, however, a point is eventually reached where metal-to-metal contact occurs in the bearing. When this occurs, an undesirable increase in friction

and wear is experienced. The use of low viscosity lubricants in bearings that were designed to run with higher viscosity oils can be catastrophic. Engine bearing performance can also be influenced by the passage of oil contaminants through the clearance of the journal bearing. The parameters that control how the bearing and the journal are separated could be clearly investigated using a controlled experimental system that provides accurate oil film thickness measurements.

In theory, the oil film thickness at a given set of conditions in a journal bearing could be compared with an extensive list of rheological properties for each oil. A statistical analysis of this information should be capable of identifying which known variables of oil dynamics have the greatest effect on the oil film thickness and whether any unknown variables are influencing the film thickness in a bearing. An accurate measurement system would allow analysis of current bearing design standards and indicate areas that need improvement.

BEARING MATHEMATICAL MODEL

Reynolds developed the theory of journal bearings based upon the concept of hydrodynamic equilibrium when the shaft is running at sufficient speed to support the shaft. Reynolds first suggested that if a lubricant is adhering to both the bearing and journal surfaces, so as to create a fluid pressure of sufficient intensity to support an external load, then the velocity, friction, and pressure distribution in the thickness of the lubricant film can be determined. These parameters can help to determine other unknown parameters required for complete journal bearing design. The hydrodynamic theory was based on the several simplifying assumptions. These assumptions are listed below.

1) Fluid film is thin enough to replace the curved partial bearings with flat bearings and the pressure variation over its thickness can be neglected.
2) Lubricants obey Newton's law of viscous fluid (Fluid is "Newtonian" or $\tau=\mu du/dy$).
3) Forces due to inertia and acceleration of lubricant are neglected ($F=ma=0$).
4) Lubricant is an incompressible fluid (flow in = flow out).
5) Viscosity of fluid is assumed to remain constant even though the inlet temperature and outlet temperature of lubricants are different (Average temperature value is used in any analysis or design).
6) The fluid experiences laminar flow with no slip at the boundaries.

With these simplifying assumptions, Reynolds began with the equations of equilibrium of forces on the fluid element shown in Figure 1. Using the above assumptions, the equilibrium of the forces shown on the differential element, and when end leakage fluid flow is included, the Reynolds equation for two dimensional flow can be developed, as shown in equation(1).

$$\frac{\partial}{\partial x}\left(\frac{h^3\partial P}{\mu\,\partial x}\right)-\frac{\partial}{\partial z}\left(\frac{h^3\partial P}{\mu\,\partial z}\right)=-6U\frac{dh}{dx} \quad \text{Eq (1)}$$

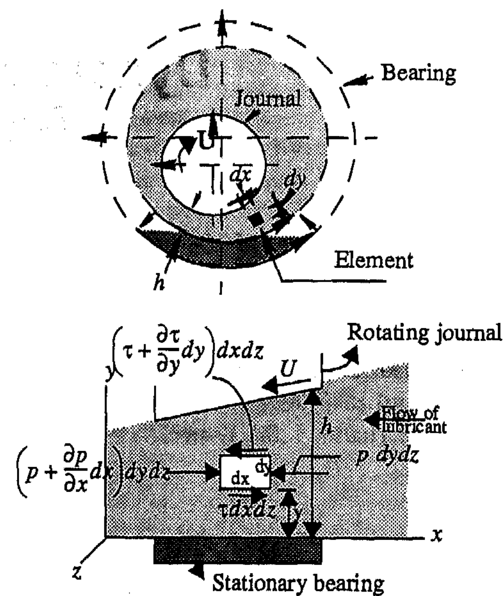


Figure 1. Shear element of oil in a journal bearing.

In Figure 1 and Equation (1), P represents the normal forces, due to pressure, acting upon the element. τ represents the shear forces, due to the viscosity, μ, and the velocity, U, and h is the thickness of the film. This differential equation has no analytical solution, therefore, the only solution that exists is a numerical solution. Raimondi and Boyd solved the Reynolds equation and presented their results in the form of several charts and graphs [5].

Dr. S. R. Bhonsle at Michigan Technological University has taken the graphical form presented by Raimondi and Boyd and digitized it in a numerical model. This model provides solutions for bearing geometries quickly and efficiently. This is a great improvement over the reading of the many charts and graphs presented by Raimondi and Boyd, which can be a tedious task and very susceptible to errors. Also, the design analysis is based on an average viscosity of the lubricant, which is based on the average temperature of the lubricant. At best, one only knows the inlet temperature of the lubricant. Therefore, an iterative process is required to arrive at the average temperature which can then be used to determine the viscosity of the lubricant used in the journal bearing. This, again, is time consuming and susceptible to errors. The output from the mathematical model can be seen in [6].

The experimental data obtained in this research was compared to the film thickness computed from the minimum oil film thickness and location of minimum oil film thickness from this mathematical model. Changes of the film thickness associated with changing the speed and load at which the bearing was operated were investigated in this research.

OPTICAL FILM SENSOR MEASUREMENT SYSTEM

The oil film thickness was measured in a journal bearing with an existing optical technique developed at Michigan Technological University under U.S. Patent Number 5,396,079. The

optical technique has been used for other studies, such as measuring the fuel film thickness in the intake port of an engine [7,8]. The film thickness sensor measures the depth of a fluid over a surface as shown in Figure 2.

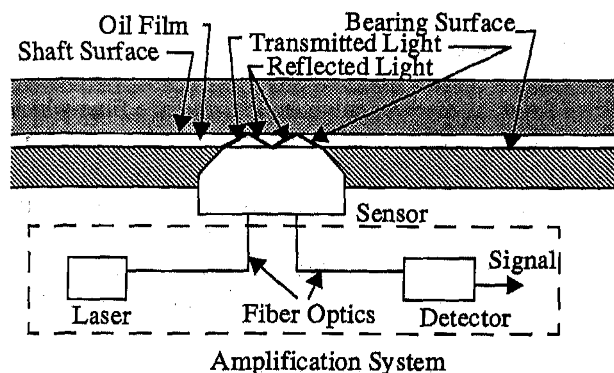


Figure 2. Overall film thickness measurement system.

The sensor is mounted flush to the bearing surface. The laser sends light to the sensor through fiber optics and forms a cone of light in the oil film. The light is reflected back to the sensor based upon the depth of the film. Fiber optics transmit the light to the detector which provides a signal based on the film thickness.

The operation of the sensor was developed using optical theory that states when a beam of light is sent from one medium to another, the degree of reflection, Θ, is determined by the angle of delivery, or angle of incidence, and the indices of refraction, n, of each medium. These angles can be determined by Snell's Law, as indicated by equation 2.

$$n_1 \sin\Theta_1 = n_2 \sin\Theta_2 \qquad \text{Eq (2)}$$

where subscript 1 denotes the characteristics of the medium in which the light is leaving and subscript 2 denotes the medium in which the light is entering. The path of a light ray according to this equation travelling from the sensor into the oil can be seen in Figure 3.

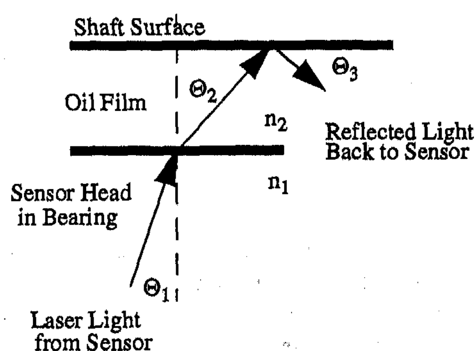


Figure 3. The effect of the different indices of refraction on a light ray traveling from the sensor into the oil film.

The angle Θ_1 is limited by a term called the critical angle. When Θ_1 is at the critical angle, the light cannot leave the sensor. The angle Θ_2 becomes 90° and results in no cone of light to be formed above the sensor. The critical angle for the sensor must not be reached to ensure proper operation of the measurement system.

Figure 4 illustrates the cylindrical optical film sensor. The optical sensor is manufactured with three main fiber optic bundles: an outer transmitter, an outer receiver and an inner receiver. The laser light enters the sensor from the transmitter bundle, placed around the perimeter of the sensor, and leaves through the outer transmitter section.

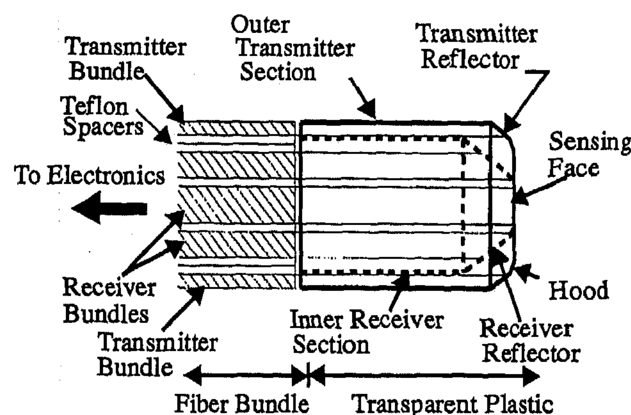


Figure 4. Sensor components and configuration.

The reflector surfaces are coated with a silvering chemical substance to enhance the reflection of the laser light and the other surfaces are painted black to reduce the reflection of stray light. The light, reflected from the cylindrical transmitter reflector, forms a cone of light in the oil film.

As the laser light enters the oil, a certain amount of its beam will be reflected back to the sensor depending upon the thickness of the oil film. The reflected light is gathered by the centrally located receiver section and sent to the detector. For zero film thickness, the voltage is also zero. As the depth increases, the amount of light returned to the receiver increases and the corresponding voltage increases.

INDEX OF REFRACTION MEASUREMENTS

The most important design variable was the angle of the transmitter reflector and the bending of the light leaving the plastic sensor head and entering the oil film. This angle was calculated from the index of refraction of oil, the plastic sensor, and the chosen maximum depth that the sensor can measure. Because there was no reference material available that gave the index of refraction for engine oils, measurements were taken to determine this parameter. Following the test standards developed according to ASTM D 1218-92 and D 1747, a Bausch&Laumb Abbe Refractometer was used to determine the index of refraction for several different weights of engine oils. Table 1 lists the indices of refraction for several different weights of engine oils and other liquids. The fluids where only one trial was given indicates a difficulty in measurement.

Table 1: Index of Refraction for Various Engine Oils

Fluid	Trial 1	Trial 2
Water	1.330	1.332
Indolene	1.4275	-
Unocal 20W-20	1.4822	1.4820
Pennzoil SAE30HD	1.4810	1.4818
Castrol 10W-40	1.4754	1.4760
Havoline 10W-40	1.4790	1.4787
Valvoline 5W-30	1.4758	1.4760
Phillips 66 SAE 40	1.4809	1.4810
Amoco SAE 30	1.4845	1.4845
ISO-32 Hydraulic	1.4760	1.4790
Meriam Oil SG=0.827	1.3220	1.4761
Dot3 Brake Fluid	1.44	-
DER331 EpoxyResin	1.56-1.57	From Textbook
Cured w/ MDA	adds 0.04 to Resin	From Textbook

MEASUREMENT ELECTRICAL SYSTEM

Figure 5 shows the flow diagram of the light signal in the measurement system.

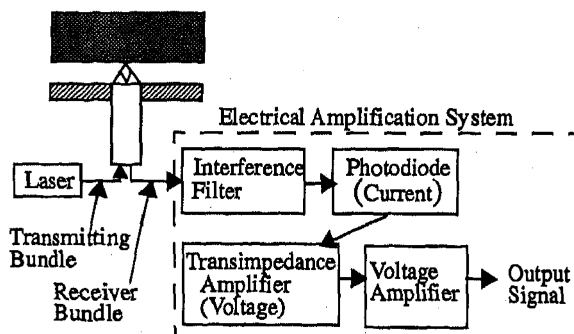


Figure 5. Diagram of film sensing signal.

A 3 milliwatt solid state diode red laser source is used to send the light signal, with a wavelength of 670 +/- 10 nanometers, from the amplification box to the journal bearing via a single plastic optical cable. This plastic cable was chosen in order to increase the light delivery efficiency and for its flexibility. For the wavelength of the laser, the cable has a loss of only 0.15 dB/meter [3]. This plastic cable is then connected to the glass fiber bundle which carries the light to the reflector surface of the sensor. Glass fibers were used in the region close to the engine because of the high temperatures that the sensor could encounter. The reflector surface sends the light at a specific angle to the sensing surface

The light from the receiver enters an interference filter which allows a very narrow bandwidth of wavelengths to continue. The bandwidth that the filter allows includes the wavelength of the laser light that was used. A filter was used to alleviate the possibility that light of other wavelengths would contaminate the measurements. The light exiting the filter directly enters the photodiode which transforms the light intensity to an electrical current. A phototransistor could also be used for this task since it provides a high level of internal gain.

However, the linear range of a phototransistor is smaller than that of a photodiode. The phototransistor also shows some sensitivity to temperature changes. Therefore, the diode was used with a higher powered, 3 milliwatt laser, to maintain a high enough output signal for the measurement range of interest. The transimpedance amplifier changes the current to a voltage signal. The interference filter, photodiode, and transimpedance amplifier were purchased as an integrated package to reduce losses between the filter and amplifier. It is also advantageous to include the transimpedance amplifier close to the diode since its output is on the order of milliamps. The final stage is a voltage-to-voltage amplifier which boosts the output level of the signal so that higher voltage readings can be acquired. A low pass filter is used after the signal is amplified to filter high frequency noise from the signal. The zero is adjusted using a voltage divider on the transimpedance amplifier when no light is entering the sensor. The gain is adjusted using the voltage amplifier with the sensor at a predetermined film thickness.

The entire package, laser and electrical circuit, was installed into one amplification box that was located near the engine. The zero and span functions were provided by the summing and gain amplifiers which could be adjusted through control knobs on the outside of the box. The output from this amplification box was sent to the Hi-Techniques IQ300 data acquisition system.

BEARING OIL FILM THICKNESS TESTING FACILITY

The testing facility for the bearing sensor required the incorporation of several key features that would allow the detailed experimentation of the journal bearing parameters. Of most interest in journal bearing research are the effects of speed, load, and viscosity of the oil. Figure 6 shows the overall setup of the bearing film thickness testing facility.

A 6KW Tecumseh horizontal shaft engine with a journal diameter of 30.15mm and a diametral clearance of 0.064mm was used for the experimental bearing. The clearance of the exper-

imental bearing was well within the designed maximum film thickness for the bearing sensor of 0.15mm. The most notable features of the test facility are the loading setup, A, and the in-bearing calibration setup, G. These features will be explained in full detail later in this section.

Other features of the bearing oil film thickness testing facility include an encoder, oil pump, oil flow meter, thermocouples, contoured sensor head, and placement of the electric motor. The Disc Instruments incremental shaft encoder, B, was included in the facility in order to increase the accuracy of the shaft rotational speed. The encoder output 360 pulses per revolution and also one top dead center pulse. The time between top dead center pulses was used to calculate the rotational speed. An electronic speed controller was used to maintain constant rotational speed.

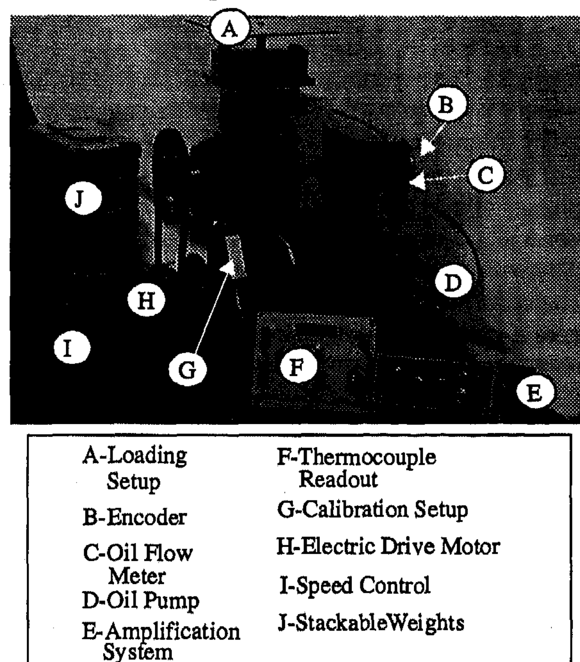


Figure 6. Overall view of the bearing film thickness testing facility.

A shaft of comparable diameter was machined to replace the crankshaft in order to allow a load to be applied. Since the original engine used a splash-type lubrication system that was obtained by the spinning of the crankshaft and dipping the connecting rod down into the oil, a different lubricating system needed to be added. The variable speed oil pump, D, was included to continuously supply the engine bearing with a controlled amount of oil. In conjunction with the oil flow meter the amount of oil fed to the bearing was precisely controlled.

A test stand was also constructed to locate the electric motor in relation to the engine. The electric motor, H, was placed below the engine in order to transmit the belt tension in the vertical direction. This would generally act as a weight applied to the shaft instead of pulling the shaft in a different direction. E-type thermocouples were also included in both the inlet of the oil pump and the outlet of the bearing in order to monitor the temperature of the oil during experimentation.

Since the calibration of the optical sensor proved to be difficult to obtain outside of the bearing, a simple device that would allow for calibration of the sensor directly in the bearing was designed and incorporated on the bearing oil film thickness testing facility. It was desired to calibrate the sensor directly in the bearing because it was necessary to maintain an oil film above the sensor at all times and, at the smaller film thicknesses, squeeze the oil film out from between the shaft and sensor. This new design was also beneficial because a calibration could be performed and then run directly without having to mount the sensor. This design would allow the sensor to be continuously calibrated and also be removed if necessary for inspection. This new calibration setup can be seen in Figure 7.

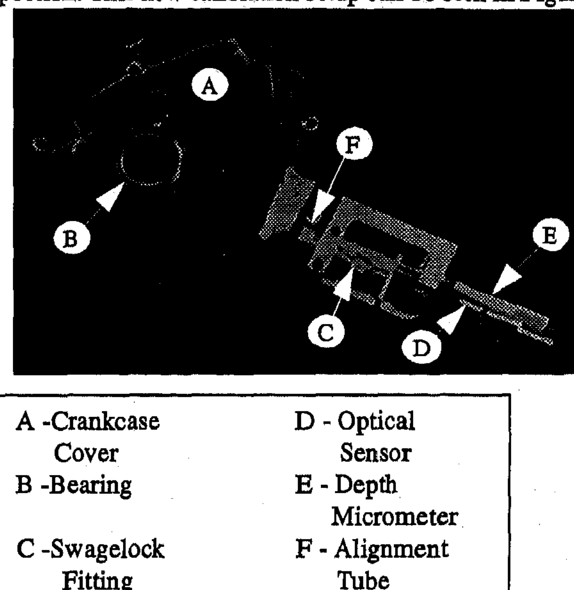


Figure 7. New in-bearing calibration setup.

From Figure 7, the bearing sensor, cast in a piece of tubing, mounts into an alignment tube that guides the movement of the sensor in and out from the bearing. A Swagelock fitting with nylon ferrules was incorporated around the bearing sensor to provide an oil pressure seal. When the Swagelock fitting was loosened, the tubing was allowed to move in or out and induce a film change directly in the bearing. A depth micrometer was used to measure this depth change. The Swagelock fitting was then tightened to provide a sufficient seal for the oil pressure to develop normally within the bearing.

From Figure 8, it can be seen that the bearing sensor was cast into a length of stainless steel tubing and the end ground to match the contour of the bearing surface. The sensor face was contoured so that the oil dynamics within the bearing would be minimally disturbed by any type of profile change.
The only effect may come from the differences in materials.

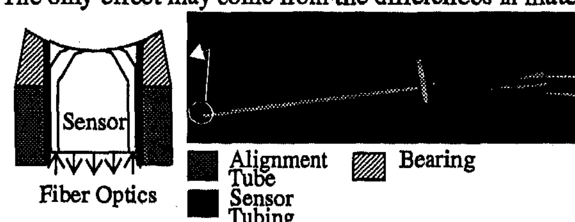


Figure 8. Final sensor mounted in tubing.

The bearing sensor installed in the bearing can be seen in Figure 9.

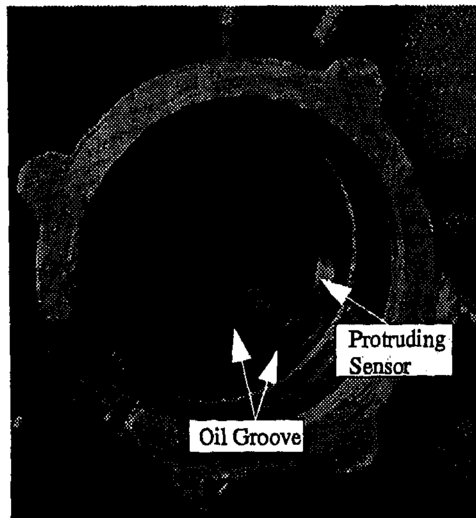


Figure 9. Bearing sensor protruding into bearing.

Another feature of the bearing film thickness testing facility was a loading setup used to load the shaft. The shaft was connected to a connecting rod and attached to another rod and plate that was used to load the shaft. An adjustable block was incorporated in the design and allowed the angle of the load to be applied at various angle from 30° counterclockwise to 30° clockwise from vertical. On top of the loading plate, weights of 9kg increments were stacked to apply a load to the shaft. This setup allowed the flexibility to apply the load while the shaft was running and also when the shaft was at rest.

CALIBRATION OF BEARING SENSOR

A calibration of the bearing sensor was performed using the in-bearing calibration setup. The engine was assembled and the shaft was pulled tight to the bearing, and held away from the sensor. The calibration was performed with the shaft not rotating and the bearing filled with oil. This calibration curve can be seen in Figure 11.

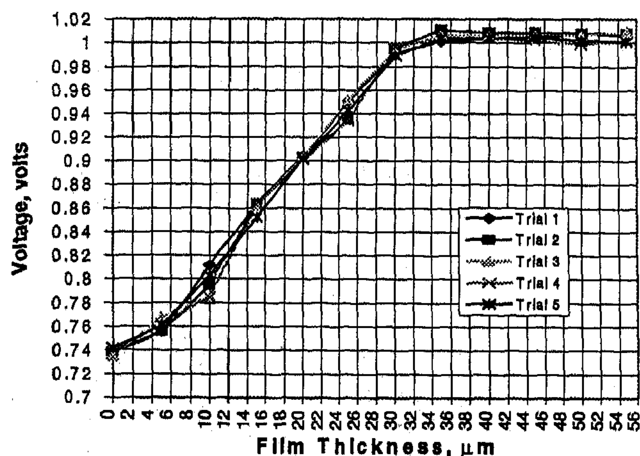


Figure 11. Calibration curve for bearing sensor.

Figure 11 shows that as the sensor moved closer to the journal, causing the film thickness to decrease, the voltage decreased. The calibration for the bearing sensor was performed five times under the same conditions. For each trial, the zero and gain of the amplifier were set when the sensor was 55µm away from the zero film condition. The overall change in voltage can be adjusted using the gain of the amplifier. Adjusting the gain, however, does not change the shape of the calibration curve.

The small variation that was seen in this calibration curve can be explained by the amplification system and calibration setup that was used. The amplifier used for calibration did have a small amount of drift. The flange on the tubing of the sensor did flex, so it was important to recheck the depth of the sensor in the bearing to obtain an accurate measure of the film thickness.

EXPERIMENTAL TESTING

STEADY STATE CONSTANT SPEED

Vertical Loads - The experimental portion of this research was to prove that the sensor was seeing trends similar to the trends predicted by the mathematical model. The first experiment performed was accomplished for a constant rotational speed and applying various loads to the shaft. For this experiment the sensor was locked into the flush position using the Swagelock fitting with Teflon ferrules. This experiment was done at constant speeds of 600 RPM and 1000 RPM for bearing loads ranging from 0 to 55 kg. The results from this experiment compared to the mathematical model for the same speeds can be seen in Figure 12.

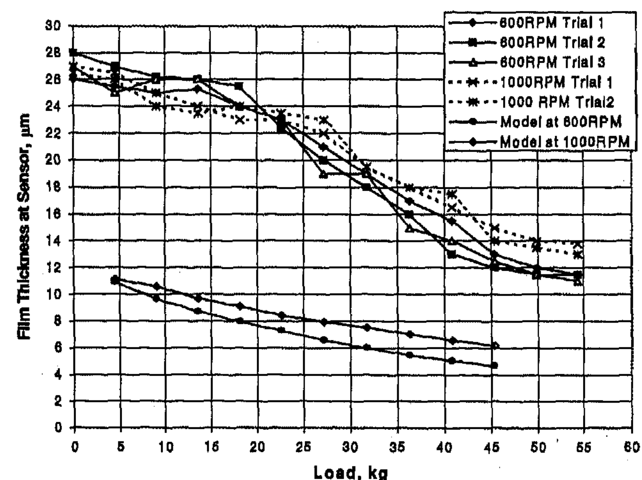


Figure 12. Experimental vs. model film thickness as a function of load.

Figure 12, shows that the general trend of increasing the load decreases the film thickness. This decreasing trend is consistent with the mathematical model. Figure 12 also shows that at the high load condition the film thickness for the 600 RPM condition is lower than the 1000 RPM condition. This trend is also consistent with the mathematical model. The large difference in magnitude between the experimental data and the mathematical model will be explained in the Differences Between Mathematical Model and Experimental Testing section.

Angled Loads - The same experiment was performed for the load applied at an angle of 30° directed toward the sensor at speeds of 600 RPM and 1000 RPM, which should cause the film thickness to decrease. The results from this experiment can be seen in Figure 13.

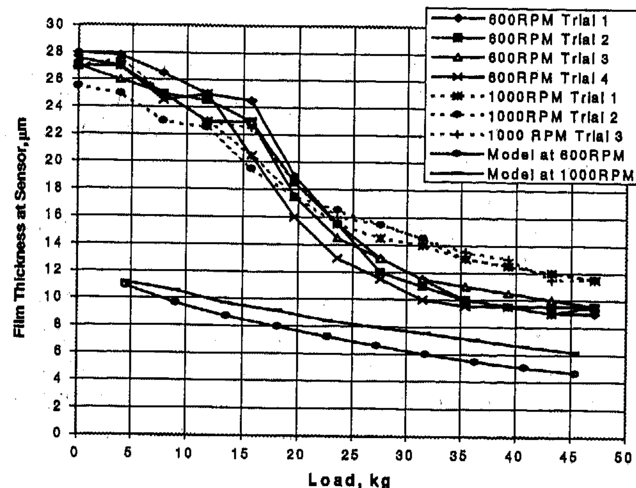


Figure 13. Experimental data for load at 30° angle vs. model data for vertical load.

Figure 13 once again shows the agreement between the decreasing trend of film thickness as a function of load at constant speeds between the experimental data and mathematical model. Another important result that can be seen from this figure is that at the higher loads, the difference between 600 RPM and 1000 RPM is more pronounced. The film thickness for the 600 RPM case was smaller than the 1000 RPM case. This was again consistent with the mathematical model predictions. Also, the film thicknesses at larger loads were less than the vertically applied load condition.

Another experiment was performed to test the effect of speed on the oil film thickness. By maintaining a constant load and increasing the rotational speed, similar trends should be achieved. This experiment was run for load cases of 18, 27, 36, and 45 kg of bearing load for a speed range of 375 RPM to 925 RPM.

STEADY STATE CONSTANT LOAD

Figure 14 illustrates the oil film thickness based on speed at constant loads. As the speed increased for all the load cases, the film thickness also increased. Also as the load decreased, the magnitude of the oil film thickness increased. The experimental data trend for each load case compares well with the predicted mathematical model trend. Once again, there is a large difference in the magnitude of the oil film thickness between the experimental data and the mathematical model.

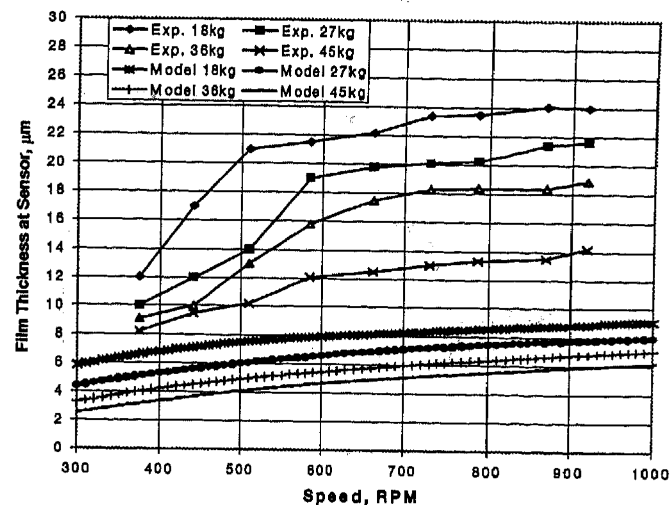


Figure 14. Film thickness at sensor as a function of speed vs. mathematical model.

Even though the trends compared well with the mathematical model, there were differences between the overall change in oil film thickness. A comparison between the change in oil film thickness for the vertical load case and speed case can be seen in Table 2 and 3.

Table 2: Change in Film Thickness for a Change in Load from 18 to 45 Kg for Constant Speeds of 600 and 1000 RPM

	Oil Film Thickness, µm	
Speed,RPM	Experimental	Model
600 RPM	12.5µm	3.5µm
1000 RPM	9µm	3.5µm

Table 3: Change in Film Thickness for a Change in Speed from 375 to 925 RPM for Constant Loads of 18 and 45 Kg

	Oil Film Thickness, µm	
Load, Kg	Experimental	Model
18 Kg	12µm	3µm
45 Kg	6µm	3µm

DYNAMIC TESTING

No Load Condition - The goal of this research was also to show the transient response during the start-up of the journal. Figure 15 shows the start-up of the bearing with no load to 260 RPM. As the speed is increased, the film thickness rises to an operating value.

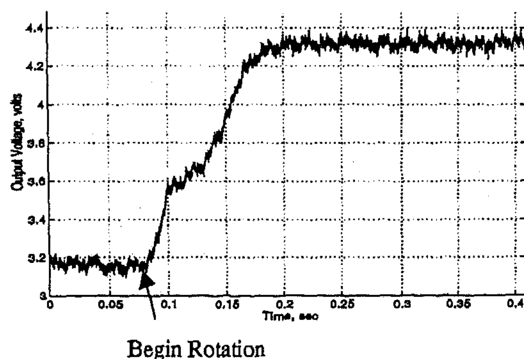


Figure 15. Start-up of bearing to 260 RPM with no load.

Loaded Condition - The effect of load on start-up was next investigated. This is shown in Figure 16.

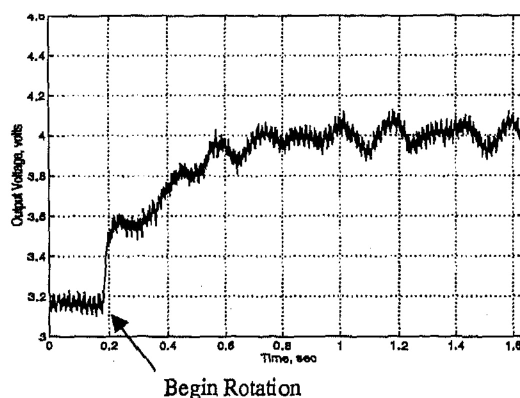


Figure 16. Start-up of bearing to 260 RPM with 36 kg bearing load.

From Figure 16, it can be seen that the start-up has a similar trend, but does not reach a constant value. The sensor seems to see the journal oscillating within the bearing. The origin of the oscillation was not investigated. Also, this average value is smaller and takes longer to reach than the no load condition. The lower voltage is a result of the load causing the film thickness to be reduced and therefore the voltage to decrease.

DIFFERENCES BETWEEN MATHEMATICAL MODEL AND EXPERIMENTAL TESTING

Several important reasons could explain the differences in the changes of thickness. The most important reasons for the differences are attributed to the assumptions used in the mathematical model. The first difference between the mathematical model and experimental conditions is the centrally located oil groove in the bearing. This oil groove can be seen in Figure 9. Oil grooving in a journal bearing provides two purposes: (1) to establish and maintain an efficient film of lubricant between the bearing moving surfaces; and (2) to provide adequate bearing cooling. The oil groove alters the pressure distribution within the bearing and significantly influences the bearing performance. The mathematical model provides conditions with no oil grooving and assumes a symmetric pressure profile across the width of the bearing. The central groove would seem to reduce the oil film thickness by altering the pressure distribution into two smaller regions across the width of the bearing.

However, an oil groove also permits more oil to pass through the bearing, thus keeping it cooler. Cooler oil temperatures result in higher viscosities. This aspect is the second major difference between the experimental data and the model. The thermocouples placed in the inlet and outlet positions of the bearing were used to substantiate the temperature of the oil. The mathematical model predicted a temperature change of 7.22°C, while the actual temperature change that was seen in the experiments was 2.78°C. When the shaft was running at 1000 RPM under a bearing load of 55 kg, the model overpredicted an increase in the outlet temperature. The resulting viscosity increase of 12.8 μreyn would tend to increase the oil film thickness.

A third influence that needs to be considered when comparing the model to the experimental data is the method of supplying oil to the bearing. The mathematical model assumes that the oil is supplied to the bearing at atmospheric pressure. Since the original engine supplies the bearings using a splash fed system, the film thickness could be influenced by supplying the oil with an oil pump. The splash fed system originally used the connecting rod to scoop oil from the crankcase and splash oil into the feeder holes of the bearing. For the oil pump system, hoses were set into the holes to feed the bearings. Even though there was no sealed system, a small increase could be attributed to the oil pressure supplied to the bearing by the oil pump.

The observed differences may also be attributed to a few more of the general assumptions used in the model. Three differences - that the oil is not fully incompressible, that the viscosity of the lubricant is not constant over the dimensions of the bearing, and that the fluid is not Newtonian - may have significant effects on the performance of the bearing. Newtonian fluids are defined as those oils for which the shear stress is directly proportional to the shear rate and for which the viscosity is only a function of temperature, pressure, and oil composition. The oil used in this research was Unocal 76 Guardol SAE 20W-20 API Service CD/SG and may contain several additives that alters the oil from being a Newtonian fluid.

ACCURACY OF FILM THICKNESS MEASUREMENTS

The accuracy with which the optical sensor can measure the oil film thickness in a journal bearing depends on the accuracy of the calibration. For the bearing film thickness testing facility, the metric depth micrometer used in the in-bearing calibration setup has an accuracy of 1μm. A statistical analysis of the calibration curve, Figure 11, shows that the observed output voltage at specified film thicknesses, on average, has a standard deviation of 0.004515 volts.

SUMMARY AND CONCLUSIONS

The following conclusions can be made regarding the evaluation of using the optical film thickness sensing system in a journal bearing application:

1) The film thickness sensing system is capable of measuring the dynamic changes of oil film thickness in a journal bearing.
2) Experimental data received from the bearing film thickness sensor is consistent with the mathematical model, which states that increasing the load decreases the oil film thickness and increasing the rotational speed increases the oil film thickness.
3) The bearing oil film thickness testing facility provided a satisfactory means to control load, speed, and oil flow.
4) The in-bearing calibration setup and technique resulted in a reproducible calibration of the bearing sensor.
5) The bearing sensor performs well, even though difficulties existed in sensor construction.

REFERENCES

[1] Murphy, C.K., and Spearot, J.A., "A Comparison of the Total Capacitance and Total Resistance Techniques for Measuring the Thickness of Journal Bearing Oil Films in an Operating Engine," SAE Paper No. 880680.

[2] D. P. Hoult, B. T. Shaw II, and V. W. Wong, "Development of Engine Lubrication Film Thickness Diagnostics Using Fiber Optics and Laser Fluorescence," SAE Paper No. 920651.

[3] Cooper, D., Moore, A.J., and Robinson, T.M., "Rheological Properties of Engine Crankcase and Gear Oil Components in Elastohydrodynamic Oil Films," SAE Paper No. 941977.

[4] Spearot, J.A., "High Temperature, High-Shear Oil Viscosity, Measurement and Relationship to Engine Operation," ASTM STP 1068, p.53, 1989.

[5] Juvinall, R.C. and Marshek, K.M., "Fundamentals of Machine Component Design," Second Edition, Hamilton Printing Company, 1991.

[6] Jaloszynski, T., "Dynamic Film Measurements in Journal Bearings Using an Optical Sensor", Masters Thesis, Michigan Technological University, 1996

[7] Bourke, M., "Fuel Film Dynamics in the Intake Port of a Fuel Injected Engine," Masters Thesis at Michigan Technological University, 1993.

[8] Coste, T., "An Optical Sensor for Measuring Fuel Film Dynamics in a Port Fuel Injected Engine," Doctoral Thesis at Michigan Technological University, 1996.

970847

Oil-Quality Prediction and Oil-Level Detection with the TEMIC QLT-Sensor Leads to Variable Maintenance Intervals

Eckard Irion
TEMIC

Klaus Land and Thomas Gürtler
Mercedes-Benz AG

Manfred Klein
Daimler-Benz AG

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The Quality-Level-Temperature-Sensor (QLT-Sensor) provides the dielectric number, level and temperature of fluids used in passenger car engines e.g. engine oils. The dielectric number can be used to indicate changes in oil-quality. Level detection can be used to determine oil consumption, minimum oil level or critical overfilling. The oil-temperature can be used in combination with the dielectric number for precisely predicting the oil drain time. The design-principles and functions, including compensation of aging and temperature effects are described. It is shown, what engine constructors and developers must consider when regarding mechanical and electrical interfaces.

INTRODUCTION

In general measuring oil quality alteration requires a large laboratory in the background, since numerous and not always the same parameters determine the oil's useful life. Essential parameters are e.g.: viscosity, additive concentration, contents of soot etc. Degradation of engine oil performance is a multidimensional problem and cannot be monitored sufficiently solely by one factor.

Today maintenance intervals of vehicles are generally set by a fixed mileage or time span due to the lack of available tools for determining exact performance degradation of various vehicle components. Normally the main maintenance determining factor is the engine oil. Its degradation is a function of how the engine is operated. Hence monitoring this operation would result in a flexible servicing system [1].

In search of a solution a practical approach was chosen: finding and measuring one parameter which showed a good correlation to oil degradation, especially to the condition of additives. In the laboratory numerous tests were conducted, which resulted in finding sufficient correlation between changes in the oil's dielectric number and its aging. This is nothing new [2] but up today no adequate methods had been available capable for doing this on-vehicle. When fresh oil is toped up or water condenses, the dielectric number is influenced. It is obvious, that such effects must be taken into account. Hence considering parameter changes in dielectric number and oil level effects in combination with a few engine parameters , oil drain period can be more precisely predicted.

Prototypes based on a capacitance measuring system are running successfully since 1993 in passenger car engines. In 1994 industrialization of the sensor started and volume production has commenced in August 1996.

FUNCTION OF THE SENSOR

The sensor cell consists of two cylindrical capacitors. One is totally immersed in oil. The capacitance depends on the dielectric number of the oil. The second capacitor is in a position, that the oil levels between the expected maximum and minimum levels can be measured. Its capacitance depends on the dielectric number of the oil and on the level in the sump. Figure 1 shows a schematic cross section of the sensor.

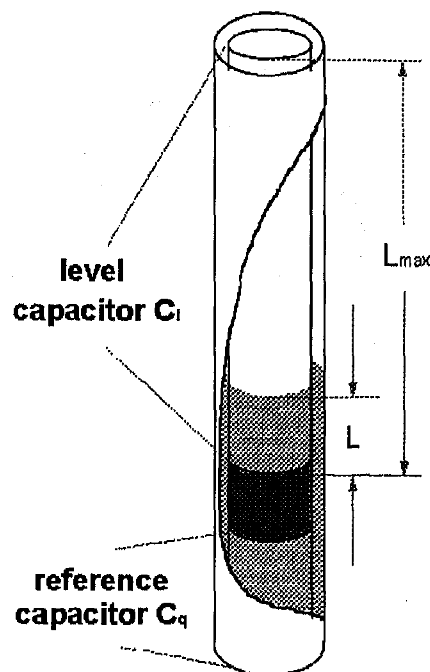


Figure 1: Schematic cross section of the sensor

The relative dielectric number of the oil is given through equation 1:

$$\varepsilon_r = \frac{C_q(oil)}{Const._1} \quad (1)$$

The oil level is given through equation 2 (derivation of the formula see APPENDIX):

$$L = \frac{L_{max} \times (C_l - Const._2)}{(\varepsilon_r - 1) \times Const._2} \quad (2)$$

Capacitance and resistance of a temperature sensitive device are measured by an electronic circuit, which converts the data into pulse width modulated signals. The signals are sent on a single signal line within a 300 millisecond time slot (100 milliseconds for each signal). After a pause of 1 second the next measuring cycle starts. The electronic circuit (Figure 2) is state of the art for resistance and capacitance sensitive circuits. It is composed of very precise reference devices like resistors and capacitors with lowest temperature coefficients, thus minimizing any temperature drift of the sensor. In a separate operation cycle the conductivity of the oil is measured. In the next cycle the capacitance are measured and the leakage current due to the conductivity of the oil is compensated through programmable current sources.

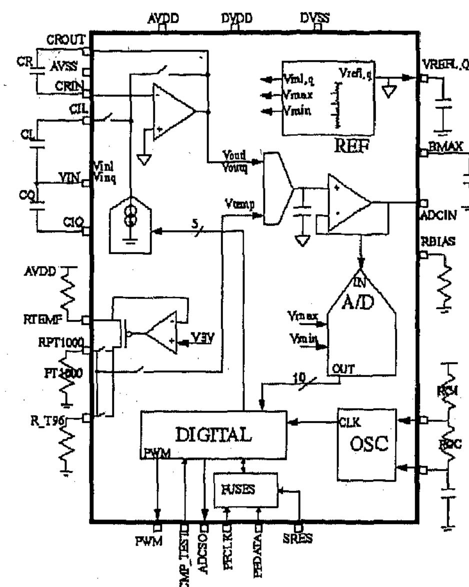


Figure 2: Block diagram of sensor control circuit

Thus a multifunctional sensor, measuring oil temperature and capacitance is realized. Dielectric number and oil level can directly be calculated with these results.

DESIGN OF THE SENSOR

The sensor has 4 major elements: a sensor cell comprising cylindrical capacitors and temperature sensor, a mechanical adapter for fixing the cell to the oil pan, a feed through including sealing elements, and an electronic housing with integrated connector.

The sensor cell consists of stainless steel tubes and spacer elements made of polyphenylene-sulfide resin. Oil tightness is realized by seal-rings made of synthetic rubbers. The tubes are connected to a printed circuit board in the electronic housing by shielded cables with crimped spring contacts. The temperature sensitive device is integrated in one of the spacer elements, forming good thermal contact to the oil. The housing of the sensor cell is made of plastic and has a well situated inlet hole for the oil, an opening for ventilation and the fixture to the oil pan. A seal-ring is designed into the sensor cell to prevent leakage through the oil pan. Of course all materials have a high temperature and chemical resistance, so that temperatures of 160°C and fluids like fuel, detergents and so on have no effects and influence on function and reliability of the components.

In addition the plastic parts, especially the part with the fixture to the oil pan were dimensioned in such a way that all expected thermal, mechanical

and chemical stress leads to no degradation of function.

The special choice of materials, as well as the parameters for molding have been chosen in accordance with FEM-simulations and verified through tests with a shaker in a climate chamber at high temperatures.

Of special interest was the material of the printed circuit board and the electronic devices, forming the electronic circuit. Well suited are epoxy-circuit boards made of FR4-05 material and passive elements like resistors and capacitors which are qualified for up to 150°C. The active part of the circuit is an application specific silicon chip (ASIC), designed in C-MOS technology. This technology has been tested over 6000h at 125°C and over 1000 h at 140°C with no failure.

Today interconnection technology is the determining factor for reliability. Only certified technologies like SMD-technology for the printed circuit board and crimping for the wiring have been chosen. The connectors are joined to the printed circuit board by pressing the contacts into the board forming a very tight reliable contact.

MODULARITY OF THE SENSOR

Various solutions for the sensor's mounting within passenger car engine's oil pans have been realized. For example, if there is not much space inside the oil pan, but enough space between oil pan and road surface, mounting from the bottom side can be recommended. Normally mounting from the bottom side solution is standard but requires a design-in time point at an early stage in the project. A mounting solution from the engine's upper side is also available, having advantages for designs with low road surface clearances. However complex the construction situation might be, there is always a way to get the sensor in the right position for sensing QLT (quality, level, temperature) since the modules as described above can be combined in various ways. As an example 3 construction types for different mounting situations are shown in the following figures (Figure 3, Figure 4, Figure 5):

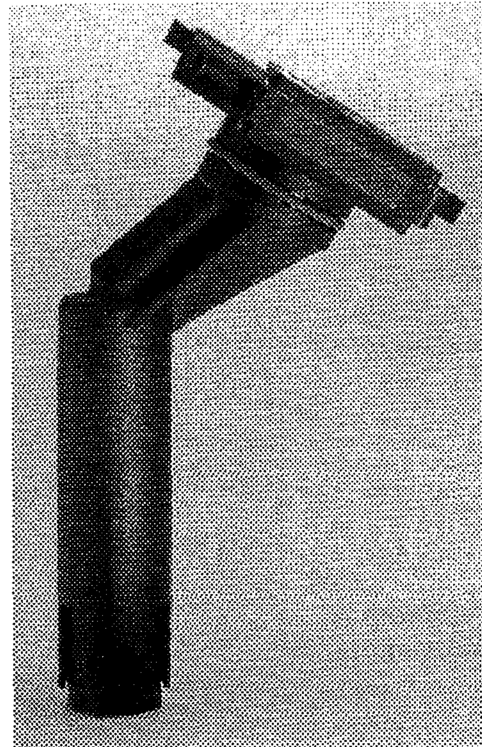


Figure 3: Sensor for mounting from the top side

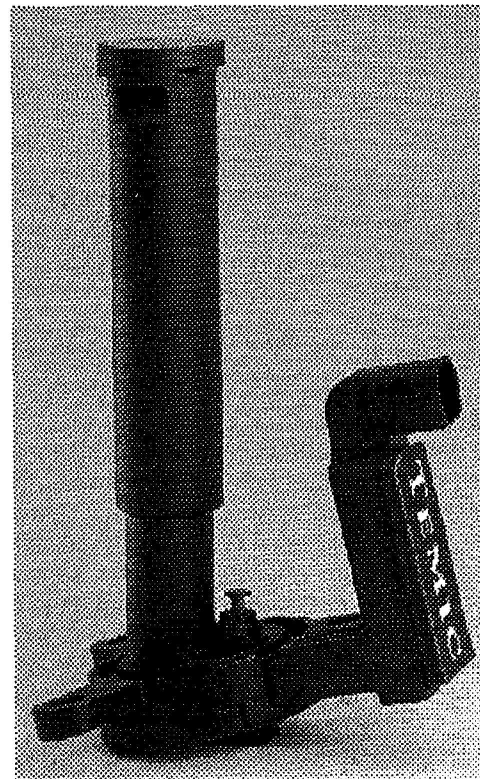


Figure 4: Sensor for mounting from the bottom side

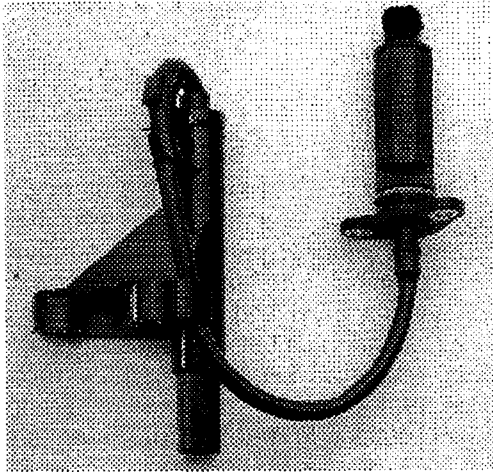


Figure 5: Separate sensor cell and an electronic housing connected through a flexible cable

MEASUREMENT RANGES

The sensor measures oil temperatures in the range of -40°C up to +160°C with an accuracy of ± 2°C. Dielectric numbers from $\varepsilon_r = 1$ up to $\varepsilon_r = 6$ are measured with an accuracy of ± 0.1. The range of the level measurement depends on the mechanical design. Ranges from 48mm up to 125mm are possible with an accuracy of about 3.8%, including the error contribution of all the signal conditioning circuits like sensor electronics, processing unit and display driving units.

MECHANICAL INTERFACE

The mechanical interface is of high interest for the engine developer. The fixture of the sensor must be able to fulfill all the high requirements of exact fitting, because the measured level also depends on the tolerances of the oil pan's fixture points . No part of the sensor should get into the critical range of moving parts inside or outside of the engine.

The mounting of the sensor has to be compatible with automation. Two assembly methods are validated: automated screwing and a pick-place and clamp-method. Design and development of these interfaces were realized according to simultaneous engineering principles.

ELECTRICAL INTERFACE

The electronic engineers have 3 major tasks to achieve: to reach the specified accuracy, to match the sensor's output with the calculating unit's input and to fulfill electromagnetic compatibility (EMC). The first two tasks are normally easy to do. To take precautions against influences like conducted distortions on the power and signal lines or electromagnetic radiation is a difficult and very complex task, especially when a capacitance measuring sensor is involved. Laboratory testing on samples is the first step to prove whether the sensor design meets the requirements. But this does not guarantee, that in the car the sensor and all other components will work well, when radiated with high electromagnetic fields. Due to special EMC-design electromagnetic interference testing was successful up to fields of 200 V/m.

TEST RESULTS

Temperature measurement with short response times and transmission of the data to the engine management unit supports functions for reducing emissions/exhaust fumes, reduced fuel consumption or fine tuning of additional mechanical functions.

The oil level in a passenger car engine is a function of engine speed, various vehicle movements and diverse oil properties. Figure 6 shows the oil level in a V6-engine as a function of engine speed. Figure 7 shows the influence of car steering.

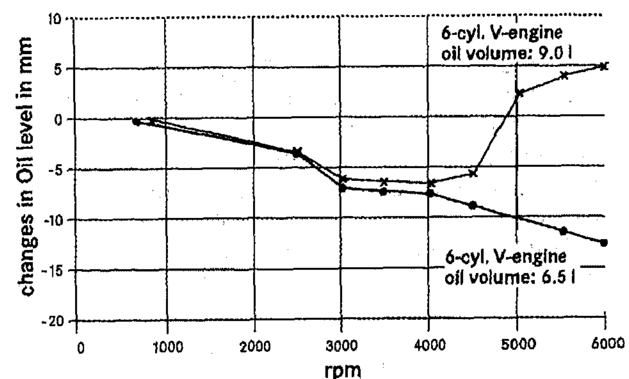


Figure 6: Oil level vs. revolutions per minute

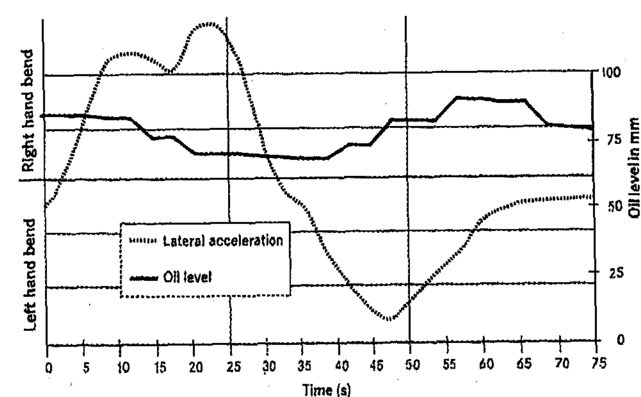


Figure 7: Oil level alteration when curving

Especially developed Algorithms can eliminate such influences. For example forming averages over a time interval or over a driving distance can eliminate influences arriving from steering or acceleration (Figure 8). Systematic effects (like engine speed) can be compensated and for extreme situations (like foaming oil, high acceleration or a high steering angle) values can be discarded. Thus the oil volume can be calculated with an accuracy of ± 1,6% taking all this measures into account.

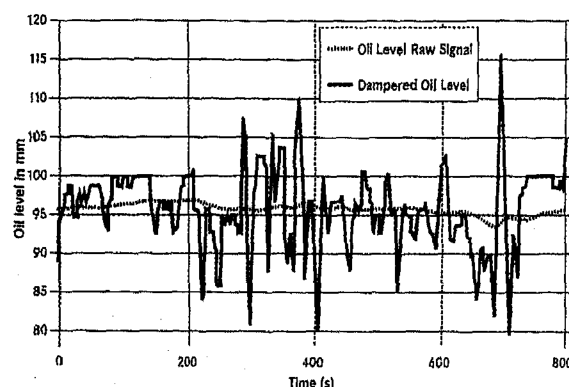


Figure 8: Data from sensor, mean value and oil level

Figure 9 shows the change of the dielectric number for several oils of different initial qualities and different driving conditions. The change depends on the quality of the oil and on the special driving conditions. It is not possible to determine oil quality only by evaluation of the dielectric number. In combination with further engine parameters customized oil drain periods can be achieved in a much more precise way.

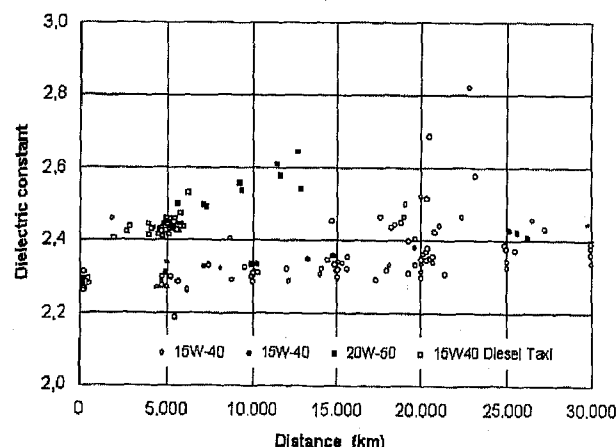


Figure 9: Change of dielectric number for several oils

It is important to know, that the oil temperature has great effects on the dielectric number. One possibility to overcome these difficulties is the evaluation of the dielectric number, measured always at the same temperature, e.g. at 80°. There are situations which lead to an increase of the dielectric number, e.g. short distance driving. The intrusion of water and fuel due to blow-by in the engine can be simulated with the so called hot and cold idling test shown in Figure 10 [3].

Fuel and water in the engine oil results in an increase of volume and therefore a higher level-signal. Driving a longer period under normal engine operating conditions with oil temperature exceeding 90°C evaporates water and fuel, so that the original level and dielectric number is reached (Figure 10). It is evident, that such situations can be recognized. Evaluation of the sensor data and other operation parameters enables to find the true interpretation of the state of the engine oil.

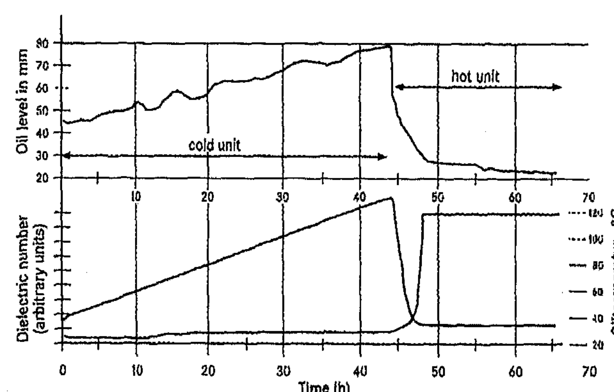


Figure 10: Hot and cold idling test

CONCLUSION

The QLT-sensor is the first system that can measure engine oil parameters on-line. Apart from oil level and temperature oil quality via detection of the dielectric number is achieved for the first time by a direct measuring instrument. Oil consumption and oil drain time can be calculated with sufficient precision contributing in saving natural resources due to less fresh and used oil. An important parameter for precise managing of other functions within the engine is oil temperature. As a result fuel consumption and exhaust emissions can be optimized.

REFERENCES

1. R. Thom, K. Kollmann, W. Warneke, M. Frend, „Extended Oil Drain Intervals: Conservation of Resources or Reduction of Engine Life". SAE Technical Paper Series No. 951035, 1995

2. G. S. Saloka, A. H. Meitzler, „A Capacitive Oil Deterioration Sensor". SAE Technical Paper Series No. 910497, 1991

3. T. Beyer, „Abstimmung eines Algorithmus zur Erkennung der Ölnachfüllmenge während der Fahrt für PKW-Motoren“ (Optimization of an algorithm for Determining on-vehicle Oil-top-up´s), Diplomarbeit HTW Dresden, 1996

APPENDIX

DERIVATION OF EXPRESSION FOR L

Capacitance between concentric cylinders:

$$C = \varepsilon_r \times \frac{\varepsilon_0 \times l}{\ln\left(\frac{r_o}{r_i}\right)} = \varepsilon_r \times Const. \quad (1)$$

with

l = length of the capacitor

r_o = radius of the outer cylinder

r_i = radius of the inner cylinder

ε_0 = dielectric constant = 8,85418782 x 10^{-12} As $V^{-1}m^{-1}$

ε_r = dielectric number of the material (here of oil)

For the sensor (assuming that capacitances in air and in vacuum are identical):

$$\varepsilon_r = \frac{C_q(oil)}{C_q(air)} = \frac{C_q(oil)}{Const._1} \qquad (2)$$

with

$C_q(oil)$ = capacitance of the reference capacitor immersed in oil

$C_q(air)$ = $Const._1$ = capacitance of the reference capacitor in air

$$C_l = Const._2 \times \frac{L_{max} - L}{L_{max}} + \varepsilon_r \times Const._2 \times \frac{L}{L_{max}} \quad (3)$$

with

C_l = Capacitance of the level-capacitor with that portion of L_{max} which is in air + that portion of L_{max} which is in oil

$Const._2$ = Capacity of the level-capacitor in air

Then

$$C_l = Const._2 + (\varepsilon_r - 1) \times Const._2 \times \frac{L}{L_{max}} \qquad (4)$$

or

$$L = \frac{L_{max} \times (C_l - Const._2)}{(\varepsilon_r - 1) \times Const._2} \qquad (5)$$

970852

Compact Magnetic Solenoid Valves Using a Composite Magnetic Material

Yoshihiro Tanimura, Keizo Takeuchi, Toshiaki Terada, Shinya Sugiura, and Yoshitada Katayama
Denso Corp.

Tsutomu Inui
Hitachi Metals, Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The recent progress of electronic control systems in vehicles is remarkable as evidenced by the development of electronic fuel injection systems,(EFI), automatic transmission control systems, and anti-lock brake systems,(ABS). The number of actuators for the systems has been increasing.

Consequently, a need has been identified for a reduction in volume and number of the system actuators for control purposes. A composite magnetic material has been developed with the aim of miniaturizing magnetic solenoid valves for actuator applications. A composite magnetic material is such that both ferromagnetic and paramagnetic sections coexist within a single material, and can contribute to optimization of the magnetic circuit of a solenoid valve.

This paper describes the development of a composite magnetic material, and its resultant characteristics. The chemical composition of the material, essentially 18% Chromium - 8% Nickel stainless steel, required to improve both ferromagnetic and paramagnetic properties and the process of producing composite magnetic properties are reported in detail.

This composite magnetic material exhibited superior properties, having a magnetic flux density B40 (Induction at H of 4000A/m) of 0.8 Tesla in the ferromagnetic section, and relative permeability μ r of 1.01 in the paramagnetic section. The application of this material to a magnetic solenoid valve for ABS could reduce by approximately 20% volume.

INTRODUCTION

Magnetic solenoid valves shown in Fig.1 are usually constructed from coil, moving core, stator core, and valve components. The function of the valve is to control the flow of a liquid or gas by moving a core using a magnetic force. In order to seal the liquid or gas, a metal component called a sleeve is used. This component is generally paramagnetic to maximize magnetic flux transfer from the moving core to the stator core.

If this component is constructed from a ferromagnetic section and a paramagnetic section as shown in Fig.2, magnetic efficiency can be improved, and the volume of the valve can be reduced. This may be accomplished by joining ferromagnetic and paramagnetic sections by welding; however, this method is costly, and presents difficulties in terms of sealing.

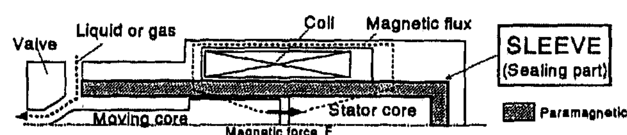


Figure 1. Typical magnetic solenoid valve

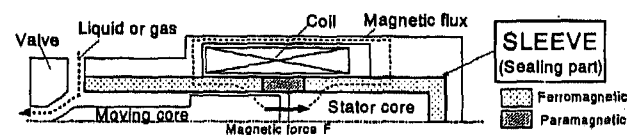


Figure 2. Efficient magnetic solenoid valve

A material has been investigated which may be either ferromagnetic or paramagnetic through the use of simple processing, and which may have both ferromagnetic and paramagnetic sections in coexistence. This material has been designated a composite magnetic material.

CONCEPT OF THE COMPOSITE MAGNETIC MATERIAL

It has been well known that austenitic stainless steels (e.g. AISI type 304) are composed of austenite after solid solution treatment, and that these materials exhibit paramagnetic properties. Furthermore, the austenite phase may be transformed to martensite through cold working, so called deformation-induced martensite, and the resultant martensite phase is ferromagnetic. Therefore, when a part of this martensite is made transform to austenite by heat treatment, a composite magnetic material is obtained as shown in Fig.3.

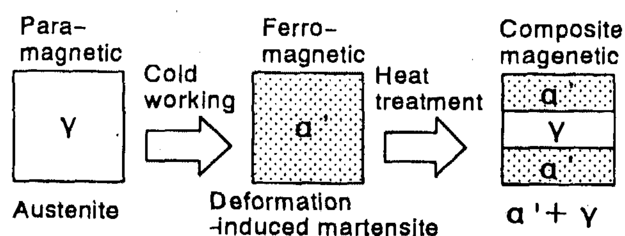


Figure 3. Schematic production process of composite magnetic material

Unfortunately, the amount of deformation-induced martensite formed upon cold working is small in typical austenitic stainless steels, such as AISI type 304, and these steels have insufficient ferromagnetic properties, a low magnetic flux density and a low maximum permeability. The development of a composite magnetic material was based upon the optimization of 18Cr 8Ni austenitic stainless steel with improved magnetic properties.

EFFECT OF CHEMICAL COMPOSITION ON MAGNETIC PROPERTIES

The effect of chemical composition on magnetic properties was investigated using various 18-8 stainless steels as detailed in Table 1.

Magnetic flux density, B40 (Induction at H=4000 A/m), for test samples after 70% cold reduction is shown in Fig.4 as a function of (Ni). Relative permeability after solid solution treatment is shown as a function of (Ni) in Fig.5. (Ni) is Hirayama's Ni equivalent [1,2] and defined as follows:

$$(Ni) = [Ni] + 0.65[Cr] + 0.98[Mo] + 1.05[Mn] + 0.35[Si] + 12.6[C] \qquad (1)$$

where [Ni], [Cr], etc. are elemental concentrations in wt% of the steel of interest. It has been known that (Ni) has a close relation to the volume fraction of deformation-induced martensite or athermal martensite.

The magnetic flux density following cold rolling increases with the reduction of (Ni) as seen in Fig.4. Additionally, the relative permeability following annealing decreases with increasing (Ni) as seen in Fig.5. Using these two factors, a chemical composition which optimizes magnetic properties may be arrived at. Magnetic flux density indicates that (Ni) should be held below 21.5 wt%. Relative permeability data indicates that optimum performance is obtained with (Ni) above 20.5 wt%.

Md AND Ms TEMPERATURES

The application of the composite magnetic material to automotive magnetic solenoid valve

Table 1. Chemical compositions of steels examined (wt%)

No.	C	Si	Mn	Ni	Cr	N	(Ni)
1	0.012	0.30	0.52	8.32	17.45	0.0018	20.5
2	0.023	0.30	0.50	8.39	17.58	0.0025	20.7
3	0.050	0.29	0.51	8.36	17.70	0.0033	21.1
4	0.030	0.10	0.50	8.34	17.84	0.0028	20.9
5	0.029	0.46	0.48	8.41	17.70	0.0042	21.0
6	0.031	0.86	0.51	8.33	17.91	0.0034	21.2
7	0.030	0.32	0.12	8.76	17.68	0.0011	20.9
8	0.031	0.31	0.79	8.38	17.72	0.0031	21.2
9	0.031	0.30	0.51	8.01	17.68	0.0028	20.5
10	0.030	0.29	0.51	8.17	17.90	0.0044	20.8
11	0.029	0.30	0.50	8.67	17.99	0.0036	21.4
12	0.030	0.28	0.50	8.37	17.23	0.0032	20.6
13	0.031	0.29	0.50	8.40	17.40	0.0022	20.7
14	0.031	0.29	0.52	8.37	18.32	0.0037	21.3
15	0.032	0.29	0.51	8.32	17.98	0.0120	21.1
16	0.032	0.29	0.50	8.35	17.98	0.0190	21.1

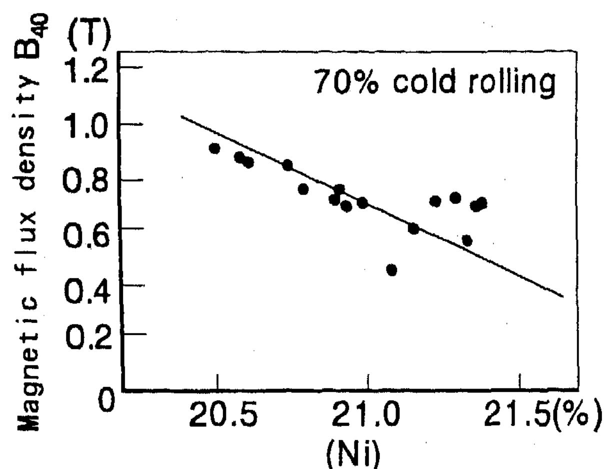


Figure 4. Variation of magnetic flux density as a function of Hirayama's Ni equivalent (Ni)

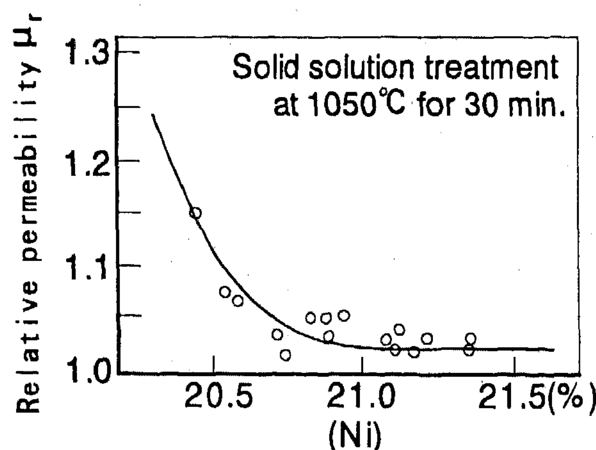


Figure 5. Variation of relative permeability μ r as a function of Hirayama's equivalent (Ni)

actuator sleeves further requires an investigation of the Md and Ms temperatures of the alloy. This is necessitated by the relationship of the Md temperature to ferromagnetism, and the Ms temperature to paramagnetism. Chemical compositions of the steels which were examined are given in Table 2.

Md TEMPERATURE - The Md temperature is defined as the temperature above which austenite will not transform to deformation-induced martensite upon mechanical working[3]. The current application dictates that the material temperature during cold deformation should be kept below the Md temperature so as to enhance the ferromagnetic properties of the material, thus promoting magnetic flux density and maximum permeability.

The martensite content variation as a function of temperature at 30% true strain is illustrated in Fig.6. Deformation-induced martensite fraction decreases rapidly with increasing temperature. The Md temperature can be estimated from Fig.6 by observing the temperature at which the volume fraction of martensite approached zero.

The variation of Md temperature with (Ni) is shown in Fig.7. As can be seen, the Md temperature is approximately 80℃, and is a very weak function to (Ni). This data indicates that working temperature is a critical property for ferromagnetic property development, and is a critical parameter for process control.

Ms TEMPERATURE - Paramagnetic austenite transforms to athermal martensite below the Ms temperature, and the magnetic property becomes ferromagnetic. The Ms temperature of the composite magnetic material should be lower than the expected operating range of the component to inhibit this transformation. If not, the relative permeability of the paramagnetic section would increase, and the advantages gained with respect to magnetic flux transfer between the moving core and the stator core would be lost.

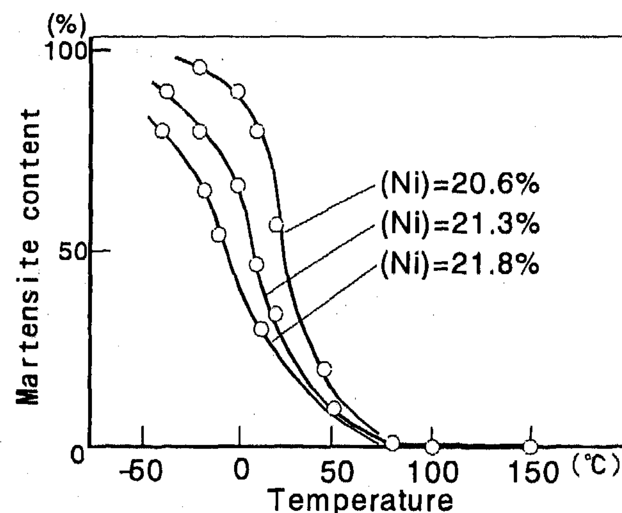


Figure 6. Variation of martensite content after 30% true strain as a function of temperature

Table 2. Chemical compositions of steels examined (wt%)

C	Si	Mn	Ni	Cr	N	(Ni)
0.028	0.17	0.43	8.20	17.75	0.0017	20.6
0.038	0.23	0.52	8.45	18.11	0.0015	21.3
0.038	0.28	0.55	8.75	18.19	0.0017	21.8

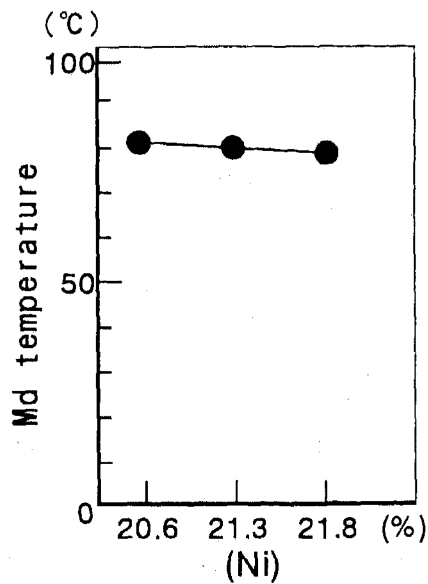


Figure 7. Variation of Md temperature as a function of Hirayama's Ni equivalent (Ni)

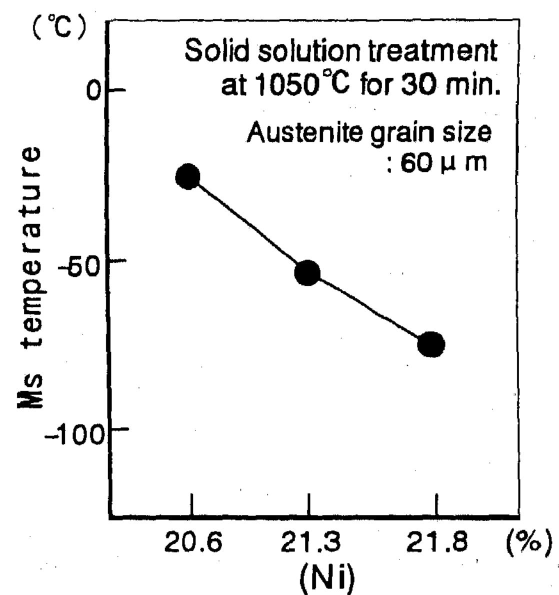


Figure 8. Variation of Ms temperature as a function of Hirayama's equivalent (Ni)

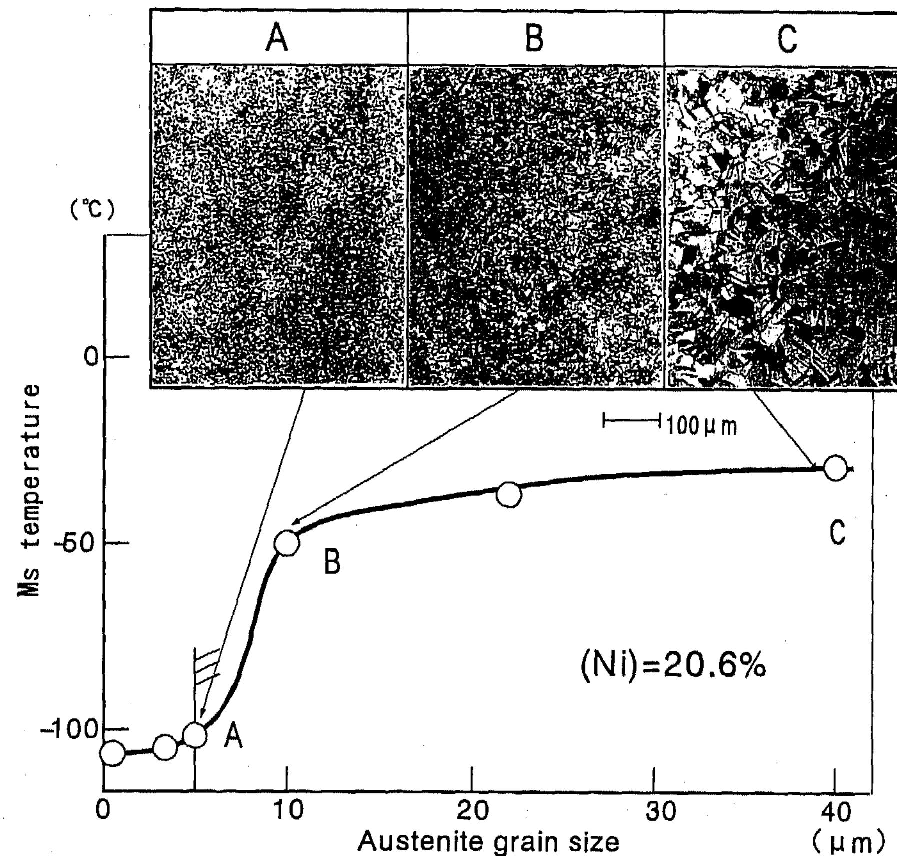


Figure 9. Effect of austenite grain size on Ms temperature

The variation of the Ms temperature with (Ni) for the materials in Table 2 is illustrated in Fig.8. The Ms temperature was measured with the differential scanning calorimeter. Ms temperature is found to decrease as the (Ni) is increased. At a (Ni) of 20.6%, the Ms temperature is found to be -28℃. As the actuator operating temperature may be below -28℃, the Ms temperature must be decreased further than is possible from strictly chemistry control.

There have been numerous studies of the Ms temperature in various steels throughout the literature[4-6], mainly conducted in Fe-Ni and Fe-Ni-C systems. Several studies have been conducted using Fe-Ni-Cr alloys, such as AISI type 304. Previous studies of Fe-Ni and Fe-Ni-C systems had concluded that the austenitic grain size had an effect on the Ms temperature of these steels [7], but that this grain size effect was small in comparison when Fe-Ni-Cr system was considered.

Based on these results, the effect of grain size on the Ms temperature of Fe-Ni-Cr steels was investigated further. The steel considered for this portion of the evaluation had a (Ni) of 20.6%. Austenitic grain size was varied through control of the austenitizing temperature of the alloy.

Fig.9 illustrates the Ms temperature as a function of austenite grain size. The Ms temperature is found to rapidly decrease with decreasing austenitic grain size when the austenite grain size is below about 10 μ m. This effect is enhanced by controlling the austenitic grain size below 5 μ m, for maximum Ms reduction.

APPLICATION TO ABS MAGNETIC SOLENOID VALVES

The composite magnetic material discussed above has been suggested for application in an ABS magnetic solenoid valve. A schematic of this type of valve is given in Fig.10. The composite magnetic materials described in Table 2 were tested for performance in the sleeve component of the valve. A production process for the manufacture of the composite magnetic sleeve is illustrated in Fig.11.

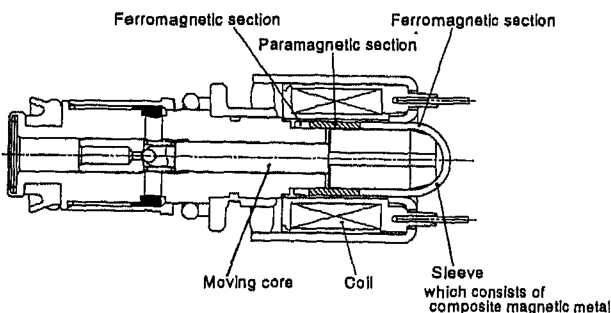


Figure 10. Schematic view of ABS magnetic solenoid valve

The sleeve was formed of the composite magnetic material, having an initial structure consisting of paramagnetic austenite. The working temperature was maintained below the 80℃ Md temperature established earlier. Mechanical working under these conditions induces a certain portion of the austenite to transform to ferromagnetic martensite as detailed previously. Subsequently, a section of the sleeve was induction heated into the austenite range to form the paramagnetic region of the sleeve. Induction heating parameters such as frequency, voltage, current, and time were adjusted to yield an austenitic grain size of less than 5 μ m, and thus obtain maximum Ms reduction.

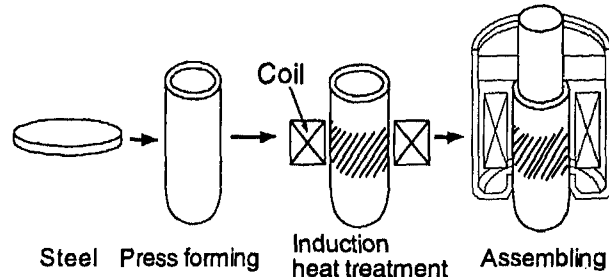


Figure 11. Schematic production process of sleeve of ABS magnetic solenoid valve

Table 3. Magnetic properties and Ms temperature of composite magnetic sleeves

(Ni)	B40(T)	μ r	Ms(℃)
20.6	0.38	1.007	-110
21.3	0.60	1.012	-105
21.8	0.80	1.020	-90
Desired value for sleeve of ABS valve	≧0.3	≦1.2	<-40

·B40 of ferromagnetic section Is measured after press forming

· μ r and Ms of paramagnetic section are measured after induction heat treatment

Table 4. Comparison of ABS magnetic solenoid valve performance

	Sleeve material	
	Composite magnetic material	AISI type 304
Outer diameter of ABS magnetic solenoid valve (mm)	ϕ 19	ϕ 24
Coil ampere turns (A·turns)	320	452
Magnetic force (N)	30	30

·Composite magnetic material has (Ni) of 21.3%

The resultant magnetic properties of the both ferromagnetic and paramagnetic sections of the composite magnetic material are given in Table 3. The resultant Ms temperature of the paramagnetic section is also given. The sleeve was found to have a

magnetic flux density, B40, of 0.38 - 0.80 T, a relative permeability, μ r, of 1.007 - 1.020, and an Ms temperature of -90 to -105℃. These properties satisfied required values.

A comparison of the ABS magnetic solenoid valve with a composite magnetic sleeve and with a sleeve of AISI type 304 is given in Table 4. Application of the composite magnetic material to the sleeve allows a 20% reduction in outer diameter as compared to the AISI type 304 sleeve. ABS valves using the composite magnetic material have been successfully used in field service since 1996.

CONCLUSION

A composite magnetic material having both ferromagnetic and paramagnetic sections has been developed from an austenitic stainless steel.

Results of this development can be summarized as follows:

(1) The coexistence of ferromagnetic and paramagnetic regions within the material may be accomplished by holding the Hirayama's Ni equivalent (Ni) within the range of 20.5 to 21.5%.
(2) Ms temperature of the alloy was substantially reduced through the control of the austenitic grain size.
(3) Application of the composite magnetic material to the sleeve of the ABS magnetic solenoid valve allowed reduction of the valve outer diameter of approximately 20%.
(4) Typical properties of the sleeve component were found to meet product expectations as follows: magnetic flux density, B40, 0.38 - 0.80 T, relative permeability, μ r, of 1.007 - 1.020, and Ms temperature of -90 to -105℃.

REFERENCES

[1] T. Hirayama and M. Odagiri, Journal of the Japan institute of metals, Vol. 34, No. 5, p. 507, 1970.
[2] T. Hirayama and M. Odagiri, Journal of the Japan institute of metals, Vol. 34, No. 5, p. 511, 1970.
[3] T. Hirayama and M. Odagiri, Journal of the Japan institute of metals, Vol. 34, No. 8, p. 826, 1970.
[4] V. Raghavan and A. R. Entwisle, Physical Properties Special Report No. 9, London, 1965.
[5] S. R. Pati and M. Cohen, Acta Met., Vol. 14 , P. 1001, 1966.
[6] A. R. Entwisle and J. Feeney, Institute of Metals Monogragh, No. 33, p. 156, 1968.
[7] M. Umemoto and W. S. Dwen, Metallurgical Transaction, Vol. 5, September, p. 2041, 1974.

970901

Complex Shape Headlamps: Eight Years of Experience

Eric Blusseau and Laurent Mottet
Valeo - Lighting Systems

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Until the late 1980s, the technology used in automotive lighting was relatively limited. Designers had little styling freedom as the main prerequisite of car makers was to reduce the size and increase the performance of lighting systems. VALEO invented SC (complex surface) technology to improve luminous efficiency. The first cars equipped with SC were the EAGLE PREMIER in the United States the CITROEN XM in Europe.

SC2 followed. SC2 incorporates reflector technology that controls both the vertical and horizontal deviation of the light and increases styling freedom especially for fog lamps.

The latest in the line of high performance lamps, SC3, was developed to optimize beam patterns in order to meet statutory requirements whatever bulb is used. Because SC3 uses a clear lens with an optical reflector, styling can be much more original and the angle of the lens can be increased to 60°. This technology is currently being made available to all VALEO customers worldwide in the form of SC3 clear lens high performance headlamps.

The photometrical performance of these new technologies requires higher precision and quality in the shape of the reflector. This is achieved by new processes developed with computer-assisted design and digitally controlled tools associated with dependable quality assurance facilities.

Since the beginning of the 1990s the higher performance and styling advantages of SC headlamps have resulted in their being used in more than 30% of new programs worldwide. VALEO's continuous growth is based on technological leadership in product innovation and manufacturing processes. The high performance of SC and HID lamps with clear lenses and automatic dynamic leveling systems form an integral part of VALEO's technological development.

INTRODUCTION

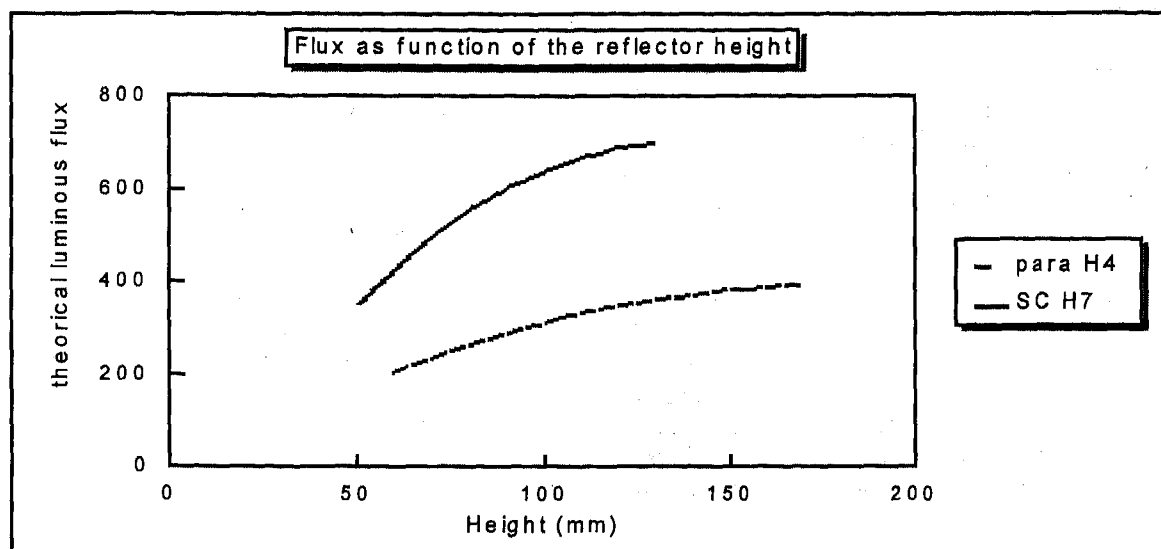
THE NECESSITY FOR COMPLEX SURFACES - Automobile lighting technology developed only slowly until the mid 1980s. Until that time the basic operating principle of headlamps had remained essentially the same. A parabolic reflector placed behind a filament directed the light ahead while a lens in front spread the light horizontally.

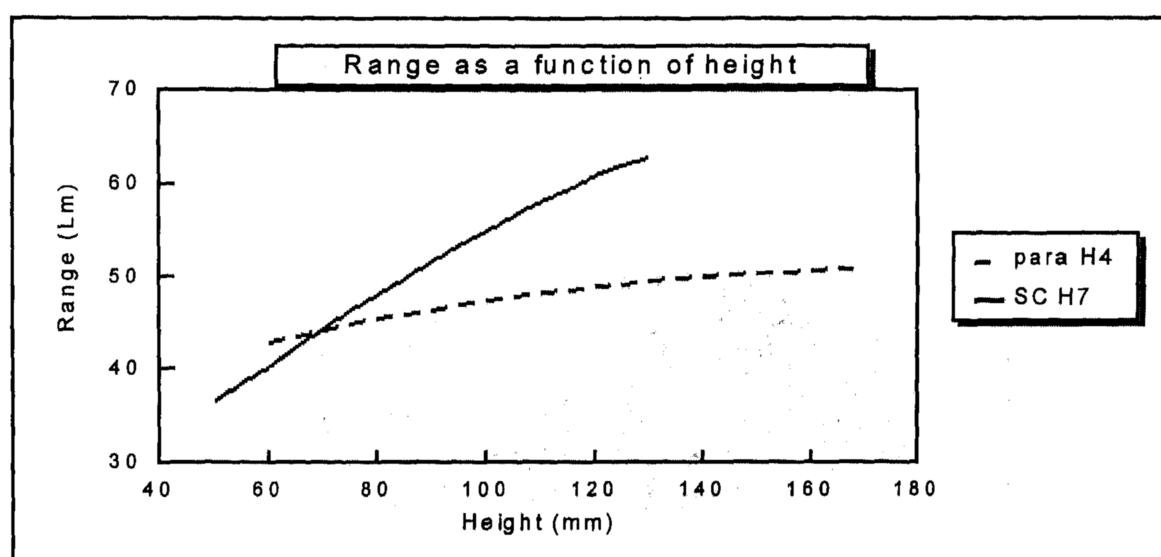
EUROPE - All over the world, regulations require vehicles to be equipped with a low beam. In Europe, in order to avoid glare, this low beam is characterized by an asymmetrical cutoff line which differentiates the illuminated area (below the HH line) from the darker area above.

European technology is based on the utilization of a light shield located below the light source. This light shield, associated with the filament, generates a cut-down beam. This technology has been proven on millions of vehicles. Nevertheless, automobile designers motivated by the need to improve aerodynamics demanded headlamps which were increasingly narrow. This powerful trend was incompatible with photometric performance which in addition was required to respond to the increased performance of the vehicles themselves. To meet these two contradictory demands, it was vital to find new solutions.



90's style: Low profiled headlamps





USA - US headlamps in 1985 utilized transversal filaments. This technique very rapidly reached its natural limits. The use of transverse light sources is not the best way to control luminous flux. Transverse light sources also require relatively high reflectors with a long focal distance if they are to provide satisfactory illumination. In the mid 1980s, axial filament bulbs appeared. These were more efficient in terms of illuminance, but their application to low beams required highly prismatic lenses which were difficult to produce. This handicap could be overcome by choosing reflectors with a long focal distance but at the cost of a significant drop in flux, which also limited styling possibilities (see SAE presentation 1987).

The demand for new solutions was therefore urgent both in the US and in Europe. Traditional technology could not meet that demand.

SC1: THE REVOLUTION

THE INVENTION OF COMPLEX SURFACES - The problem was clear, but the solution had yet to be found. As it happened, the identification of the problem coincided with the massive expansion of computer assisted design and optical simulation in photometric laboratories.

The idea was to "align the highest points of reflected images". The search for a mathematical definition of the surfaces took 2 years and resulted in an initial series of patents in 1982. The potential of this new concept of surface, which disassociated the horizontal and vertical focus lines from the reflecting surface while maintaining continuity, was considerable. The problem was solved in a manner which was both elegant and efficient. For a given dimension, the product is considered to be 50% more effective than a traditional headlamp. This means that for a given performance, the height of the headlamp can be divided by 2. Any combination of these two extremes is of course possible.

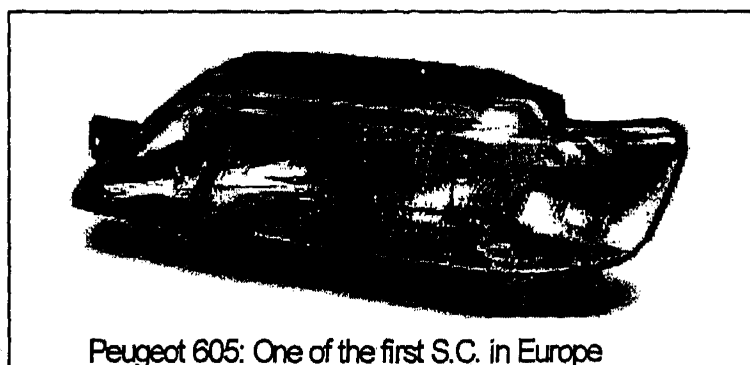
The immediate applications of this surface concept are immense.

- improvement of fog lights by simplifying distribution lenses,
- definition of a European low beam illuminance shape,
- definition of a US low beam illuminance shape with the possibility of optimizing the image position.

THEORY INTO PRACTICE - Once the concept had been discovered and developed, it had to be put into practice. A large number of mock-ups and prototypes were produced in order to improve control of light distribution. Lens trials were carried out. The impact of dispersion of filament positioning was measured.

The concept was presented to manufacturers all over the world. They immediately saw the importance of the contribution this technology could make and were able to see to what extent it fulfilled to their demands as regards both styling and performance.

The first orders were filled for the EAGLE PREMIER in the US, the CITROEN XM and PEUGEOT 605 in Europe.



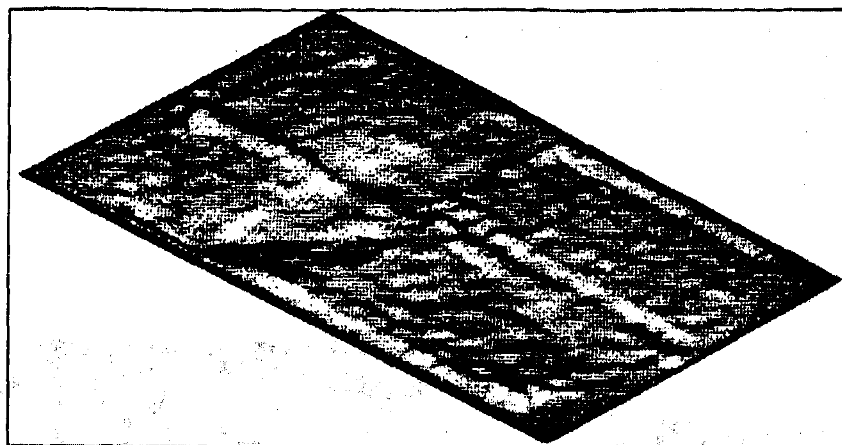
Peugeot 605: One of the first S.C. in Europe

INDUSTRIALIZATION

The next difficulty was the mass production of these headlamps, which were precision optical instruments. The utilization of CAD/CAM in the machining of these complex surfaces became common. Digital control was indispensable in the production of the machine tools.

Newer, more dimensionally stable materials with improved resistance to operational temperature requirements were developed for the reflectors.

Research was carried out particularly on thermo-hardened materials with excellent heat-resistance characteristics. In fact, the reduction in the size of headlamps resulted in an increase of some 50°C in the temperatures inside the headlamp. Composite materials with precisely adjusted expansion coefficients were perfected. The parameters of injection presses became increasingly sophisticated in order to be able to obtain optimum mold performances for reflector injection.



Visualization of the difference between the reflector and the theoretical shape

The use of industrial robots in varnishing processes also improved the surface treatment of the reflectors.

At the same time, new tools had to be developed to check the surfaces in order to be able to assure the optical quality of these products. Plungers had to be accurate to within 5 microns and the surface of the reflectors to within 1/100 mm.

THE LIMITS

The complex surface technique has enabled the efficiency of low beam lamps to be increased to a significant degree both in the US and in Europe. It has opened up new fields of possibilities in terms of styling particularly by allowing smaller headlamp reflectors while maintaining or even improving optical performance. Today around the world, 30% of the new automobile models presented by manufacturers are fitted with complex surface headlamps (i.e. non-parabolic surfaces). Nevertheless, with ever bolder and more original styling trends, it has proved necessary to advance even further. And to meet new demands.

The initial limits were revealed by shaped fog lights. In fact, the combination of light deviation ribbing with a lens inclined at more than 30° posed a problem. However, complex surfaces, confronted with this new challenge, would provide the solution.

SC2: ANOTHER STEP FORWARD

Complex surfaces didn't allow the production of a high performance beam with very inclined glass with optics. The overall beam suffered unacceptable deviations.

THE PRINCIPLE - The idea consisted in seeing that there were only two possible solutions to the dilemma that though inclined ribbing was unsatisfactory, inclined headlamps were required by the manufacturer,:

1) either incorporate a second lens, with the inner lens handling the distribution of the light, while the smooth outer lens ensures continuity of line and closure.

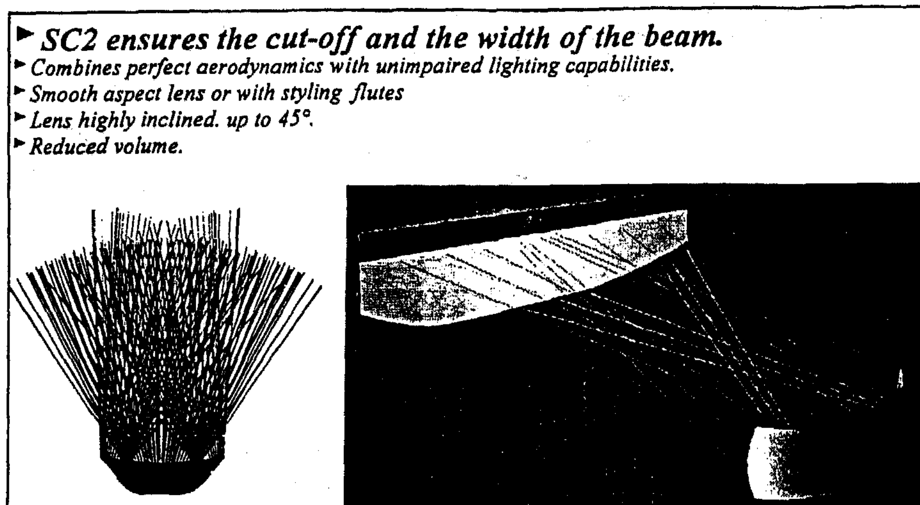
2) have the light distributed as much as possible by the reflector, with the external lens only slightly ribbed or completely smooth.

The first alternative was rejected by engineers because the double lens caused problems of temperature, volume, appearance and cost.

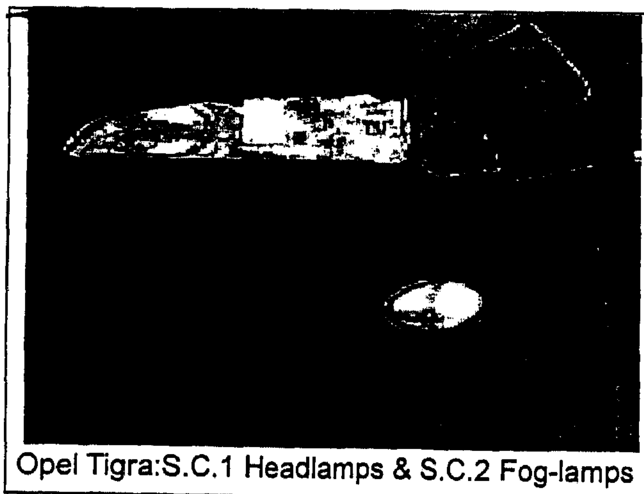
The second path was therefore the one that was explored.

It took more than two years of research to find the technical solution. The work concentrated on the optimization of the shape of the reflector in order to ensure the desired distribution of the light. After numerous trials, SC2 was presented in 1991, and promptly won the 1st prize of INNOVATION for automobile equipment in Paris 1991. This was not without reason, as the product's multiple functions were interesting from numerous points of view. It ensured the lateral distribution of light for the vehicle and took account of the external environment of the vehicle to preclude any unlit areas.

The design took account of the interior structure of the headlamp in order to avoid interference between the reflected light and the bulb or any headlamp cache. Thermal constraints were overcome by optimizing parallel light levels in one lens.



APPLICATIONS - This principle was widely employed in fog lights. The combination of an SC2 surface with a smooth lens enabled a range of variously shaped standard reflectors (round, rectangular) which could then be installed in different vehicle configurations. These products have been fitted to various European and American model ranges (Renault, Ford, Saab, Mercedes, G.M., etc.). Standardization of the mirrors and the simplicity of the lens have led to a reduction in R&D costs as well as in production costs.



Opel Tigra:S.C.1 Headlamps & S.C.2 Fog-lamps

SC3: EXTENDING THE CONCEPT

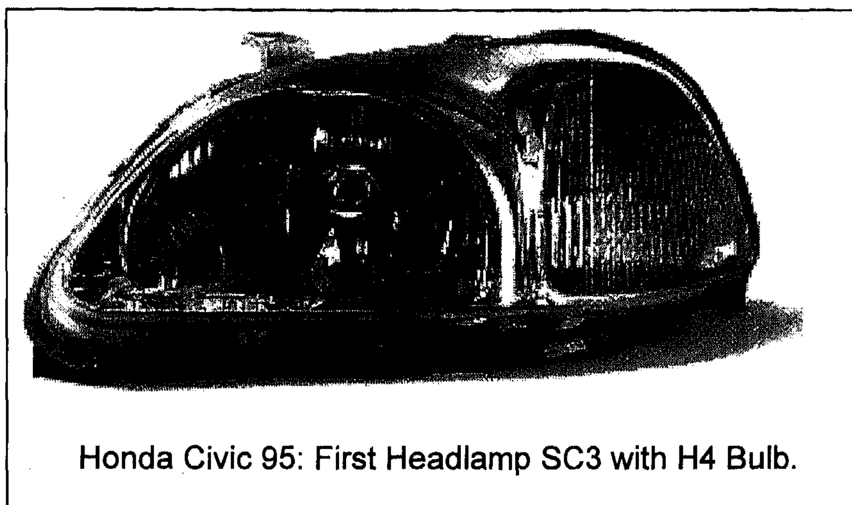
Styling demands were increasingly exacting. As regards main headlamps, the increasingly pressing need for originality expressed by the inclination of the lens prompted VALEO engineers to go even further in their research into optical surfaces.

The aim was to optimize portions of locally modified SC2 surfaces, sometimes by tilting them, or by adding optics to the standard surfaces.

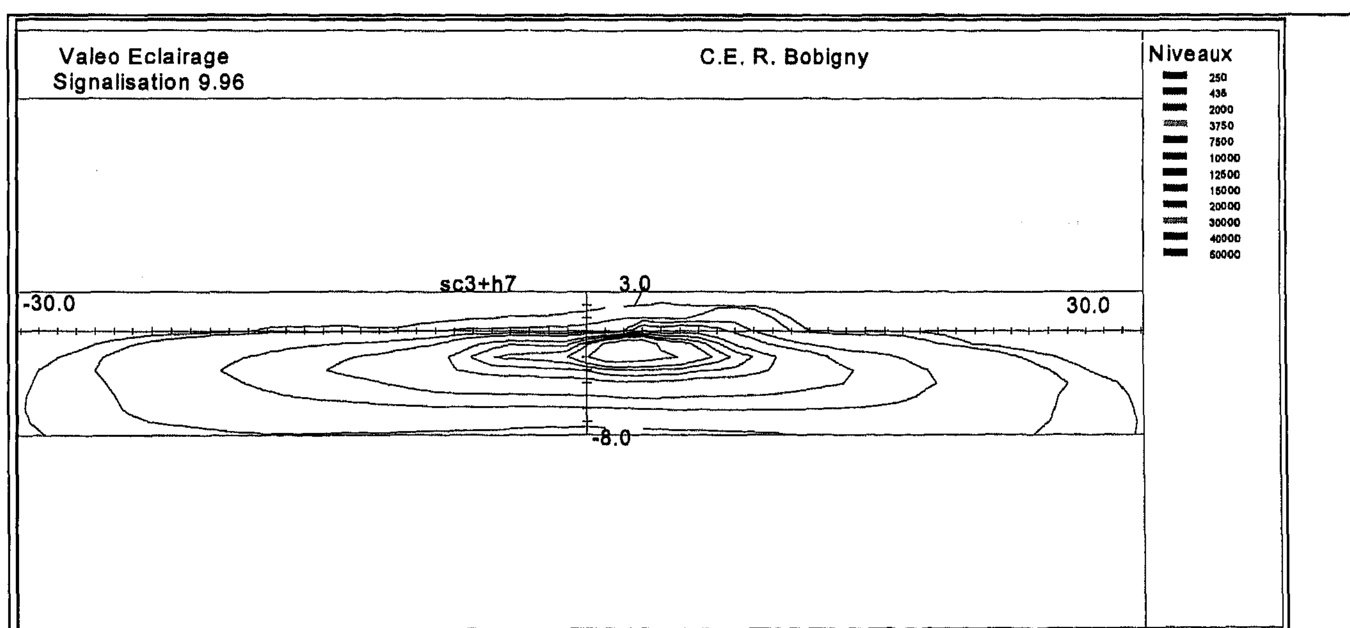
Research into optical surfaces was carried on for 2 more years. The complex surface concept was refined and adapted to all types of bulbs (9005/9006, 9007, HB2) and for European and US beams. In Europe, 6 different types of surface were defined.

This new concept of complex surfaces brought about breakthroughs in styling. In fact, the clear lens associated to a faceted reflecting surface gives a deep, shimmering (glistening) appearance to the headlamp. The shininess of the aluminum and the surface facets give the lamp a jewel-like quality. The optically neutral clear lens can be inclined at more than 60° and even be curved in without impairing the beam. Yet this freedom of styling is not achieved at the expense of the efficiency of the headlamp, and SC3 performance is at least equivalent to an SC1 surface of the same dimensions. This new generation of complex surfaces has required developments in industrial tools. The quality criteria for surface states have been pushed back and requirements in terms of the appearance of headlamps are increasingly draconian. Dust-free production lines have had to be created.

APPLICATIONS - SC3 headlamps were first released in Europe and Japan in 1995. In Europe, the Renault Safrane and in the U.S., the 1997 Saturn Coupé were fitted with this new technology, which is assured a promising future.

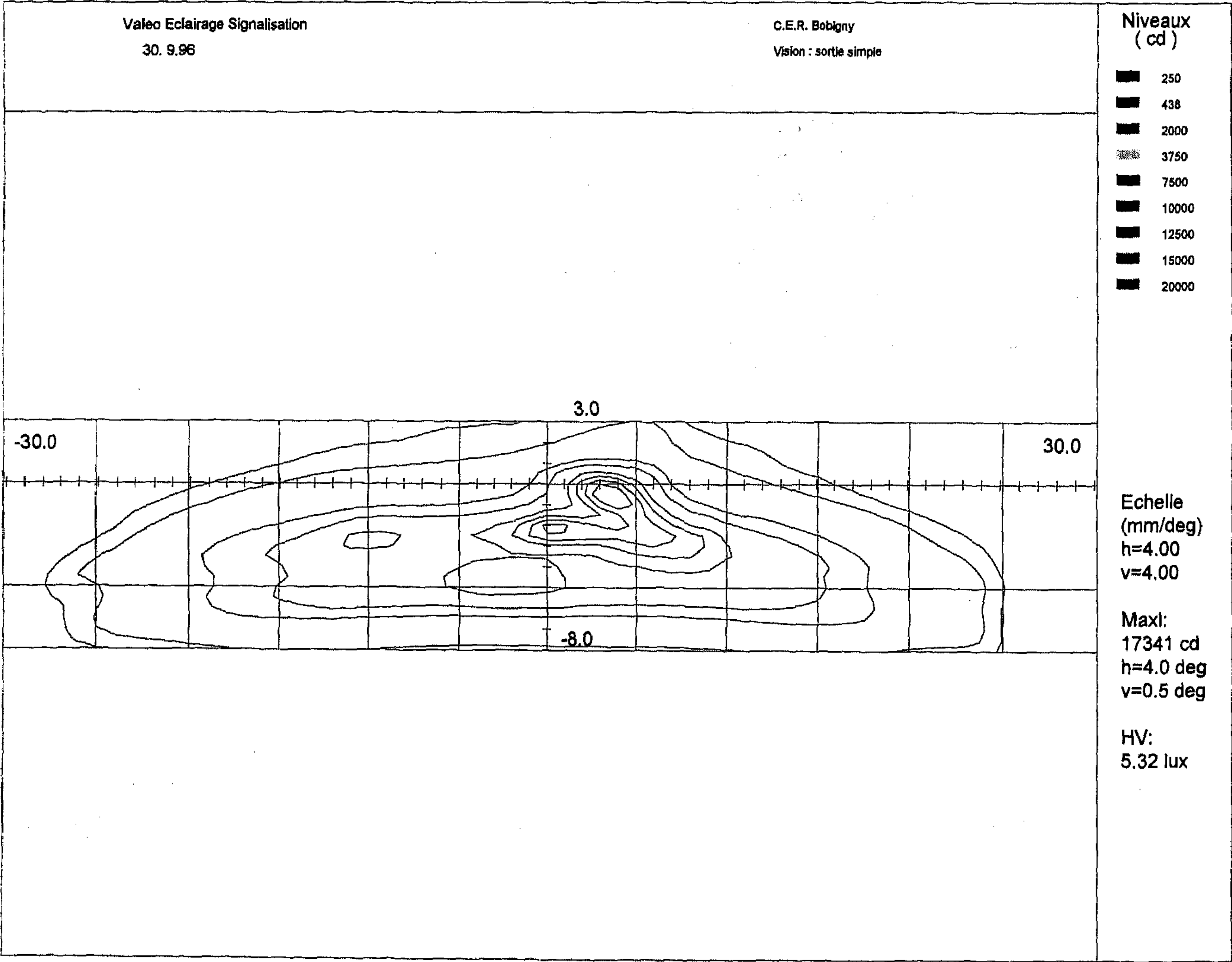


Honda Civic 95: First Headlamp SC3 with H4 Bulb.

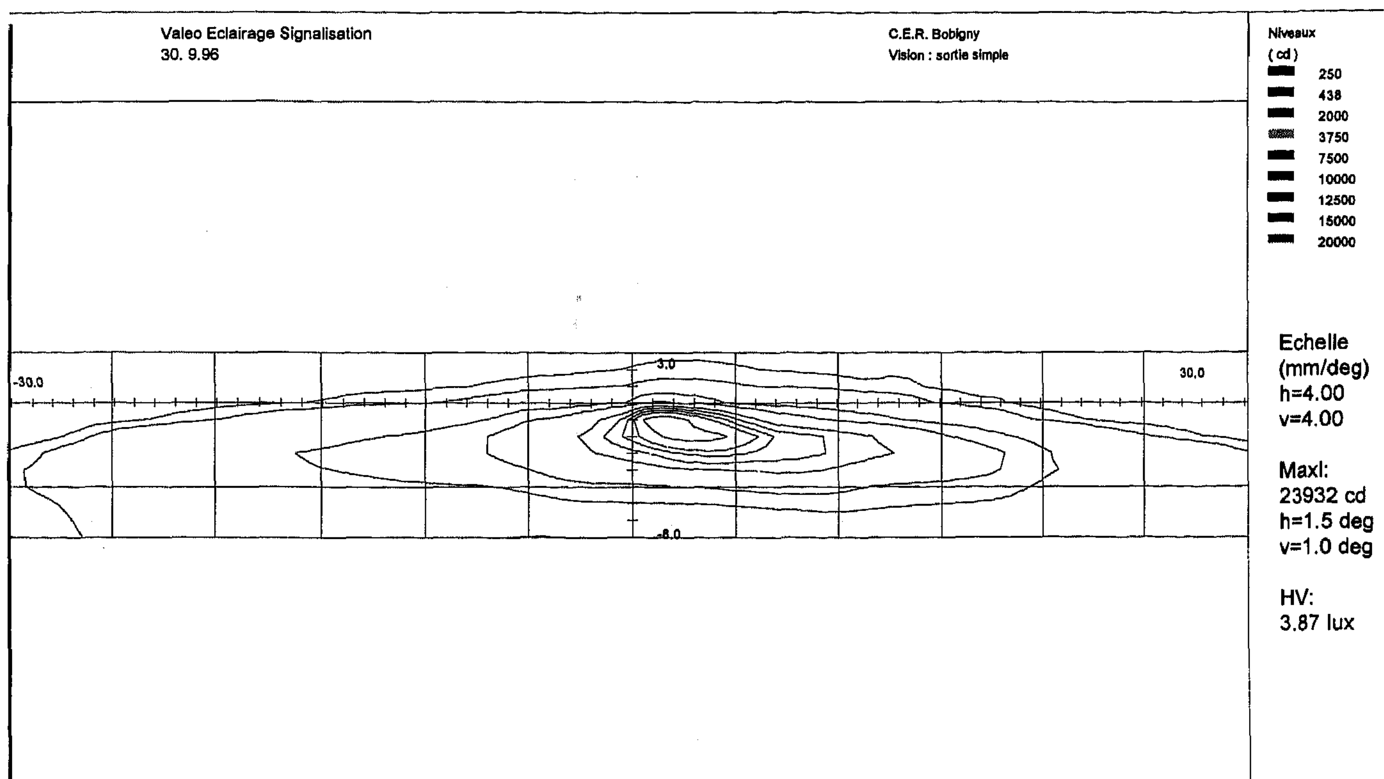


Isolux curve European beam pattern SC3

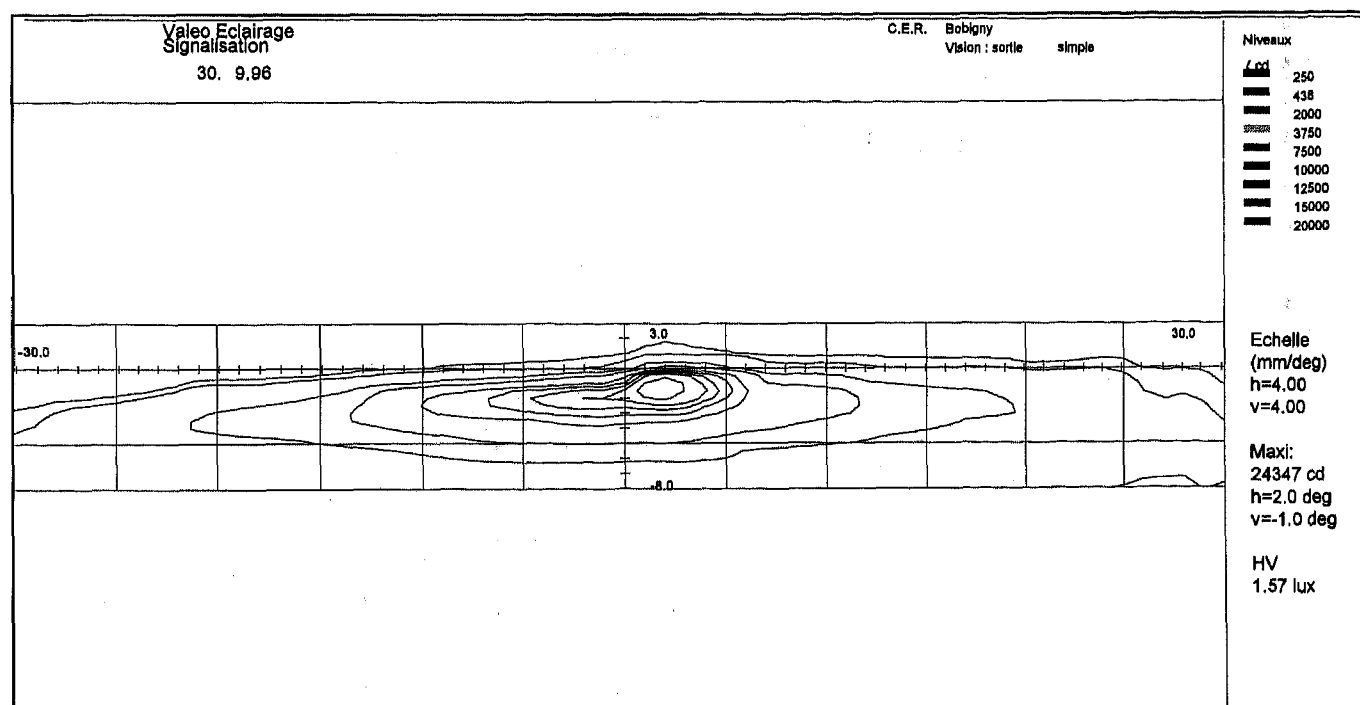
American Beam Pattern :



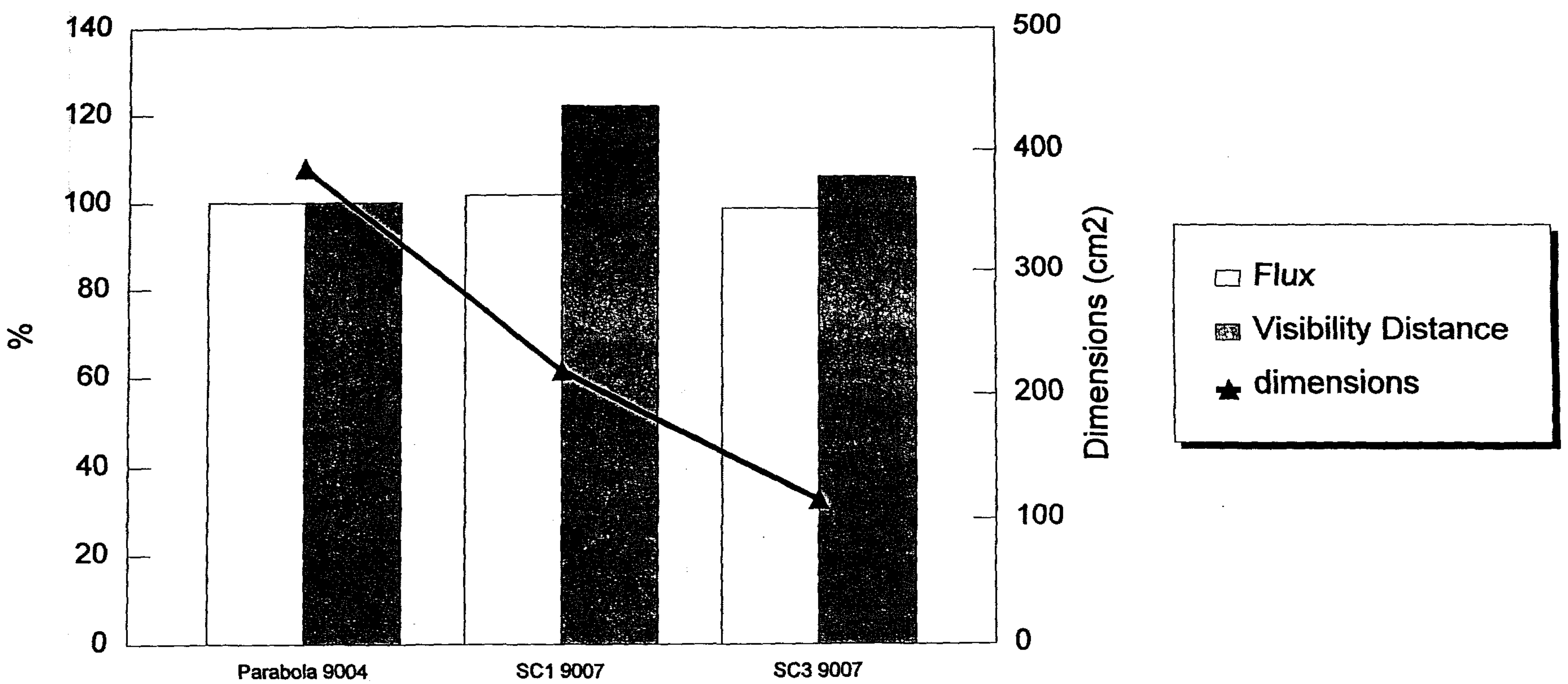
Isolux curves Parabola 9004



Isolux curves SC1 9007



Isolux curves SC3 9007


Evolution of low beam
performances/dimensions
%
140
120
100
80
60
40
20
0
500
400
300
200
100
0
Dimensions (cm2)
Parabola 9004
SC1 9007
SC3 9007
Flux
Visibility Distance
dimensions

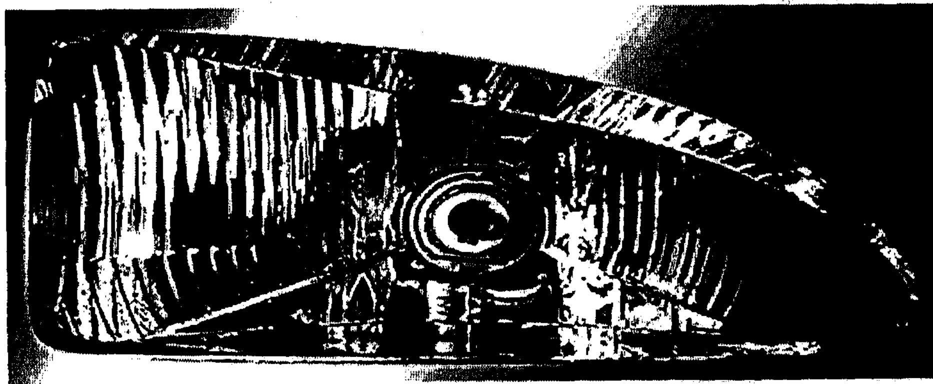
HOW ABOUT TOMORROW?

Complex surfaces have brought about a revolution in automobile lighting techniques. At this time, parabolic technologies still represent 70% of all new vehicles manufactured. By 2000 this proportion will have fallen to 50%. Of the remainder, SC1 lamps will account for 35% and SC3 lamps 22%.

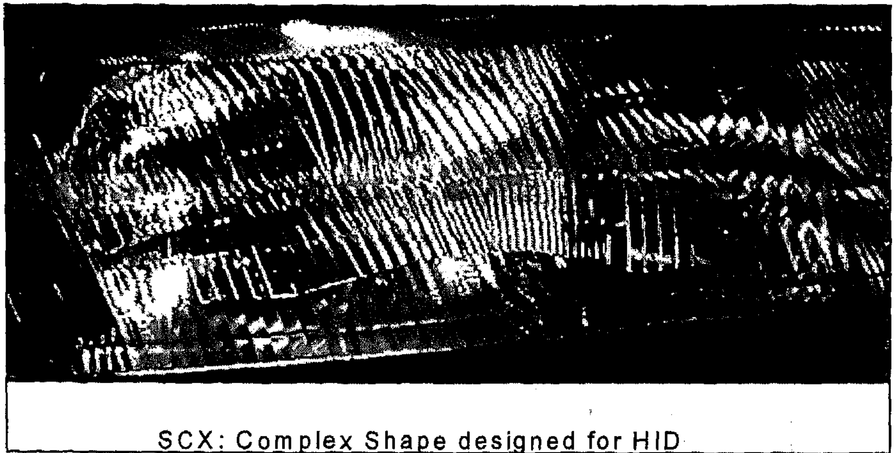
Experience and success prompted further research in this area. Development work is currently being concentrated in two main areas:

1) In order to anticipate and contribute to the ongoing global harmonization of photometric regulations, Valeo has created a new SC concept which will enable the principle of "design to cost" to be applied to our complex surfaces worldwide. For example, it is already possible to use the same tool cavity for the U.S., Europe, and Japan. Only the plunger is still specific depending on the sense of the traffic (left or right-hand drive). Over 50% of tooling costs can thus be saved.

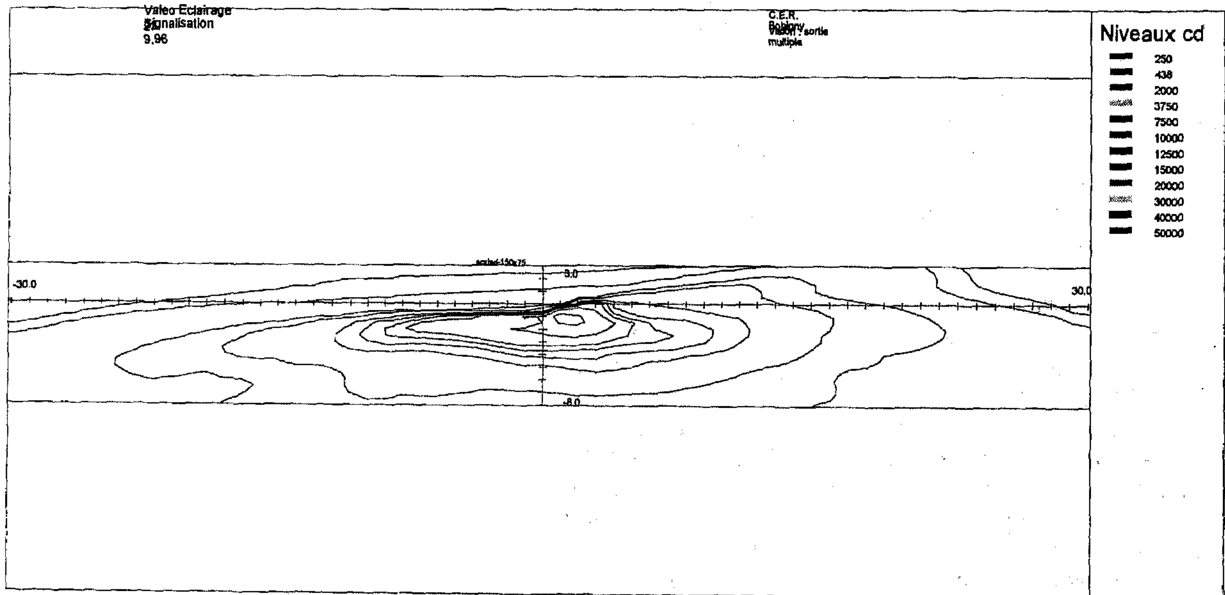
2) The emergence of H.I.D. represents a new opportunity for the SC concept to demonstrate its potential for development. This super-powerful lamp, with indistinct lighting contours, makes the computation of optical surfaces much more difficult. Even so, the complex surface concept has been adapted to meet the challenge. SCX (lens with optics) and SC3X (smooth lens), are complex surface products undergoing validation and are already being proposed to manufacturers around the world for their new headlamps. Yet another step in the reduction of dimensions, with a consequent improvement in performance that will be of benefit to all.



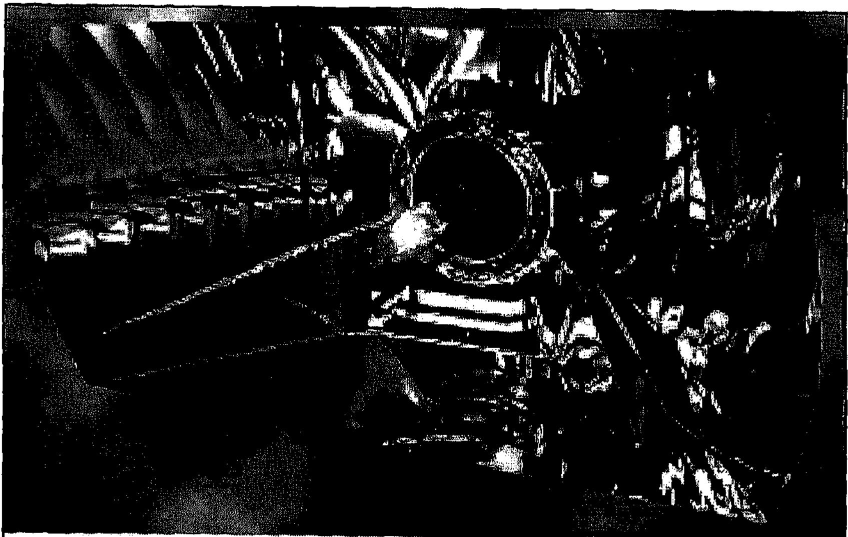
New S.C.3 Worldwide concept



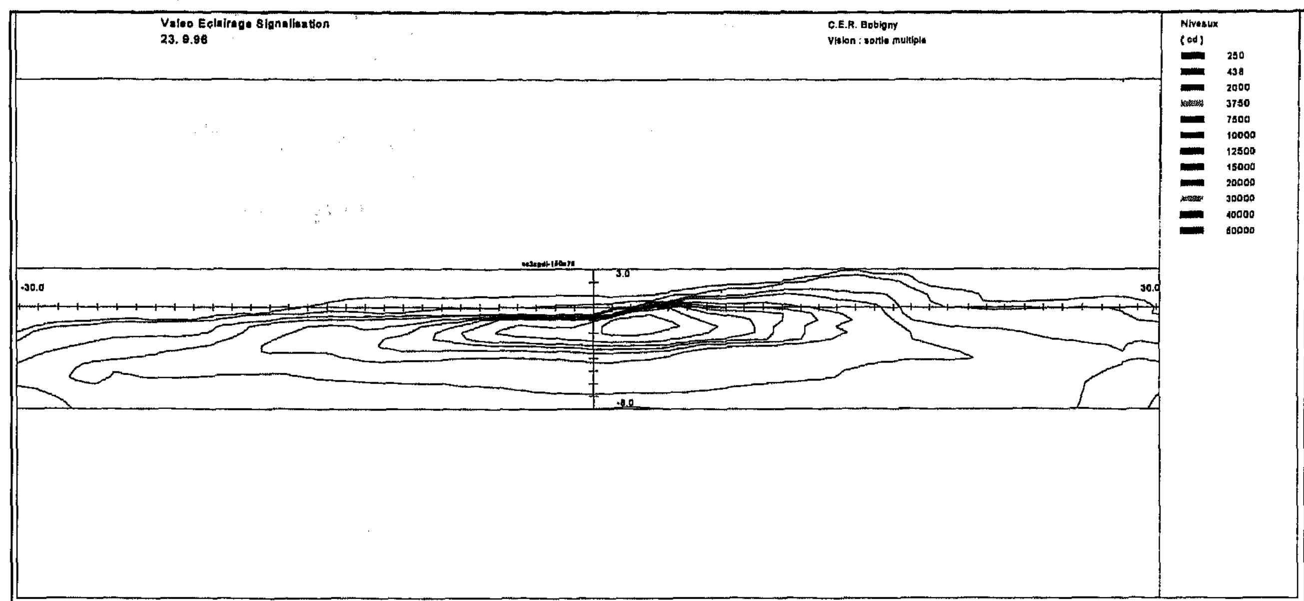
SCX: Complex Shape designed for HID



Isolux curves SCX with HID light source.



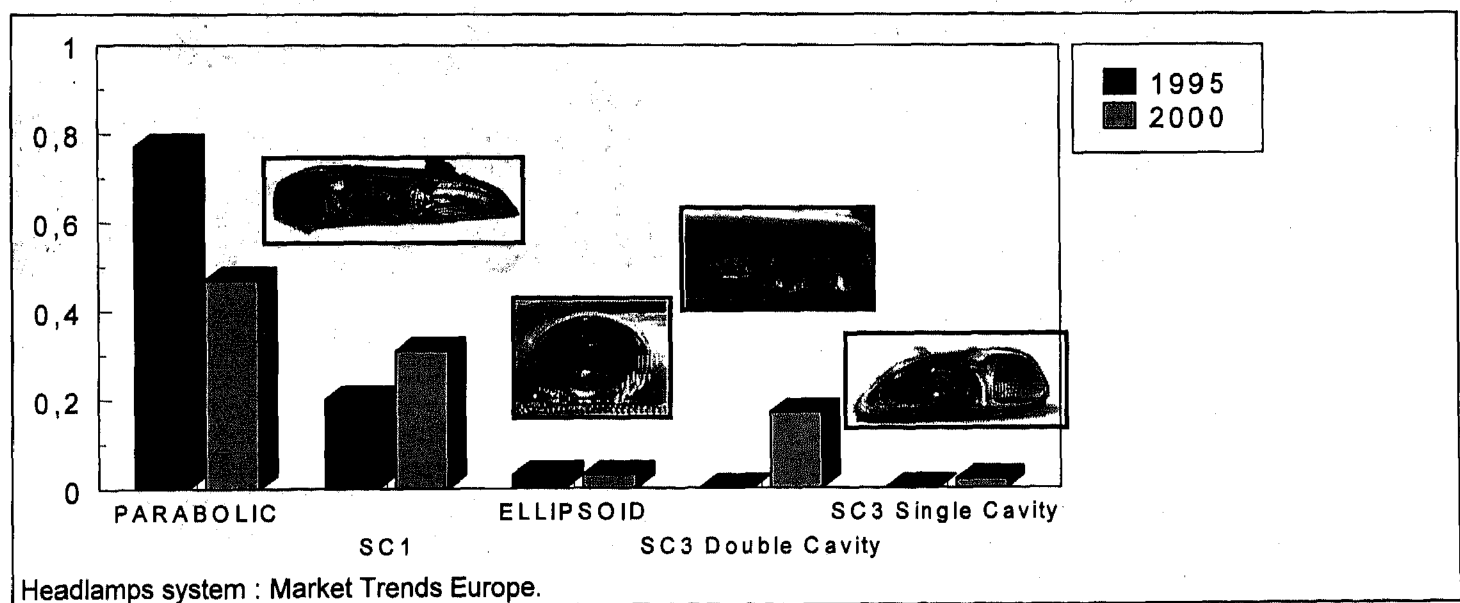
SC3X: Clear lens complex shape for GDLS.



Isolux curves SC3X with HID light source

CONCLUSION

Today's situation as well as future trends both in Europe and in the U.S.A confirm the success of the concept of complex surface, whatever their evolution. They fulfill the designers aspirations while optimizing the lighting efficacy and downsizing, without sacrificing the cost. In a rare example, progress in technology made it possible to make happen in the real world the most daring styling demands.



Headlamps system : Market Trends Europe.

970902

Design Strategy for Free-Form Reflector Head Lamp Using High Intensity Discharge Light Source

Ben Wang, Douglas Kreysar, and Jianzhong Jiao
North American Lighting Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A design strategy for implementing High Intensity Discharge (HID) light sources in low beam head lamps with Free-Form Reflectors (FFRs) is presented. A theoretical analysis of a light source image formed by each portion of a reflector is presented as well as a strategy for using those images to create a head lamp beam pattern. Simulation results are shown for a head lamp design using this strategy.

INTRODUCTION

Styling and performance requirements have directed the automotive industry towards head lamps with clear lenses and free-form reflectors (FFRs)[1,2]. Further increases in head lamp performance demands have led to much interest in high intensity discharge (HID) light sources[3,4]. In order to marry the two technologies into one product, several technical problems must be overcome. Two of the most difficult issues in designing a FFR low beam headlamp with an HID source are A) the control of glare light and B) the control of excess amounts of foreground light. These difficulties stem from the extended discharge area of the HID source (4.4 mm x 4.4 mm x 7.1 mm ellipse, see Fig. 1) compared to a typical tungsten halogen (TH) bulb (5.2 mm x 1.2 mm cylinder). The extended discharge area translates into a large effective light source size, which causes a large angular spread of light from a point on the reflector surface. This large angular spread can lead to excess amounts of glare light or foreground light.

The angular spread of light (i.e., image size) from a point on the reflector depends on the size of the light source and its distance to the point on reflector. In order to decrease the image size produced from a point on the reflector, the distance between the light source and the point on the reflector should be increased. This paper presents a theoretical analysis of the image size formed by a free-form reflector and a method of categorizing these images whereby the reflector is divided into "zones" based on the image size formed by different areas of the reflector. Furthermore, this paper discuss how the difficulties of the HID light source are overcome by using different zones of the reflector for different areas of the beam pattern.

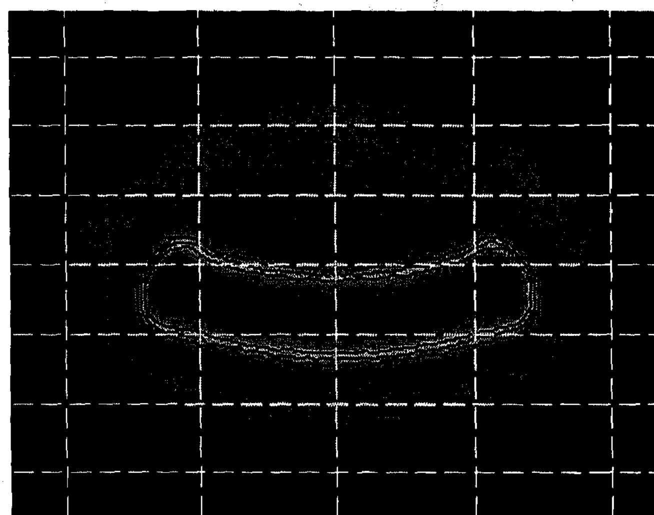


Fig. 1 Image of D2S HID Light Source

THEORETICAL ANALYSIS OF THE IMAGE SIZE FORMED BY DIFFERENT PORTIONS OF A FREE-FORM REFLECTOR

A reflector can be considered an imaging device. If only specular reflector is assumed, images of a light source

(filament or arc) are formed by each point on the reflector. When the light source geometry and location are defined, the image varies with respect to the points on the reflector. In geometrical optics, the relationship between object, image and reflecting surface can be described using vector notations.

Let $\mathbf{a}_i$ be the direction of incident light from a light source to a surface, $\mathbf{a}_r$ be the direction of reflected light, and $\mathbf{n}$ be the normal to the surface at the point of incidence. The vectors $\mathbf{a}_i$, $\mathbf{a}_r$, and $\mathbf{n}$ are unit vectors, and p is the point of incidence on the reflector surface (see Fig. 2). For specular reflections, Snell's Law[5] states that

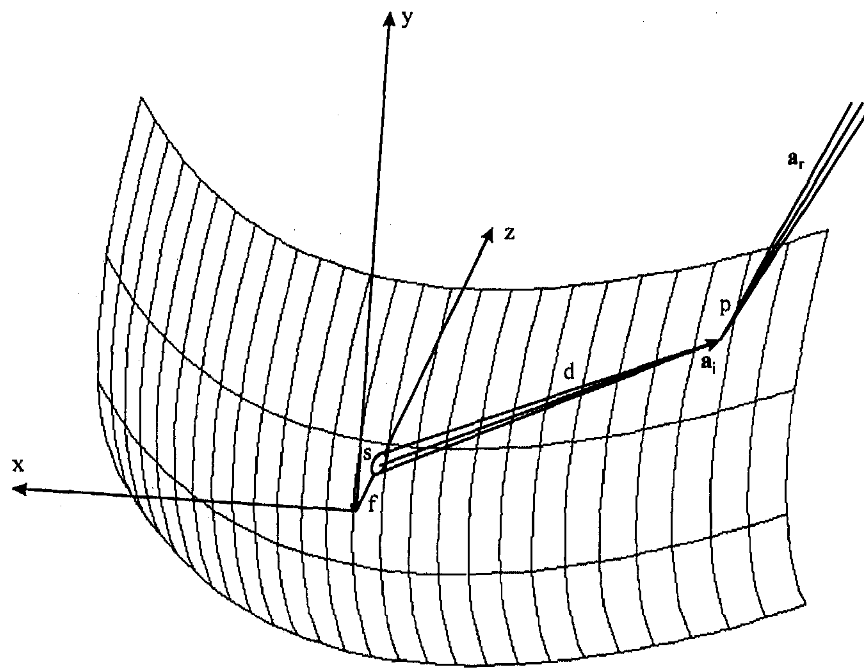


Fig. 2 Diagram of Light Rays from Source to Reflector to Screen

$$-\mathbf{a}_i \cdot \mathbf{n} = \mathbf{a}_r \cdot \mathbf{n} \qquad (1)$$

where the vectors $\mathbf{a}_i$, $\mathbf{a}_r$, and $\mathbf{n}$ are coplanar, which means that $\mathbf{a}_i$, $\mathbf{a}_r$, and $\mathbf{n}$ are linearly dependent:

$$-\mathbf{a}_i + \mathbf{a}_r = r\mathbf{n} \qquad (2)$$

where r is a coefficient.

For a given reflector, if the center of the light source is located at point (0, 0, f) and extends from $-l_x$ to l_x, $-l_y$ to l_y, and $-l_z$ to l_z, then (Δl_x, Δl_y, $f+\Delta l_z$) can represent an arbitrary point of the light source. The value for the incident vector from an arbitrary point of the light source to an arbitrary point on the reflector, p, can written

$$\mathbf{a}_i = \left(\frac{p_x - \Delta l_x}{d}, \frac{p_y - \Delta l_y}{d}, \frac{p_z - f - \Delta l_z}{d} \right) \qquad (3)$$

where

$$d = [(p_x - \Delta l_x)^2 + (p_y - \Delta l_y)^2 + (p_z - f - \Delta l_z)^2]^{1/2} \qquad (4)$$

At the center of the light source,

$$\mathbf{a}_i = \left(\frac{p_x}{d}, \frac{p_y}{d}, \frac{p_z - f}{d} \right) \qquad (5)$$

For simplicity, the reflector is approximated as a parabola with the center of filament on at the focal point (i.e., the light is reflected straight ahead from the reflector surface), so that $\mathbf{a}_r$ = (0, 0, 1). That is approximately the case for the high intensity area of a low beam head lamp centered at 2R 1.5D, which is near the areas of greatest concern for both excess glare light and excess foreground light (see next section). The normal of the reflector surface is thus

$$r\mathbf{n} = -\mathbf{a}_i + \mathbf{a}_r = \left(\frac{-p_x}{d}, \frac{-p_y}{d}, \frac{d - p_z + f}{d} \right), \qquad (6)$$

Solving for $\mathbf{n}$ and normalizing, we obtain for a parabolic surface:

$$\mathbf{n} = \left(\frac{-p_x}{[p_x^2 + p_y^2 + (d - p_z + f)^2]^{1/2}}, \frac{-p_y}{[p_x^2 + p_y^2 + (d - p_z + f)^2]^{1/2}}, \frac{d - p_z + f}{[p_x^2 + p_y^2 + (d - p_z + f)^2]^{1/2}} \right). \quad (7)$$

The direction of reflection for any point of light source is

$$\mathbf{a}_r = \mathbf{a}_i - 2\,\mathbf{n}\,(\mathbf{a}_i \cdot \mathbf{n}) = (\,I - 2E\,)\,\mathbf{a}_i, \quad (8)$$

where

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (9)$$

and

$$E = \begin{pmatrix} n_x^2 & n_x n_y & n_x n_z \\ n_y n_x & n_y^2 & n_y n_z \\ n_z n_x & n_z n_y & n_z^2 \end{pmatrix} \quad (10)$$

Finally,

$$(a_{r_x}, a_{r_y}, a_{r_z})^T = \begin{pmatrix} -2n_x^2+1 & -2n_x n_y & -2n_x n_z \\ -2n_y n_x & -2n_y^2+1 & -2n_y n_z \\ -2n_z n_x & -2n_z n_y & -2n_z^2+1 \end{pmatrix} (a_{i_x}, a_{i_y}, a_{i_z}). \quad (11)$$

Equation (11) determines the vector of the reflecting rays from any point on the reflector, (p_x, p_y, p_z), given the incident ray vector, $\mathbf{a}_i$. For any point of the light source, the incident ray vector, $\mathbf{a}_i$, can be calculated by using equation (3). Thus given an arbitrary point on the reflector and an arbitrary origination point in the light source, we have a method to calculate $\mathbf{a}_r$. By calculating $\mathbf{a}_r$ for light source points at a specific reflector point we can calculate an image size (in degrees) for that specific reflector point. The image size of the HID light source for a FFR using a 24 mm "focal distance" has been calculated. The results are shown in Fig. 3.

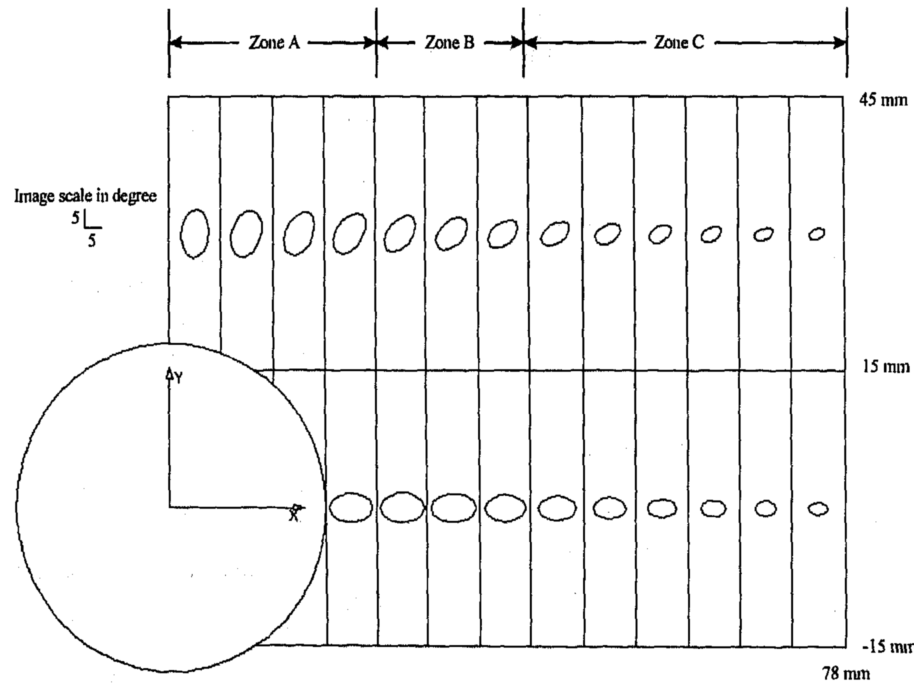


Fig. 3 Image Size of HID Light Source for Points of a Reflector

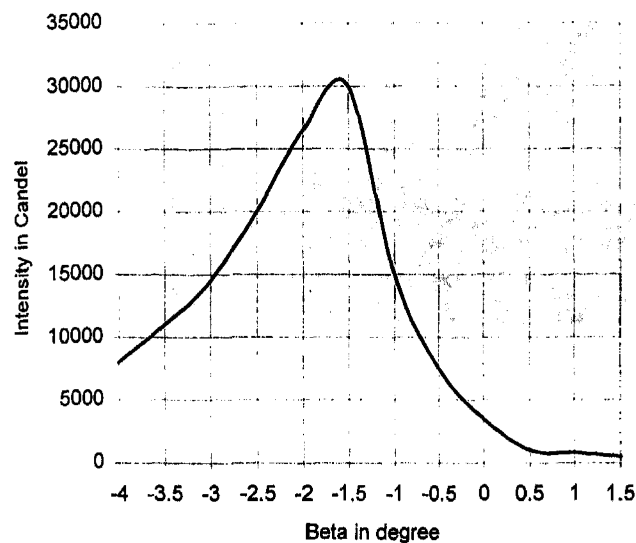


Fig. 4 Intensity Distribution Curve for a Vertical Section at the high intensity Area of an SAE Low Beam

INTENSITY DISTRIBUTION CURVE OF THE HIGH INTENSITY AREA

Based on SAE[6] and FMVSS108[7] photometry requirements for the low beam head lamp, an intensity distribution curve of a vertical section through the high intensity area (i.e., 4D to 1.5 U at 2R) can be devised (Fig. 4). To construct the high intensity area of the beam pattern using a FFR with an HID D2S source, the image size must match the intensity distribution curve. The intensity distribution for a given area of a beam pattern is a superposition of all the images produced by the reflector and projected to the area. Therefore all the images superimposed on that area must be restricted in their size, or else the intensity distribution shown in Fig. 4 can not be achieved.

ZONAL DIVISION OF THE FFR SURFACE

The design strategy used here to divide the reflector surface into three zones based on local angular spread (Fig. 3). The lumen collection at each area of the reflector are calculated and shown in Fig. 5. In order to achieve a high performance beam pattern, Zone A is used for horizontal light spread since it has the largest image size. Zone B has intermediate image sizes and is used for smoothing the beam pattern. Zone C has the smallest image sizes and is used for the high intensity area. The beam pattern simulation results from the FFR are shown in Fig. 6.

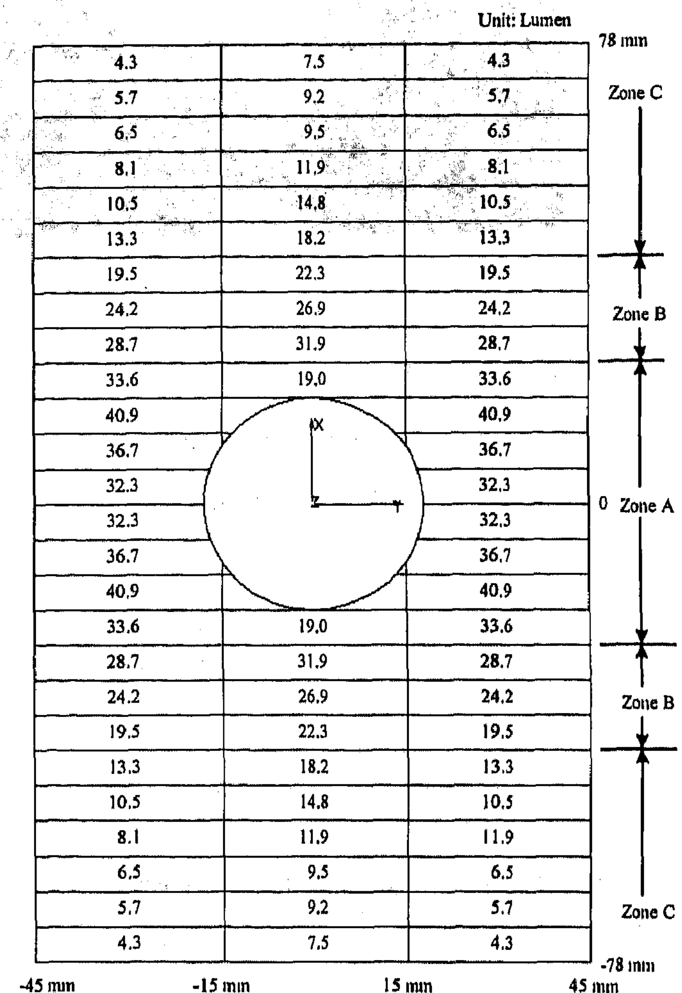


Fig. 5 Light Collected from an HID Source at Each Area of the Reflector (Focal distance = 24 mm)

CONCLUSION

The design difficulties for a FFR using an HID source have been discussed, and solutions have been proposed. The strategy used here allows a high degree of freedom to manipulate the reflector surface so that light can be collected and projected to a desired beam pattern. More importantly, the relationship between the source (object), the spread (image) and the reflector surface has been mathematically defined. This mathematical definition can be used to develop computer-aided engineering (CAE) programs to precisely design FFRs with an HID source and conduct accurate simulations which in turn can to help further reduce the optical design lead-time in product development.

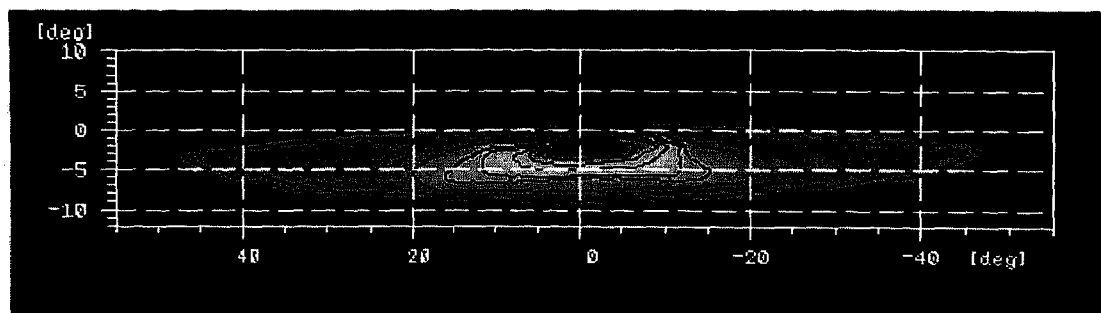


Fig. 6 Beam Pattern of FFR using HID Source

ACKNOWLEDGMENTS

The authors gratefully acknowledge the computer programming help of Sam Banerjee. Also we thank Franz-Josef Kalze of Hella KG for several helpful discussions.

REFERENCES

1. Hogrefe, H. and Neumann, R., "Application of Free Form Reflectors in Modern Headlamp Systems," SAE Technical Paper #950593, 1995.

2. Brandenburg, W., "The Replacement of Parabolic Reflectors by 'Free Form' Reflectors," SAE Technical Paper #960926, 1996.

3. Huhn, W. and Hege, G., "High Intensity Discharge (HID) - Experience for More Than 3-1/2 Years of Commercial Application of Litronic Headlamps," SAE Technical Paper #950591, 1995.

4. Hendrischk, W., "Trends in the Development of Vehicle Front Lighting," SAE Technical Paper #960923, 1996.

5. Optical Society of America Handbook of Optics, Bass, M., ed., McGraw-Hill, Inc., New York, pp.1.28 (1995).

6. SAE Ground Vehicle Lighting Standards Manual, SAE HS-34, "Performance Requirements for Motor Vehicle Headlamps SAE J1383", pp. 19-99 (1996).

7. United States Code of Federal Regulations, Title 49, Section 571.108 (1994)

970903

Styling and Reflector Design - New Reflector Shapes for Low Beam, High Beam and Foglamps

Doris Boebel
Robert Bosch GmbH

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

For the automotive industry, the styling of a car is an important feature. Since the appearance of a car is strongly characterized by the headlamp, stylistic aspects have an increasing influence on the development of headlamps. At the same time, the trend goes toward reducing the constructional volume available to the headlamp and also integrating other functions such as foglamps. To fulfill all these requirements, new reflector shapes had to be created.

In this paper, different reflector concepts for low beam, high beam and foglamps are presented. They will be discussed under consideration of stylistic aspects (e.g. space for the required bulb shield, semi-clear outer lens) and technical aspects (e.g. thermal stress of plastic lenses).

INTRODUCTION

Modern headlamps are playing an important role as stylistic elements of cars. Simultaneously, a high quality light pattern is very important for the safety and comfort of drivers. Because of this situation, the challenge for the headlamp manufacturer is to develop attractive headlamps from the stylistic point of view and to improve the quality of light patterns.

Free-form reflectors have the potential to fulfill these requirements [1,2]. The geometry of a free-form reflector shape can not be described in a simple mathematical way like a parabola or an ellipse. But parts of the reflector behave like these conventional geometries. In the horizontal cross section the basic concepts for the calculation of a free-form reflector shape can be described as continuous transitions of parabolic, hyperbolic and elliptical areas. In the vertical cross section they are nearly parabolic. To connect the horizontal and the vertical concept to a complete reflector shape and to achieve the required light pattern, a program is necessary which allows the local variation and optimization of all reflector areas. Our optical design program CAL (Computer Aided Lighting) [1,3] is a tool for the calculation of such reflector shapes. The calculation can be done fully automatically, semi-automatically or manually by the optical engineer. With CAL it is also possible to calculate the light distribution in the plane of the outer lens. This is important for the design of headlamps that incorporate plastic lenses.

In the following sections reflector shapes for low beam, high beam and foglamps are presented. Some technical and stylistic aspects are discussed and specific examples are shown.

REFLECTOR SHAPES FOR LOW BEAM

There are trends in the USA and in Europe which require new reflector shapes for low beam applications. In the USA, visually aimable headlamps will be introduced in the near future. In Europe, car designers are asking more and more often for headlamps with clear outer lenses. Both leads to a demand for low beam reflectors which produce high quality light distributions with cutoff lines.

Reflectors with smooth surfaces are a good choice to achieve the required quality of light patterns. Faceted reflectors have the disadvantage that the steps between the individual segments can cause stray light. Especially the horizontal steps are critical in view of keeping glare values low and maintaining a good contrast at the cutoff line.

In principle, two different reflector concepts are applied to create reflector shapes for low beam light patterns: the diverging (Fig.1) and the converging

reflector concept (Fig.2). In both concepts the light from the outer part of the reflector is directed nearly parallel to the optical axis (parabolic behavior). This light is used to form the center of the light distribution with the highest illuminance. The light from the central part of the reflector is used to form the horizontal spread of the light distribution. In this point the two concepts differ.

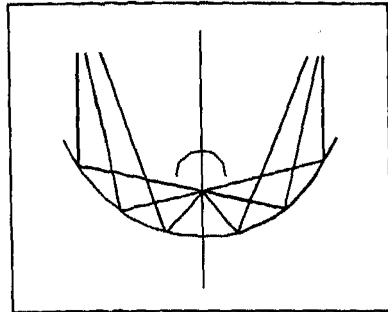


Fig. 1: Diverging reflector concept for low beam (horizontal cross section).

The horizontal cross section of the diverging concept (Fig.1) shows that the rays from the central part of the reflector are forming the horizontal spread of the light distribution without crossing the optical axis (hyperbolic behavior). This feature is important if a bulb shield is required as styling element or to block the direct light of the bulb. Furthermore, the diverging reflector concept is a good choice if the depth of the reflector is limited by the dimensions of the headlamp. The hyperbolic characteristic in the central part leads to flat reflector shapes in comparison to converging-concept reflectors. The light distribution of an SAE application using the diverging reflector concept is shown in Fig. 2.

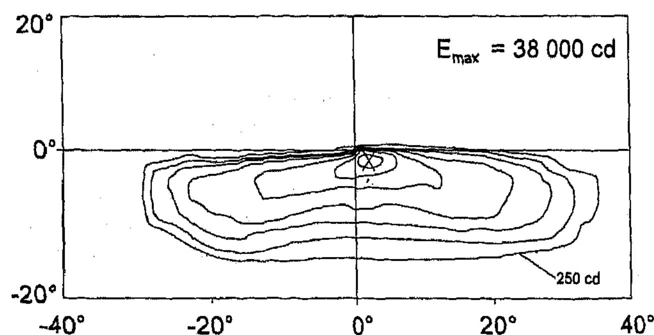


Fig. 2: Light pattern of an SAE low beam, diverging reflector concept. HB4 bulb. Rectangular reflector 160 mm wide, 100 mm high.

In the converging reflector concept shown in Fig.3, the rays from the central part of the reflector are crossing the axis to form the horizontal spread of the light distribution (elliptical behavior).

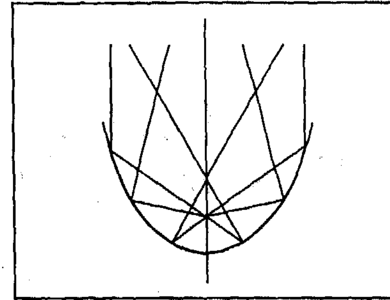


Fig. 3: Converging reflector concept for low beam (horizontal cross section).

The converging reflector concept is a good choice for small reflectors. Typically, they have a small focal length. Therefore a high amount of the luminous flux of the bulb can be used for the light distribution. But the depth of converging concept reflectors is higher than that of reflectors of comparable width calculated with the diverging concept. Fig. 4 shows the light distribution of an ECE low beam, calculated with the converging concept.

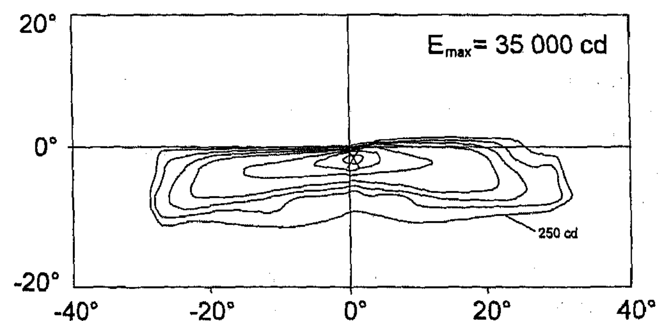


Fig. 4: Light pattern of an ECE low beam, converging reflector concept. H7 bulb. Rectangular reflector 110 mm wide, 100 mm high.

REFLECTOR SHAPES FOR HIGH BEAM

In view of customer's demand for clear outer lenses for the entire headlamp, new high beam reflector shapes have to be designed as well. On the other hand there are customer design requests for semi-clear outer lenses for high beam, in most cases combined with projector-type units for low beam. In this case the optical engineer has to develop a reflector shape which forms the light distribution together with optical profiles in the lens, but all profiles with the same scattering angle.

The concepts applied to the calculation of high beam reflectors are in principle the same as those for the calculation low beam reflectors. They differ in the central part of the reflector. Due to the smaller horizontal spread of high-beam light distributions, rays reflected in the central part of the reflector have smaller angles to the

axis than in low beam reflectors. The horizontal cross sections of a diverging and a converging reflector concept are shown in Fig. 5.

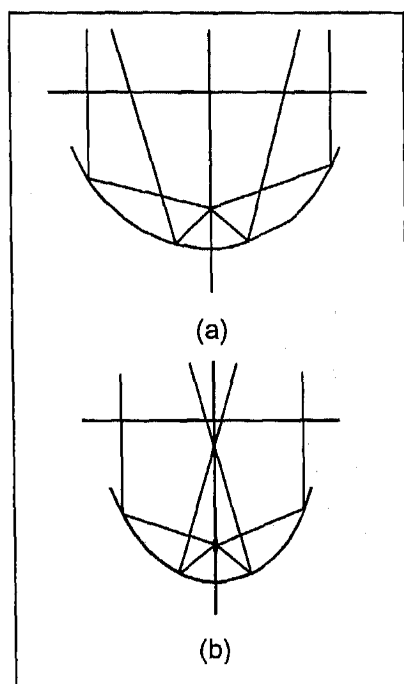


Fig. 5: Diverging (a) and converging (b) reflector concepts for high beam applications (horizontal cross sections).

The sections in Fig.5 show outer lenses in front of the reflector. For two specific examples, the light distributions in the outer lens plane are shown in Fig.6.

The distance between the filament and the lens is 100 mm in both cases. The rays of the central part of the reflector have an angle of about 15 degrees with respect to the axis. In this special case, the diverging reflector concept generates two hot spots on the left and the right side respectively, whereas the converging reflector concept generates a single hot spot in the middle of the lens plane. The maximum illuminance of the spot in the middle (Fig.6a) is about 1.5 times higher than that of the spots on the sides (Fig.6b).

For headlamps that incorporate plastic lenses, the thermal stress of the outer lens is sometimes critical. The light distribution in the plane of the outer lens has a major influence on the temperature in the lens. Therefore it is important to choose the reflector concept not merely in accordance with the space for a bulb shield or the total luminous flux. The light distribution in the plane of the lens has to be taken into consideration as well.

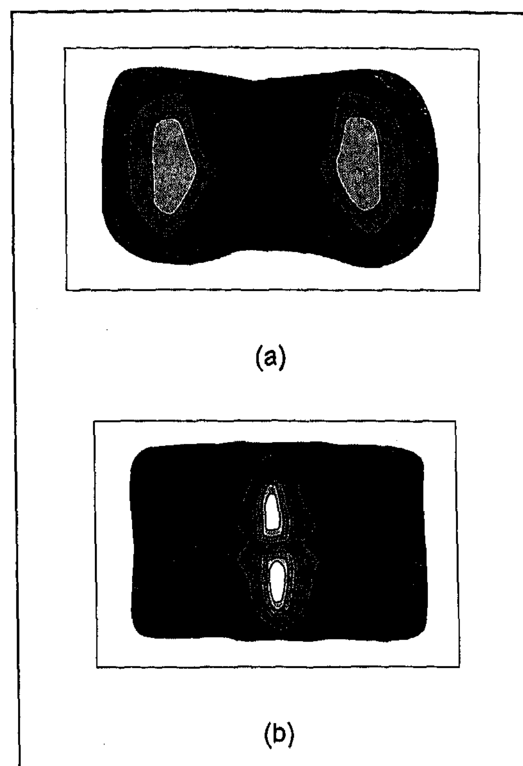


Fig. 6: Light distributions in the plane of the outer lens.

The following example shows a high beam reflector with a semi-clear outer lens. The reflector is calculated with the converging concept. The horizontal cross section in Fig.7 shows that the angles of the rays in the central part of the reflector are smaller than those in reflector shapes for clear outer lenses (Fig.5).

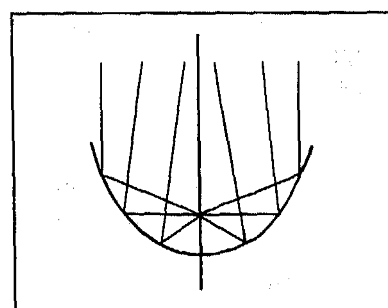


Fig. 7: High beam reflector concept for semi-clear outer lens (horizontal cross section).

The pre-scattered light distribution formed by the reflector (a) and the final light distribution completed by the outer lens (b) are shown in Fig.8.

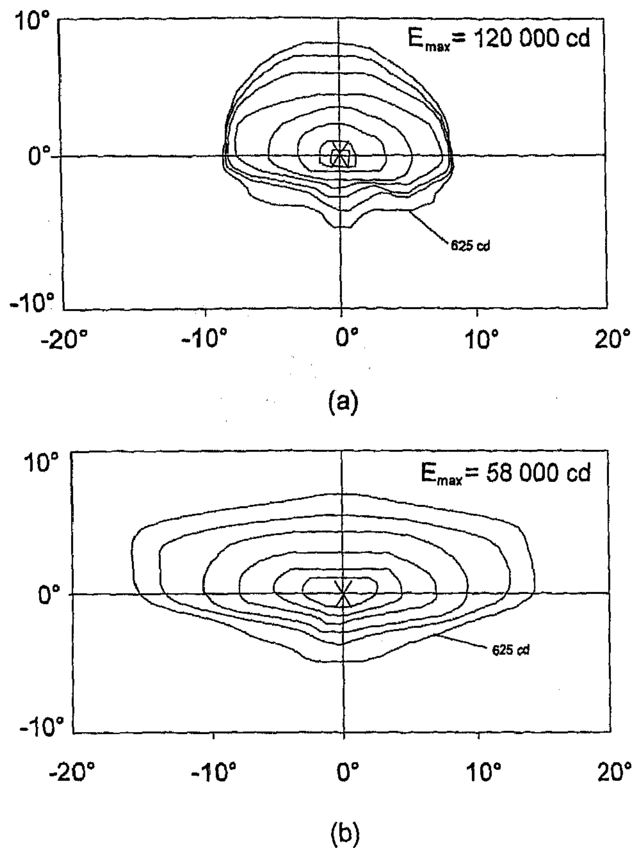


Fig. 8: Pre-scattered light distribution (a) and final light distribution (b) completed by the outer lens.

REFLECTOR SHAPES FOR FOGLAMPS

The trend towards clear or semi-clear outer lenses is the same for foglamps. From the technical point of view, this trend is helpful. The aerodynamic shape required of modern cars also influences the contour of the headlamp outer lens. The resulting high inclination angles make it difficult to form the light pattern with optical profiles in the lens. This is especially true in the case of foglamps, where a very wide spread is required for the light pattern.

For the calculation of reflector shapes for foglamps, a converging reflector concept similar to the one shown in Fig.3 is applied in most cases. Since most foglamps are quite small, it is important to utilize as much of the available luminous flux as possible for the final light pattern.

The first example is a foglamp with a semi-clear outer lens. For this application a H1 bulb is used. A bulb shield is necessary to block the direct light of the bulb and to prevent stray light from those parts of the reflector, which are not designed to contribute to the light pattern. The size of this bulb shield is determined by the available space in the center of the reflector. Here, the available space is very limited because the central rays cross immediatelly in front of the bulb. Hence the bulb shield has to be small and conical in shape. In this case, it is no styling element of the headlamp. The completed light pattern formed by reflector and lens is shown in Fig.9.

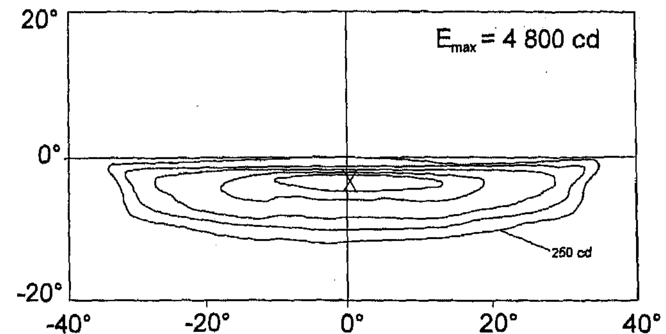


Fig. 9: Light pattern of a fogbeam with a semi-clear outer lens, converging reflector concept. H1 bulb. Reflector 90 mm wide, 70 mm high.

The second example is a foglamp with a clear outer lens. In this case, the light source is a H3 bulb with a transverse filament. For the same reasons as in the example before, a small conical bulbshield is used.

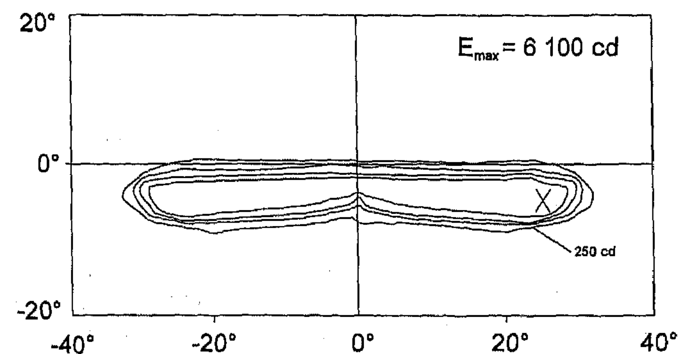


Fig. 10: Light pattern of a fogbeam with a clear outer lens, converging reflector concept. H3 bulb. Round reflector, diameter 85 mm.

Other light sources with transverse or axial filaments are also suitable for foglamps. For each application with a different light source or a different size of the reflector a new reflector shape has to be calculated to achieve a high quality light pattern.

CONCLUSION

For every headlamp, it has to be determined if the stylistic and technical requirements can be fulfilled simultaneously. A selection of typical reflector characteristics has been presented, together with appropriate choices in reflector design. These examples do not offer complete, ready-to-use solutions. The best reflector concept has to be chosen for each headlamp project individually.

REFERENCES

[1] H. Hogrefe, R. Neumann:
Application of Free Form Reflectors in Modern Headlamp Systems
SAE Technical Paper Series 950593, 1995

[2] Willi Brandenburg:
The Replacement of Parabolic Reflectors by 'Free Form' Reflectors
SAE Technical Paper Series 960926, 1996

[3] R. Neumann, H. Hogrefe
Computer Simulation of Light Distributions for Headlamp Systems
SAE Technical Paper Series 910827, 1991

970904

Headlamp Design Variations Using HID in Projector Type Headlamps

Heike Eichler
Robert Bosch GmbH

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

For some years now the new HID technology has been used for headlamp systems to improve safety and comfort during night time driving. A new generation of projector type headlamps starts series production in 1997. The characteristics of the HID bulb and the specific requirements of the light distribution make it necessary to calculate special optical systems to control the high amount of light. There are different possible design strategies. On one hand, it is possible to improve the light pattern, e.g. increase the luminous flux and the range. On the other hand it is possible to build a system with extremely small dimensions. Such a system ensures an excellent light quality even in smallest headlamps and gives way to new styling possibilities. To match such different targets, some parameters of the reflector and the lens have to be optimized. Some examples of projection headlamps which are designed for different applications will be discussed.

INTRODUCTION

In 1991 Bosch introduced the first headlamp with a High Intensity Discharge bulb (Litronic) to the market, it was a projector type headlamp (1). Since then, a lot of test drivings have been done, and the Litronic system which is also known as Xenon Light has been improved continuously (2). The number of cars equipped with a HID lamp has grown. Nowadays nearly all major car manufacturers have already introduced the HID headlamp to the market or are preparing series production. With the validation of the homologations for gas discharge bulbs and headlamps using HID, the floor is open for the development of a new generation of optical systems which give a higher amount of light on the street and a higher range. The high luminous flux of the HID lamp can be used more efficiently. The optics are also developed in a more global way, to give way to harmonization efforts of the SAE and ECE light pattern. According to the customers' wishes, one of the most important aspects is to make the system smaller so that less space is needed and a maximum styling freedom is achieved.

LIGHT PATTERN - ASPECTS TO IMPROVE

The high light output of the HID lamp opens a wide range of completely new possibilities for the light pattern design. However, due to the high luminances that can be achieved with the HID lamp, the light pattern has to be shaped very carefully and in a rather controlled fashion. The advantages in comfort and safety should be given in all driving situations, even under adverse road or weather conditions. A characteristic feature of the light pattern is the strong asymmetry between the right and the left side of the road. The hot spot on the right side of the street has a high luminance and thus results in excellent range. Early recognition of street direction and possible obstacles enables the driver to choose the appropriate cruise speed. Since the superposition of the two light patterns for both vehicle headlamps can lead to zones of high illumination, the 15-degree slope is cut off above the horizontal line. Thus, glare for the oncoming traffic is controlled efficiently.

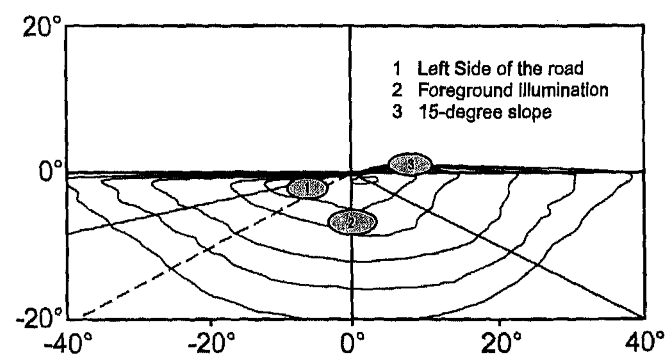


Fig. 1: Zones in the light pattern which need special control.

The illumination of the left side of the street is always a compromise between range and glare. A high amount of

lateral dispersion can reach 60 degrees. This is realized by the Compact PES System which has an excellent lateral dispersion.

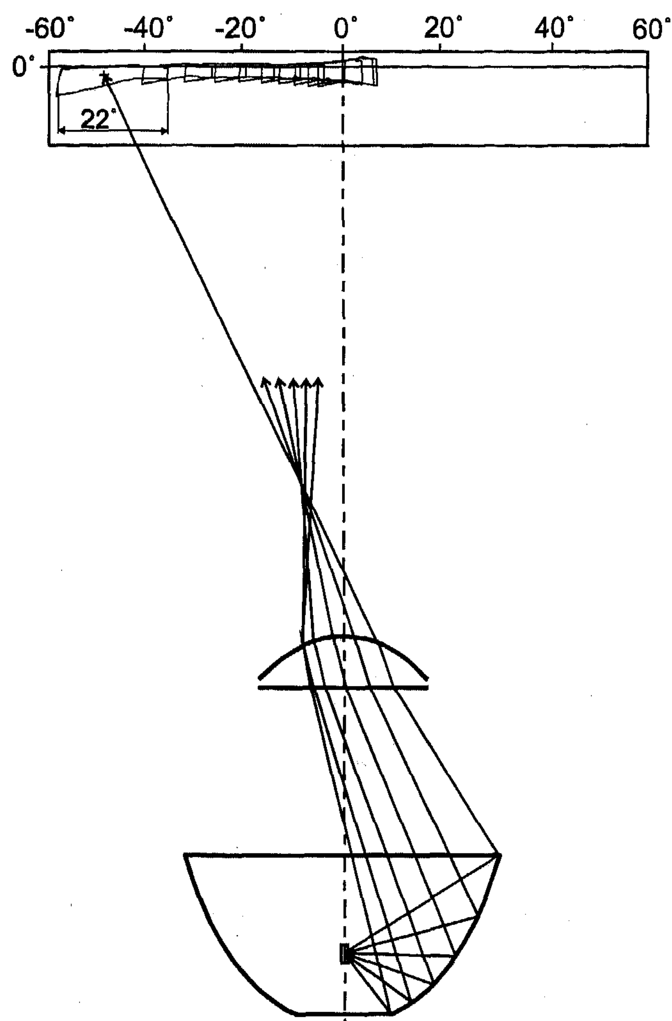
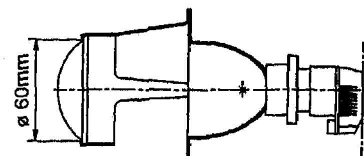


Fig. 3 shows how the high lateral dispersion of the Compact PES is achieved.

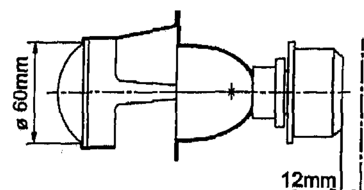
NEW PROJECTOR TYPE SYSTEMS

There are two new PES systems developed by Bosch. The comparison of these systems to the current series PES system is shown in Fig. 4. The overall length as well as the volume has been reduced by using smaller focal lengths. These systems can be used more flexibly and in headlamps with small dimensions. This might become an important point for lighting systems of the future because of better aerodynamics, packaging advantages and design freedom.

Current PES



New PES
- improved range
- lower foreground illumination



Compact Litronic PES

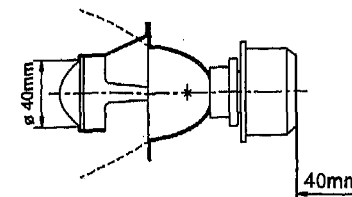


Fig. 4: Litronic PES low beam.

PES WITH LENS DIAMETER 60MM - The light pattern has been improved. As shown in the isocandela light pattern, the maximum luminous intensity has been increased from 30000cd to 40000 cd which leads to improved range. Because the cut-off line is built by a diaphragm the glare can be precisely controlled. The foreground illumination has been reduced further, so that the ratio between hot spot and foreground becomes more comfortable for driving. The lateral dispersion and the luminous flux remain roughly the same. The isocandela light pattern and the bird's-eye view are shown in Fig. 5 and Fig. 6.

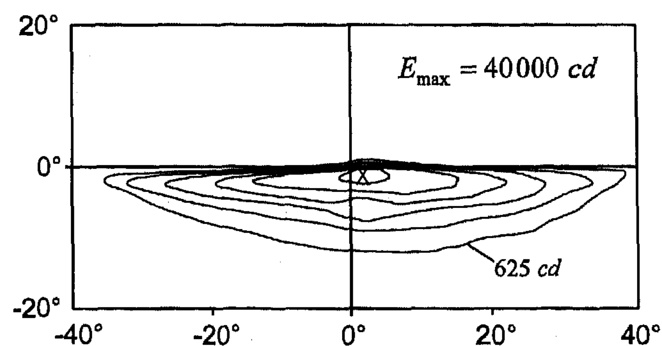


Fig. 5: SAE isocandela light pattern of new PES lens diameter 60mm.

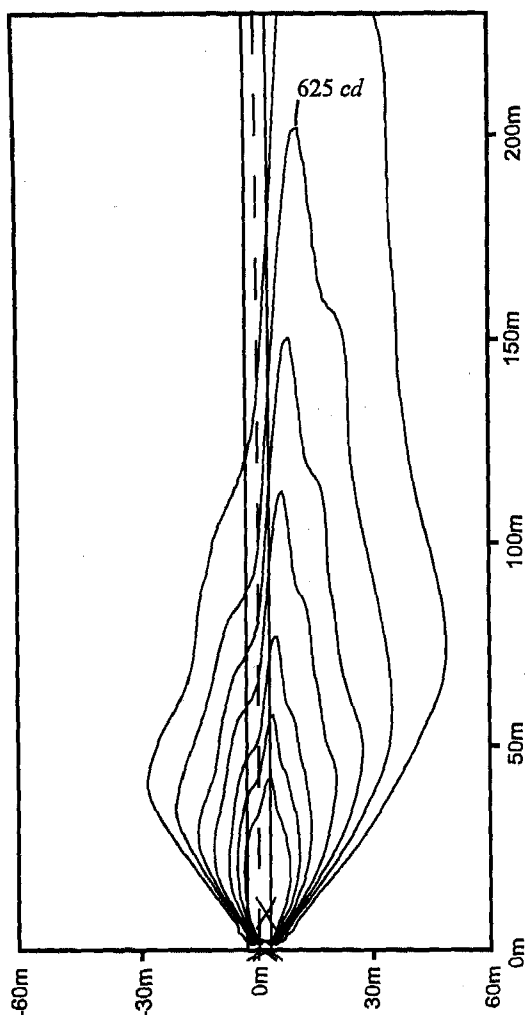


Fig. 6: Bird's-eye view of SAE light pattern lens diameter 60mm.

COMPACT PES - LENS DIAMETER 40MM - A completely different design strategy is necessary when minimizing a Litronic PES system down to the smallest possible dimensions. If done sucessfully, the light output is nearly equal to that of the PES system with 60mm diameter lens, in quality as well as in luminous flux. The isocandela light pattern and the bird's-eye view are shown in **Fig.** 7 and Fig. 8.

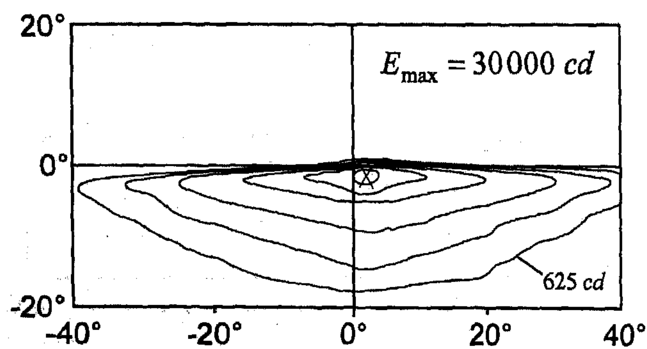


Fig. 7: SAE isocandela light pattern of Compact PES.

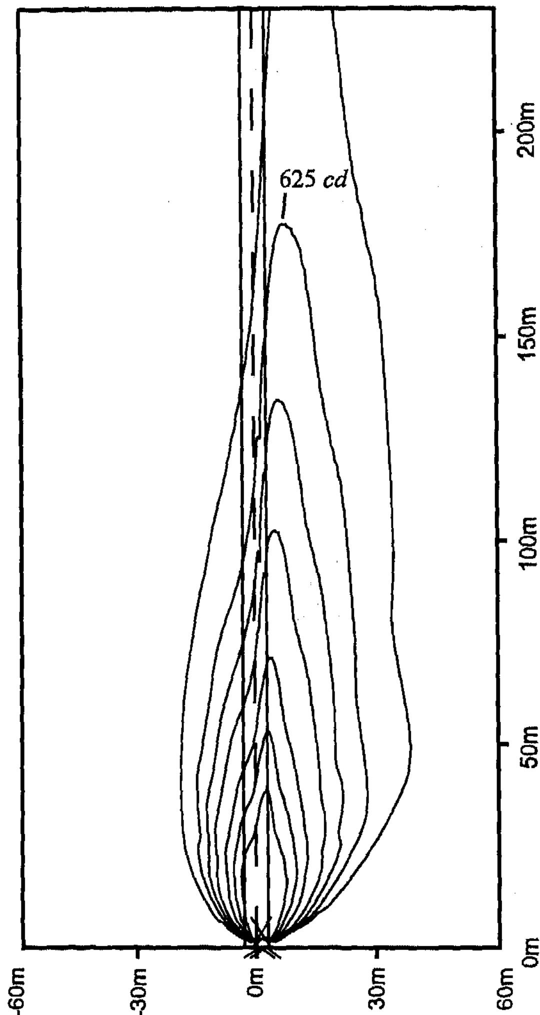


Fig. 8: Bird's-eye view of SAE light pattern Compact PES with 40mm lens diameter.

The light pattern shows a maximum of 30000 cd and a moderate foreground illumination. The system accomodates even in the most restrictive space requirements.

SIZE OF SIGNAL IMAGE - Especially with the HID lamp, the signal image of the headlamp should have a minimum illuminated size. Several investigations in recent years have proven that small illuminated surfaces lead to an increase of discomfort glare. This effect, which was investigated and confirmed also for Litronic applications (6), can easily be avoided by enlarging the signal image. There are several possibilities of creating an illuminated area around the lens, one of them being the implementation of an additonal ring-shaped reflector (7). With the Compact PES which uses only a 40mm diameter lens, enlarging the signal image becomes even more important than with a 60mm diameter lens. By using a signal image closer in size to that of conventional reflector-type headlamps, the discomfort glare can be reduced drastically. The design possibilities of the illuminated surface are virtually unlimited.

CONCLUSION

More light from even smaller and more compact headlamps is a trend in modern vehicle lighting. Better aerodynamics, packaging advantages and more design freedom are requested. The new generation of projector type headlamps offers a convincing solution. The compact PES system reduces the overall length by 25% compared to the current series system. Luminous flux and range are essentially the same. The lens diameter is 40mm instead of 60mm, and the headlamp volume has been cut in half. This Compact PES sets new standards and facilitates introduction of Litronic technology to the compact and subcompact class. With the new 60mm compact PES. the overall length is reduced by 10%. Nevertheless, the range and the light distribution is improved.

REFERENCES

(1) B. Woerner, R. Neumann:
Motor Vehicle Lighting Systems with High Intensity Discharge lamps
SAE Technical Paper Series 900569, 1990

(2) W. Huhn, G. Hege:
High Intensity Discharge Headlamps (HID)-Experience for More Than 3-1/2 Years of Commercial Application of Litronic Headlamps
SAE Technical Paper Series 950591, 1995

(3) R. Neumann, H. Hogrefe:
Computer Simulation of Light Distributions for Headlamp Systems
SAE Technical Paper Series 910827, 1991

(4) H. Hogrefe, R. Neumann:
Application of Free Form Reflectors in Modern Headlamp Systems
SAE Technical Paper Series 950593, 1995

(5) ECE Regulation No. 99:
Uniform Provisions Concerning the Approval of Gas-Discharge Light Sources for Use in Approved Gas-Discharge Lamp Units of Power-Driven Vehicles, 1996

(6) M.J. Flannagan, M. Sivak, D.S. Battle, T. Sato, E.C. Traube:
Discomfort Glare From High Intensity Discharge Headlamps: Effects of Context and Experience, (Report No. UMTRI-93-10),
University of Michigan Transportation Research Intstitute, 1993

(7) R. Neumann:
Improved Projector Headlamps Using HID (Litronic) and Incandescent Bulbs SAE Technical Paper Series 940636, 1994

970906

Headlighting - Toward a Model of Customer Pleasing Beam Patterns

Stephen M. O'Day, Christopher H. Stone, Daniel D. Jack, and Vivek D. Bhise
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The headlamp beam pattern development process contains both subjective and objective evaluations. The subjective evaluation, communication between the customer and the engineer, was developed in previous work [1][2]. This paper presents exploratory models used in the identification of objective photometric variables of a beam pattern that relate to the subjective impression of the beam pattern. Additional research will allow use of the photometric variables and their selected ranges for designing and evaluating beam patterns to achieve improved customer pleasing beam pattern driving experiences.

BACKGROUND

For many years engineers have recognized two important conditions for the meaningful evaluation of the products they produce: 1) the evaluation should be done by the intended customer and 2) the evaluation should be conducted in the environment in which the product is intended to be used. In the past many headlamp beam pattern designers have experimented with different ways of satisfying both of these conditions. Such differences in practice have lead to considerable variations in the visual quality of beam patterns. The semantic differential approach to headlamp beam pattern evaluation [1][2] satisfied both of the above stated conditions and was the first step in the process of improving the quality of beam patterns based on customer input. Our previous work has shown the value of acquiring customer input during the headlamp design process. The information gained from the customer can positively influence headlamp beam pattern design and result in a higher quality beam pattern as measured by customer satisfaction. This paper reports on the next steps taken toward the development of a model for customer pleasing beam patterns. These steps involved the analysis of the photometric characteristics of the beam patterns and relating them to the previously acquired customer subjective evaluations of the headlight beam patterns.

ANALYZING THE BEAM PATTERN

The analysis of the physical aspects of headlamp beam patterns started with photometric measurement of each headlamp system using a goniometer (see Figure 1). Twelve beam patterns representing a wide cross-section of currently available headlamp technologies including conventional, complex reflectors with smooth form back cans, complex reflectors with faceted back cans and projectors were measured.

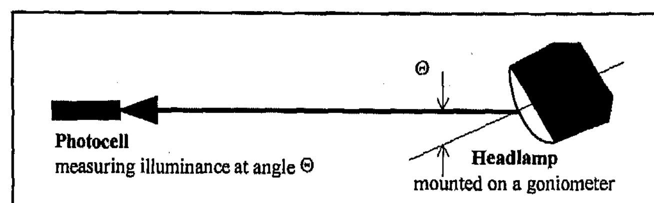


Figure 1 - Photometric measurement of a headlamp

Each headlamp was measured on the goniometer over a horizontal range from 40° left to 40° right at .25° intervals and over a vertical range of 4° up to 15° down at .25° intervals. Photometric data from the right and left headlamps was input into a computer program that combined them to form a complete beam pattern. The program then "projected" the beam pattern (i.e. illumination) by interpolating the beam pattern data onto the road surface at every 0.1° as it would appear from the view point of the driver. Road reflectance was calculated according to research by Bhise et al. [3]. Multiplying the computed reflectance coefficients by incident illumination provided pavement luminance values. These luminance values are here referred to as brightness. The 0.1° grid of brightness values thus provided a more complete reconstruction of each beam pattern. A brightness, iso-foot-Lambert(fL), contour plot of one of these computer reconstructed beam patterns is shown in Figure 2.

The plot is presented in polar coordinates with respect to the driver's eye position. The negative numbers on the vertical axis refer to what the driver would see looking DOWN from the horizon that many degrees. On the

horizontal axis the negative numbers refer to what the driver would see looking LEFT and the positive numbers correspond to what the driver would see looking RIGHT. Five milli-foot-Lamberts (mfL) was selected as the lowest brightness level for plotting. Each succeeding contour represents a brightness level at least 50% higher than the previous contour. The assumption was that a 50% increase, or decrease, in road brightness will be perceptually noticeable to the human eye. Thus these polar contour plots should closely represent the discernible levels of brightness on the road surface as perceived by the driver.

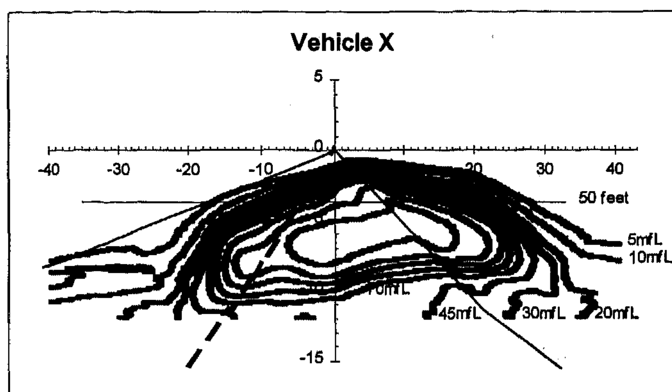


Figure 2 - Brightness (iso-fL) contour plot of Vehicle X's beam pattern

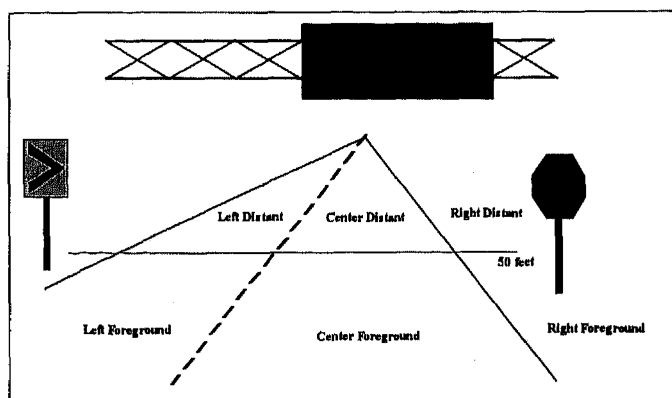


Figure 3 - Road scene showing the location of the six zones.

PHYSICAL CHARACTERISTICS

Based on a previously reported factor analysis of subjective data [1][2] it was determined that four beam pattern factors were considered important to the customers: uniformity, aim, power and comfort. The goal was to find physical measures within the photometry data that would characterize these factors. Though not present in the factor analysis it was decided that another factor, beam shape, should be included to arrive at a more complete analysis of a beam pattern. Each of these factors, derived from customer input, describes a distinctly different aspect of a beam pattern. Note that they may not be mutually exclusive of one another.

The physical measures described below were thought to have a high probability of being important in quantifying each factor before the analysis and model building process began. However, many of these measures were not included in the models due to criteria that will be described later.

BEAM PATTERN UNIFORMITY - Uniformity is perhaps best quantified by a measure of the changes or gradients in brightness of a beam pattern. Vertical gradients (along the direction of travel) and horizontal gradients (perpendicular to the direction of travel) were measured for each beam pattern. Maximum and average gradients as well as maximum and average second order gradients occurring in each zone (see Figure 3) of the beam pattern were also considered.

Each of these simple gradient measures could have a problem characterizing uniformity for both bright and dim beam patterns. A bright beam pattern, for example an HID lamp, would always have higher gradients compared to a conventional lamp regardless of uniformity. To address this potential problem a relative difference in brightness and a second order difference in brightness was calculated for each beam pattern. These measures were calculated as a sum total of all the relative differences for various areas on the road surface:

Relative Difference in Brightness =

$$\sum_{\substack{i=vertical\\angle}} \sum_{\substack{j=horizontal\\angle}} \frac{(b_{i,j+1} - b_{i,j})}{b_{i,j}} \qquad (1)$$

2nd Order Difference in Brightness =

$$\sum_{\substack{i=vertical\\angle}} \sum_{\substack{j=horizontal\\angle}} \left[\frac{(b_{i,j} - b_{i,j-1})}{b_{i,j-1}} - \frac{(b_{i,j+1} - b_{i,j})}{b_{i,j}} \right] \qquad (2)$$

where $b_{i,j}$ is the brightness on the road surface at angles i and j (incremented at every 0.1 degrees) relative to the driver's eye position.

BEAM PATTERN AIM. Many measures used in the study were affected by aim. The most direct measure of aim was the center of mass of the beam pattern. This was defined as the point on the road surface that had the same amount of light around it in any direction. The center of mass was calculated by weighting each illuminated point on the road surface according to the amount of light it received. The vertical and horizontal average then gives the polar coordinates that describes the center of mass of the total beam pattern.

$$\text{Center of Mass}_{\text{thorizontal}} = \frac{\sum_{\substack{i=vertical\\angles}} \sum_{\substack{j=horizontal\\angles}} j * b_{i,j}}{\sum_{\substack{i=vertical\\angles}} \sum_{\substack{j=horizontal\\angles}} b_{i,j}} \qquad (3)$$

$$\text{Center of Mass}_{vertical} = \frac{\sum_{\substack{j=horizontal\\angles}} \sum_{\substack{i=vertical\\angles}} i * b_{i,j}}{\sum_{\substack{j=horizontal\\angles}} \sum_{\substack{i=vertical\\angles}} b_{i,j}} \quad (4)$$

where $b_{i,j}$ is the brightness at angles i and j on the road surface relative to the driver's eye position.

BEAM PATTERN POWER. The power of the beam pattern referred to the amount of light the headlamps put on the road surface. It was measured four ways. One measure was the amount of brightness generated within the 6 zones on the road surface (see Figure 3), as well as, the total amount of brightness within the photometry space. Various brightness ratios were also calculated, including the ratio of foreground brightness to distant zone brightness and the ratio of brightness in the center distant zone to total beam pattern brightness.

A second measure was the calculation of areas bounded by various brightness contour levels projected on the road surface. The power of a beam pattern was assumed to be proportional to the area of a beam pattern at a given brightness level.

The third measure of power was the calculated perimeter of the brightness contours. It was thought that the power of a beam pattern would be proportional to the perimeter of any given iso-footLambert contour.

Lastly, a calculation of the horizontal and vertical dimensions (overall width and height) of the brightness contours was used as a measure of power.

BEAM PATTERN SHAPE Quantifying the shape of a beam pattern was accomplished by borrowing some measures used in the growing field of computer character recognition. These measures included: circularity, ellipsity and convexity. In addition to these measures the area on the road surface covered by the beam pattern and the width of the beam pattern at 6 degrees down (as viewed from the driver's eye) were also included.

Circularity measures how circular a given contour is by looking at the perimeter and the area enclosed by a given contour. A circle has the following properties:

$$\text{Area: } A = \pi R^2 \quad (5)$$

$$\text{Perimeter: } P = 2\pi R \quad (6)$$

where R is the radius of the circle.

From this it can be seen that the ratio P^2/A is always 4π for a circle. The circularity for an iso- footLambert contour is defined as $C = P^2/(4\pi A)$. Therefore, a circle would have a circularity measure of 1.0. Since a circle encloses the greatest area with the smallest perimeter, any other shape would have a circularity greater than 1.0. For instance, a square would have a $C \approx 1.2$ and a triangle would have a $C \approx 1.44$.

Since the shape of a beam pattern is usually elliptical, a measure of ellipsity was employed. An ellipse has the following properties:

$$\text{Area: } A = \pi ab$$

$$\text{Perimeter: } P = 2\pi\sqrt{\frac{a^2+b^2}{2}} \quad (8)$$

where a and b are the lengths of the semiaxies.

Thus, the ratio

$$\frac{P^2}{A} = \frac{2\pi(a^2+b^2)}{ab} \quad (9)$$

for an ellipse. The ellipsity measure, E, was defined as:

$$E = \frac{abP^2}{2\pi A(a^2+b^2)} \quad (10)$$

This measure will equal 1.0 for an ellipse (as well as a circle), but for a square $E \approx 1.27$ and for a triangle $E \approx 1.84$.

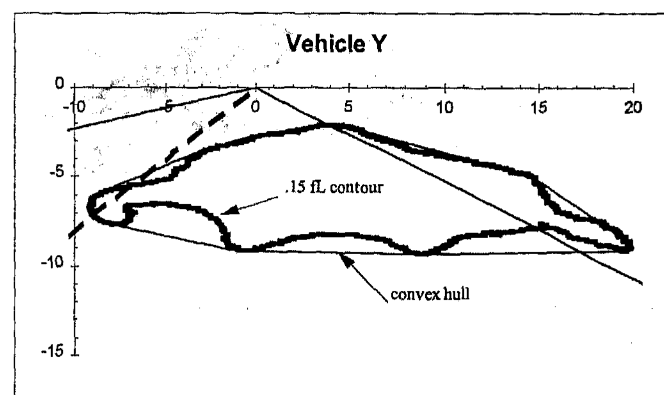


Figure 4 - Convex hull circumscribing the .15 fL contour.

To quantify the bumpiness of a beam pattern a measure of convexity was employed. This measure is a ratio of the perimeter of an iso-footLambert contour to the perimeter of it's convex hull. A convex hull is a shape with no indentations. Figure 4 shows the .15 fL contour of a beam pattern and the convex hull that circumscribes it. The convexity measure was intended to quantify the jaggedness and bumpiness of the beam pattern. A cursory examination of some of the beam pattern shapes suggested that a smooth convex shape may be more pleasing to the customer than a rough or non-convex shape.

TOWARD A MODEL

A preliminary model of customer pleasing beam patterns can be summarized as follows:

Subjective Experience of a Pleasing beam pattern =
f (Physical Characteristics of the beam pattern)

This equation states that the customer's subjective experience of a pleasing headlight beam pattern is equal to some function of the measurable physical (ie. photometric) characteristics of the beam pattern. Measures of the subjective experience of a pleasing headlight beam pattern have been previously acquired and summarized using semantic differentials. Physical measures of the evaluated beam patterns (mentioned above) have been identified and calculated. The only piece of

the model left to determine was the function that relates the subjective measure(s) to the physical characteristic(s).

The identification of the functional relationship was explored in the following manner. Physical photometric measures were calculated for 12 beam patterns. These 12 beam patterns represented a cross-section of lamp technologies and exhibited a wide range in their subjective evaluation scores. Five of the beam patterns were subjectively judged in field evaluations to be pleasing, five were judged to be unpleasant and two were judged to be not especially pleasing but not unpleasant.

Choosing this mix of beam patterns was an acknowledgment that beam pattern design is in essence a "turnip" (see figure 5). There are a few ways to make a good beam pattern and a few ways to make a bad beam pattern, but there are many ways to make a mediocre beam pattern. A quick examination of the beam patterns on the road today show this to be the case. This over abundance of mediocrity could confound attempts to find a functional relationship between subjective and physical measures. Therefore, the 12 beam patterns were specifically chosen to balance the good and the bad and minimize the confusion in between.

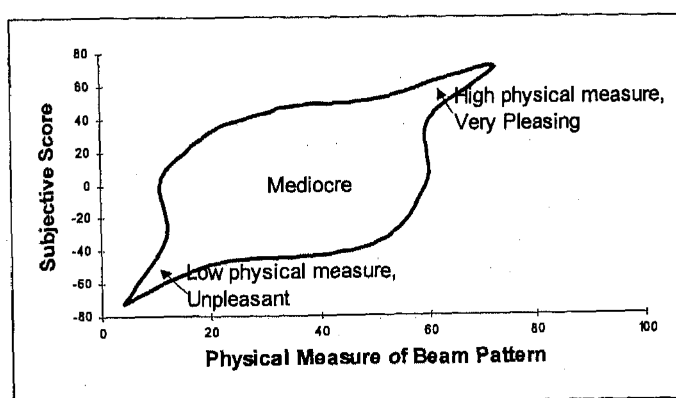


Figure 5 - The Bulge of Mediocrity in headlamp design

On the subjective side of the equation each beam pattern was evaluated by at least 5 to as many as a hundred subjects during a series of dynamic night drives. The evaluations consisted of a set of 25 word pairs related to an aspect of the beam pattern (this process is detailed by Jack et al. [1]). Subject responses were obtained using a five point scale. For example with the word pair "Uniform-Spotty", the subject could respond in one of the following ways: very uniform, somewhat uniform, neither uniform nor spotty, somewhat spotty, very spotty. The subject's response was encoded as a number from +2 to -2. These responses were then averaged over all subjects for each of the 12 beam patterns. Some of these subjective word pair averages were combined to form a summary score for each of the five factors: uniformity, aim, power, shape, comfort. The composition of the subjective score for each factor is shown in Table 1. Each factor's constituent word pair averages were added together and scaled to arrive at the summary score. These summary subjective factor scores were then correlated with the physical measures.

The physical measures mentioned above were calculated for each beam pattern and along with their subjective scores were subjected to a stepwise regression analysis [5]. The results of these analyses served as a starting point to identify the functional relationships between the subjective impressions and the measured physical attributes of the beam patterns. To complete each regression model physical measures were added to or subtracted from those identified by the step-wise regression analysis based on the following criteria:

1. Each model should be composed of physical measures that in some rational way make sense with respect to the subjective measure it is attempting to describe.

2. Each model should be as simple as possible with little or no redundancy.

Table 1 -Composition of the subjective summary factor scores

UNIFORMITY =	Uniform-Spotty score + Smooth-Choppy score + Blended-Streaky score + Unified-Divided score + Balanced-Uneven score + Uniform-Streaky score.
AIM =	Balanced-Uneven score + Elongated-Truncated score + Long-Short score + Aimed-Misaimed score + Right-Left score.
POWER =	Sufficient-Insufficient score + Long-Short score + Strong-Weak score.
SHAPE =	Spacious-Defined score + Long-Short score + Elongated-Truncated score.
COMFORT =	Comfortable-Uncomfortable score + Sufficient-Insufficient score + Secure-Insecure score + Pleasant-Unpleasant score.

The results of this step-wise regression process are shown in Table 2 along with the correlation coefficients and R^2 values between the physical measures and the 5 subjective factors. The correlation between the subjective factors and the corresponding regression models are also shown in Table 2 and are marked with an asterisk. These individual regression models for each factor can be restated as follows:

Subjective UNIFORMITY = $a_u(\Delta_2) + b_u(\Delta_{max}) + c_u$
Subjective AIM = $a_a(M_h) + b_a(M_v) + c_a(W_6) + d_a(C_{30}) + e_a$
Subjective POWER = $a_p(B_{cd:t}) + b_p(W_6) + c_p(P_{30}) + d_p(HV_{20}) + e_p$
Subjective SHAPE = $a_s(W_6) + b_s(R_{70}) + c_s$
Subjective COMFORT = $a_c(B_{cd:t}) + b_c(\Delta_{max}) + c_c(HV_{20}) + d_c$

where

Δ_2 = the sum of the second order difference in brightness
Δ_{max} = the maximum difference in brightness
M_h = the horizontal center of mass
M_v = the vertical center of mass
W_6 = the width of the beam pattern at 6 degrees down.
C_{30} = the convexity of the 30 mfL line.
R_{70} = the circularity of the 70 mfL line.
P_{30} = the perimeter of the 30 mfL line.

$B_{cd:t}$ = the ratio of the center distant brightness to the total brightness of the beam pattern.

HV_{20} = the horizontal * vertical dimensions of the 20 mfL line.

and a, b, c, d, e are regression coefficients.

Table 2 - Model factors/physical characteristics and statistics

UNIFORMITY	Correlation	R^2 Value
sum 2nd order Difference in Brightness	-0.802	0.643
Max Difference in Brightness	-0.557	0.310
UNIFORM regression model *	**0.825**	**0.681**
AIM		
Horizontal Center of Mass	0.118	0.014
Vertical Center of Mass	-0.499	0.249
Width of the beam pattern at 6 degrees down	0.659	0.434
Convexity of the 30 mfL line	0.755	.570
AIM regression model *	**0.943**	**0.888**
POWER		
ratio of brightness in the Center Distant zone to Total brightness	-0.406	0.165
Width of the beam pattern at 6 degrees down	0.672	0.452
Perimeter of the 30 mfL line	0.109	0.012
Horizontal * Vertical dimensions of the 20 mfL line	0.661	0.437
POWER regression model *	**0.921**	**0.849**
SHAPE		
Width of the beam pattern at 6 degrees down	0.655	0.429
Circularity of the 70 mfL line	-0.196	0.038
SHAPE regression model *	**0.776**	**0.602**
COMFORT		
ratio of brightness in the Center Distant zone to Total brightness	-0.503	0.253
Max. Difference in Brightness	-0.534	0.285
Horizontal * Vertical dimensions of the 20 mfL line	0.767	0.588
COMFORT regression model*	**0.792**	**0.628**

The R^2 values shown in Table 2 indicated the regression models explained the following amount of variance in the data for the five factors: uniformity(68%), aim(89%), power(85%), shape(60%), and comfort(63%). The F statistic for each model exceeded the critical value indicating the amount of variation in the data explained by each model was significant.

FUTURE RESEARCH ISSUES

Verification of the proposed models should be conducted by an evaluation using additional headlamp systems. Photometry data should be collected on a number of new "good" and "bad" (high and low subjective scores) headlamp beam patterns. This data would serve as input into the models to predict the factor scores. Then a subjective beam pattern evaluation should be conducted and the data compared to the predicted factor scores. Following successful verification of the models, limits on each photometric independent variable could be developed to assure that beam patterns can be created that will provide improved levels of subjective impressions on visual quality.

When completed the beam pattern design and evaluation process should be considered as a potential supplement to other photometric requirements (eg. SAE and federal standards).

ACKNOWLEDGMENTS

The research reported in this paper was conducted jointly by the Lighting Technology Group and the Human Factors Engineering and Ergonomics Department of Ford Motor Company. We gratefully acknowledge the effort of Jose Cruz of LTG for the photometric measurements on the headlamps and the help of evaluators who participated in the many night drives.

REFERENCES

1. **Jack, D. D.; O'Day S. M; and Bhise, V. D.** "Headlight Beam Pattern Evaluation Customer to Engineer to Customer", SAE paper No. 940639 (SP-1033), March, 1994

2. **Jack, D. D.; O'Day S. M; and Bhise, V. D.** "Headlight Beam Pattern Evaluation Customer to Engineer to Customer - A Continuation", SAE paper No. 950592 (SP-1033), March, 1995

3. **Bhise, V.; Farber, E.; Saunby, C.; Troell, G.; Walunas, J.; Bernstein, A.** "Modeling Vision with Headlights in a Systems Context", SAE paper No. 770238, March 1977

4. **Johnson, N. L.; Leone, F. C.** Statistics and Experimental Design in Engineering and the Physical Sciences, John Wiley & Sons, New York 1977

5. **Draper, N.; Smith, H.** Applied Regression Analysis, John Wiley & Sons, New York 1966

970908

Daytime Running Lamps with Low Power Consumption

Beate Weigelt
HELLA KG Hueck & Co.

Hans-Ulrich Rienäcker
Philips GmbH

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

During the last few years the function of daytime running light (DRL) has been applied more frequently. According to road safety experts it is an important factor in reducing daytime collisions.The number of countries allowing or requiring the DRL continues to increase. A frequent argument against DRL is the increase in fuel consumption caused by the rise in power consumption.

The development target, therefore, was a separate daytime running lamp, which would have a reduced power consumption in comparison to the low beam function frequently used.

A prototype with an approved light source was to be developed to allow immediate series production.

The increasing quality standards in the automotive industry are also challenging for the component life of the bulb. With the optimisation of the "bulb - lamp" system, new concepts may be introduced, which will comply with future component life requirements.

1 Introduction

1.1 Reliability of the Daytime Running Light

Studies (Source: 1.3) show, that the daytime running light increases road safety. Since the introduction of the DRL in several countries, a reduction of approx. 10 % of the daytime collisions was noted.

It is noticeable, that particularly northern countries, where the daytime light periods are short, and twilight, rain and snow are frequent, have shown interest in this subject.
As a result, the DRL has become mandatory in Sweden, Norway, Finland, Denmark - but also in Canada since 1989, in Poland and Hungary in the 1990s.

In the USA, Australia and Great Britain, DRL is permitted, but not mandatory.

Since 1992 the European Union (EU) have a uniform regulation stipulating that the daytime running light is permitted as a signal lamp.

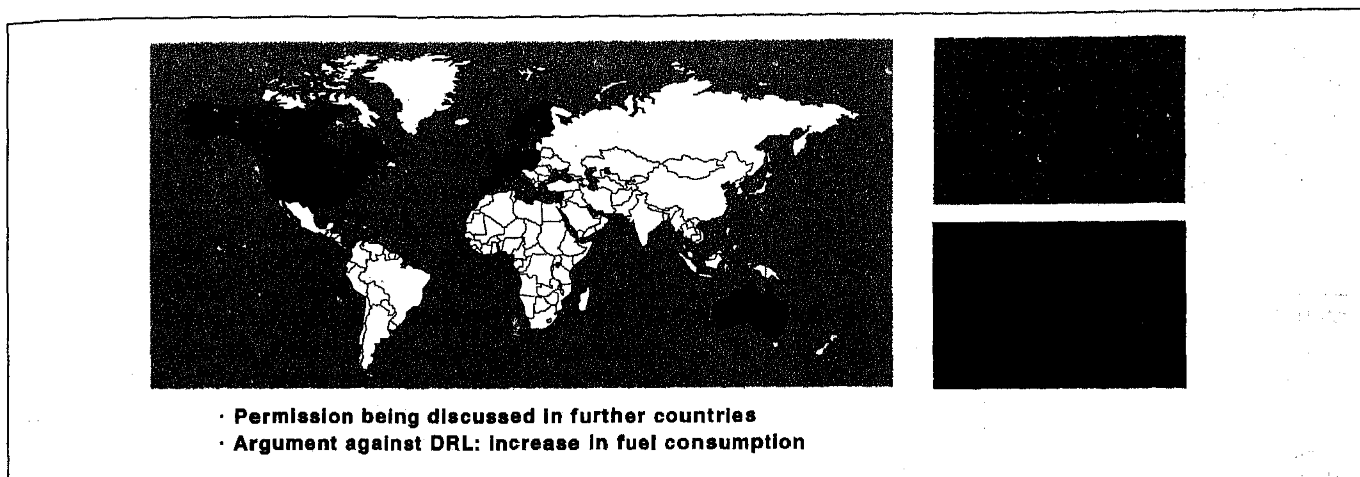


Fig. 1 Daytime running light - current situation

One of the most frequent arguments against DRL is the increase in fuel consumption. The minimum values given in this context are 0.1 l per 100 km with switched-on headlamps. For all of Europe an additional consumption of approx. 288 million l fuel is given per year (Source: 1.5). Therefore, special attention needs to be paid to the energy requirement of this continuous function.

In this project, the use of light sources with drastically reduced power consumption was to be realised. The bulb life and possible optimisations were analysed for this purpose. The example of a separate daytime running lamp allowing for series use in EU and US was the result.

1.2 Definitions and Regulations

According to ECE-R No. 87, the DRL is defined as follows:

"Lamp for daytime running light" is a lamp pointed in the driving direction used to make the vehicle more visible, when it is driven during daytime light.

In national definitions (e.g. in Sweden) this special lamp function is also called **"perception lamp"**.

DRL may be generated by the low beam function, a reduced high beam, front fog lamp, front direction indicator or clearance lamps or by special additional daytime running lamps (depending on the regulations).

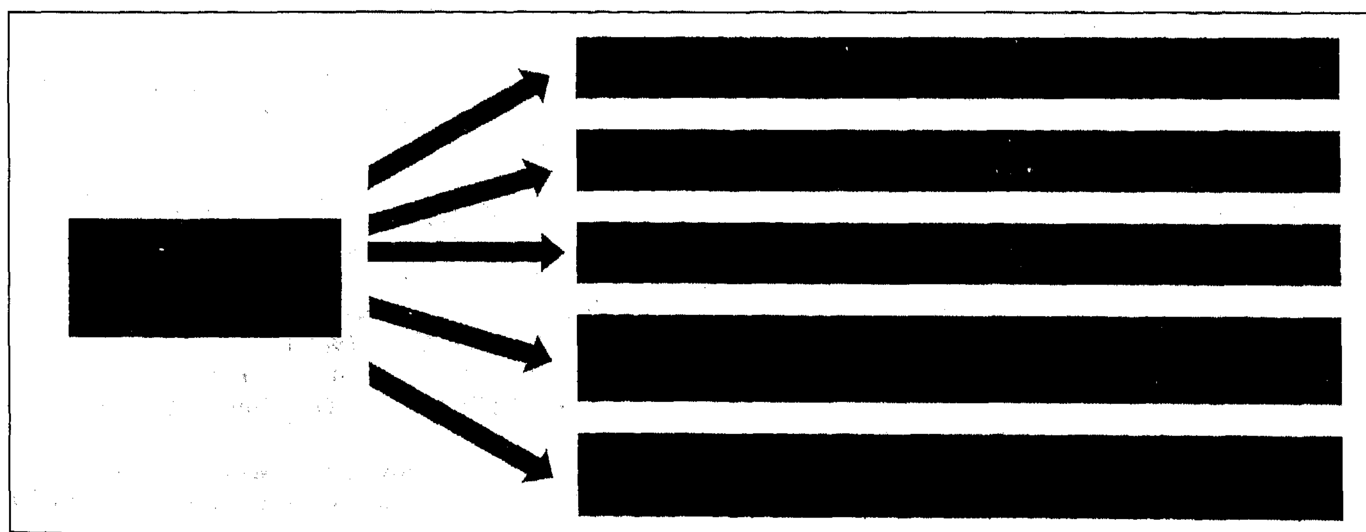


Fig. 2 Daytime running light - Existing solutions

In Europe the daytime running lamp is permitted within the installation directive ECE-R 48 (2/1995), i.e. the responsibility lies with the vehicle manufacturer. In addition to the various national regulations, ECE-R No. 87 is accepted. (see fig. 3).

Country: **Europe**		Light intensity **[cd]**			**Daytime running lamp - white**		
Basis of requirement:					**ECE-R 87**		
	-20°	-10°	-5°	0°	+5°	+10°	+20°
+15°							
+10°			80	80	80		
+5°	40	80		280		80	40
0°	100	280	360	400	360	280	100
-5°	40	80		280		80	40
-10°			80	80	80		
-15°							
Remark: permitted maximum 800 cd							

Fig. 3 Regulation on daytime running lamp - ECE-R 87

Canada's light distribution requirement, in contrast, is for a narrow light strip according to CMVSS 108. Europe as well as Canada have a minimum requirement of ≥ 40 cm² with regard to the illuminated surface.

Country: **Canada**		Light intensity **[cd]**			**Daytime running lamp, white to amber**		
Basis of the requirement::					**CMVSS 108**		
	-20°	-10°	-5°	0°	+5°	+10°	+20°
+10°							
+5°							
0°		250	250	500	250	250	
Remark: permitted maximum 1200 cd							

Fig. 4 Daytime running lamp requirements - Canada

1.3 Studies

The articles listed below refer to the findings about the reduction of daytime collisions since the introduction of the daytime running light:

- "Daytime Running Light- Reduce Daytime Collisons" (PAL-Symposium 1995) by Kåre Rumar, Swedish National Road Administration

The following table documents the reduction of daytime collisions since the introduction of the daytime running light.

Year	Test fleet / Area	Accident reduction
1962	Greyhound Bus	-10 %
1965	US truck fleet	- 44 %
1968	NY car fleet	- 18 %
1970	Chicago taxi fleet	- 12 %
1976	**Nation Finland**	**- 21 %**
1979	US truck fleet	- 7 %
1980	Nation Schweden	- 11 %
1985	US company fleet	- 7 %
1986	**Nation Norway**	**- 15 %**
1989	Canadian company fleet	- 12 %
1990	**Nation Denmark**	**0 % (left turn - 37 %)**
1991	Cars Israel (Campaign)	+ 19 %
1993	**Nation Norway**	**0 % (- 15 % summertime)**
1994	**Nation Canada**	**- 11 % (oncoming)**

Fig. 5 Reduction of daytime collisions due to application of the DRL [Kåre Rumar]

- "Study of Daytime Running Light - TP 6716 E" (Canada 1985)

- "Effectiveness of Daytime Running Lights in Canada - TP 12298 (E)" von H. Arosa, D. Collard, G. Robbins, E. R. Welbourne, J. G. White (Dez. 94)

This analysis shows, that with the application of the DRL, collisions have been reduced by 8,3 %.

1.4 Development of a separate lamp

In the daytime running light version used primarily - the combination with the low beam - a long life bulb is required. This means that the filament is larger, the reach of the beam is reduced and the glare caused by straylight can increase. Therefore, the use of a separate daytime running light function is to be preferred. One advantage of a single function lamp is a more homogeneous signal when compared to DRL realised with headlamps

For the development of the prototype, the light source and the lamp were considered together. The target was the fulfilment of the European and American lighting and surface requirements.

The power input should be as small as possible when compared to the usual H4/HB2 bulb with its 68 W.

The following elements were used:

• Lamp

One of the targets was to develop a lamp which meets all the requirements of the various countries.

The reflector is a rotationally symmetrical spherical reflector with metallised prisms. The minimum requirements for the illuminated surfaces of $\geq$ 40 cm² are fulfilled (Europe, FMVSS 108).

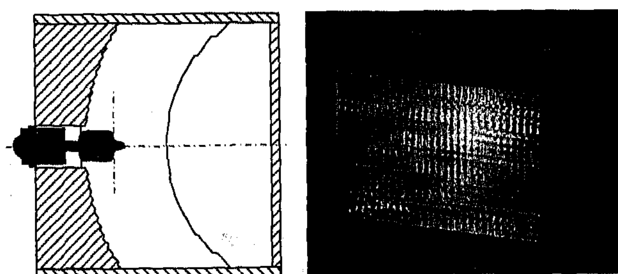


Fig. 6 DRL / schematic drawing and illuminated example with stepped spherical reflector

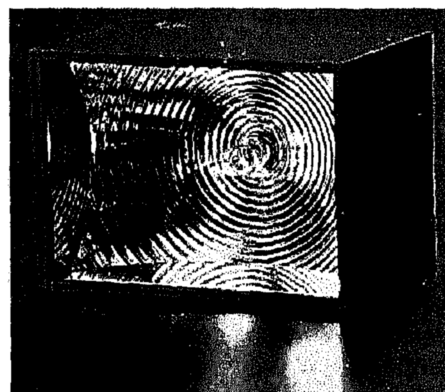


Fig. 7 View of the stepped spherical reflector without lens

A particular challenge was the ECE light distribution requirement. Here the cd value at $\pm$ 20° horizontally is 25 % of the max. value.

Alternatively, a paraboloid was considered as a reflector shape. This example is deeper and more efficient.

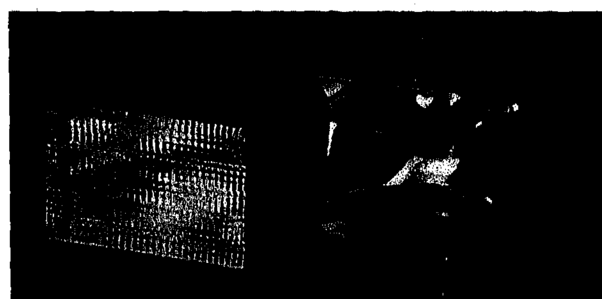


Fig. 8 Spherical reflector with lens (left) Paraboloid without lens (right)

• Light bulb

To support an immediate series application with type-approved bulbs, bulbs which are permitted according to ECE-R 37 were taken into account. Upon consideration of appropriate types and an adequate power consumption, the H6W bulb with a light flux of 125 lm was chosen. Its properties are discussed in point 1.6.

1.5 Power consumption

The additional power consumption of the DRL is one of the major arguments held against this function. The values given lie between 0.1 l and 0.2 l per 100 W of electrical power. A Canadian study of 1995 [Source: Preliminary economic evaluation of the costs and benefits of the DRL regulation - TP 12517 (E)] quotes a typical excess consumption of 0.19 l per 100 km at an average speed of 40 km/h. Power consumption lies between 40 W and 200 W. Vehicle manufacturers usually speak of 0.1 l of excess consumption.

According to the German Government, an additional 288 million l of fuel per year would be consumed in Europe, should all passenger vehicles use DRL during daytime driving [Source: Verkehrsnachrichten 3/96 - Ministry of Traffic].

In the Hella analysis, a relative reduction in consumption is the decisive parameter. The application of the halogen bulb **H6W** having a power consumption of 7 W would result in a **90 %** reduction when compared with the 68 W **H4/HB2** bulb. Looking at the standard bulb **P21W/1156** for lamps with a consumption of 25W, the savings would still amount to **72 %**. - In other words, a major reduction in primary energy consumption can be achieved.

1.6 DRL bulb life

Consideration of the bulb life is a primary concern especially with regard to warranty periods. In the US, the current warranty period is 3 years, however, the target should be 4 years.

Based on a driving performance of 20,000 km per year [Source: Highway Statistic 1994], an average speed of 50 km/h and an average use of the headlamps of 9.2 h per 1000 km [Source: Use of Car Lighting, Mercedes 1995], and taking into account a warranty time of 4 years, we reach a calculated value of 864 illuminated hours for the DRL function.

Until today it is normal practice to characterise the bulb life by means of the B_3 and T_C values. These two values have proven to be useful since they accurately characterise the failure well, while keeping the number of required measurements reasonable. Smaller values (e. g. $B_{0.1}$) would require a far larger number of bulbs to be tested in order to achieve a high level of confidence. Nevertheless, these data should allow the derivation of B_3 and T_C values from a theoretical $B_{0.1\ theor.}$ value. The characterics can be determined in line with these thoughts, if the $B_{0.1\ theor.}$ value is set ≤ the operation time.

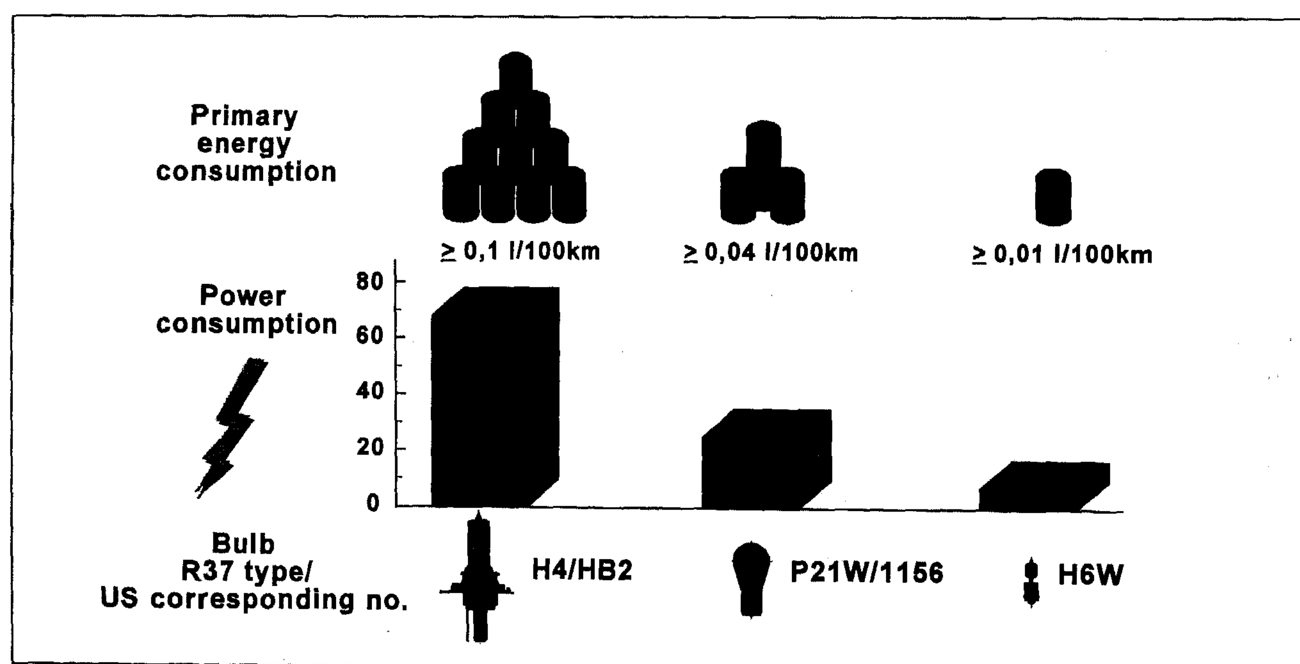


Fig. 9 Power consumption of the DRL function
Comparison of various bulb types

Considering a typical slope of the Weibull curve, we can calculate the corresponding B_3 and T_C values using the following formula:

$$t = \sqrt[b]{\ln(1-F(t))/\ln(1-F(T))} * T$$

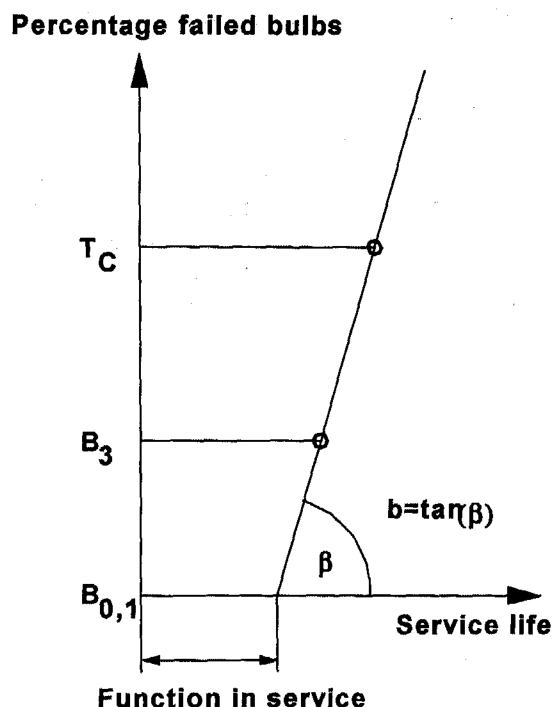


Fig. 10 Model of a *Weibull* graph

t = Bulb life to be calculated

b = Slope of *Weibull* curve

F(t) = Desired cumulative probability B%/100

T = Known bulb life

F(T) = Known cumulative probability B%/100

The result of this calculation are values of B_3 = 2200 h and T_C = 5300 h.

Obviously, if the luminous flux losses are not to be excessive and the performance over the period of the bulb life is to be acceptable, such life values can only be achieved with high pressure halogen bulbs.

The H6W, being a halogen miniature bulb with special design features offers the best conditions for use in DRL.

- Shock resistance
  - Solid inner mount
  - Short clamping of the filament
- Best optical qualities
  - Tight tolerances
- Service life
  - High pressure xenon filling

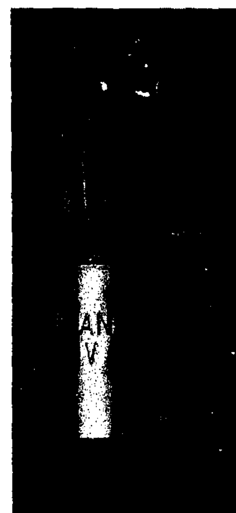


Fig. 11 Halogen miniature bulb HxW

Considering the service life we must notice, that this in accordance with ECE-R37 authorized bulb is specified in his operating range for other applications. If the obtainable values for the service life are compared with these which are required after above considerations, we notice, that the requirements are fulfilled only to approximately 12 %. It is obvious, that development work is required on this subject in the near future.

Values at 14 V	B_3 [h]	T_c [h]	$B_{0,1}$ [h] theoret.
H6W	210	480	100
HxW	2200	5300	864

Fig. 12 Comparision of H6W and HxW bulb, fulfilling the life requirements

To allow the immediate introduction of the DRL with low energy consumption, the system must be designed for the H6W, but at the same time, a bulb needs to be developed, which is comparable to the H6W in its photometrical properties. This would allow a change to a longlife bulb without having to change the lamp.

Based on an identical component volume, the same filament alignment and length, we assume for the long life bulb the following changes in parameters related to the H6W data:

at 13.5 V	Value	Differences [%]
Luminous flux	133 lm	+ 7 %
Power consumption	10.5W	+ 50 %
Filament temperature		- 7 %
Filament diameter		+ 80 %

Fig. 13 Parameter differences between H6W and HxW

1.7 Summary

The prototype of a separate DRL equipped with halogen bulb H6W indicates an imminent series application for the countries where its use is permitted.

The savings achieved in power consumption amount to 90 % of that of the H4/HB2 bulb.

Increasing bulb life demands require further optimisation on the manufacturer's side. The consideration of the power consumption will be of increasing importance for future developments - particularly with view to the US requirement that approx. 10 % of the produced vehicle will have "0-emission vehicles" or electrics from the year 2003 onwards.

970910

Effects of Large-Radius Convex Rearview Mirrors on Driver Perception

Michael J. Flannagan, Michael Sivak, and Eric C. Traube
University of Michigan Transportation Research Institute

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

The U.S. currently requires that rearview mirrors installed as original equipment in the center and driver-side positions be flat. There has recently been interest in using nonplanar mirrors in those positions, including possibly mirrors with large radii (over 2 m). This has provided additional motivation to understand the effects of mirror curvature on drivers' perceptions of distance and speed. This paper addresses this issue by (1) reviewing the concepts from perceptual theory that are most relevant to predicting and understanding how drivers judge distance in nonplanar rearview mirrors, and (2) reviewing the past empirical studies that have manipulated mirror curvature and measured some aspect of distance perception.

The effects of mirror curvature on cues for distance perception do not lead to simple predictions. The most obvious model is one based on visual angle, according to which convex mirrors should generally lead to overestimation of distances. But convex mirrors affect other perceptual cues (including vergence and accommodation) in ways that lead to predictions of underestimation.

Empirical investigations of the effects of mirror curvature have produced a strong consensus that convex mirrors cause overestimation of distance, but several factors can moderate or compensate for that effect. All quantitative studies of the effects of the radius of convex mirrors have demonstrated less overestimation of distance than predicted by the visual-angle model. Shorter-radius (more strongly curved) mirrors generally lead to greater overestimation of distance. Previous studies have examined the effects of mirror radius up to 2 m. There is strong evidence that 2-m mirrors still cause substantial overestimation, and little indication that reductions in overestimation have asymptoted at that radius.

New empirical efforts to study the effects of larger-radius mirrors, beyond 2 m, could both contribute to basic understanding of how mirror curvature affects distance perception, and provide practical information about the possible benefits of using large-radius rearview mirrors.

Introduction

A major issue in optimizing the safety of rearview mirrors is how to provide a field of view that is large enough to eliminate blind zones near a vehicle without distorting perceptual information about the speed and distance of rearward vehicles. Planar mirrors provide undistorted information about speed and distance, but relatively small fields of view. Convex mirrors can provide much larger fields of view, but they alter information about the size and distance of objects (Seeser, 1974). The most important effect is probably a reduction in image size that may lead to an increase in the perceived distances to objects. This has led to a concern that drivers who use convex mirrors may misjudge the locations of other vehicles. It is therefore difficult to provide rearview mirror fields of view that are acceptable both in terms of quantity and quality.

Certain aspheric mirror designs may offer a good solution to this dilemma. Aspheric mirrors, which typically combine an inboard section that is relatively gently curved with an outboard section that is progressively more strongly curved, provide a large field of view primarily because of the more strongly curved outboard section. The images of vehicles seen in the outboard section are strongly distorted. However, vehicles that appear in that part of the mirror will be close enough to the observer's vehicle that, it can be argued, the driver merely has to detect their presence to know that a maneuver such as a lane change is not safe. Quantitative judgments about their speed and distance are moot. Vehicles that are some distance behind the observer's vehicle will be imaged in the inboard, more gently curved section of the aspheric mirror, and their images will therefore be less distorted than they would be in a spherical convex mirror with an equal total field of view.

Aspheric mirrors have been used on the driver's side in Europe for a number of years (Pilhall, 1981). They meet European regulations as long as they incorporate a spherical convex section that by itself meets the regulatory requirements. The aspheric section is then considered an allowable, unregulated addition. A certain type of aspheric mirror would be permitted on the driver's side in the U.S. under a similar rationale. However, because U.S. Federal

Motor Vehicle Safety Standard (FMVSS) 111 specifies that the driver-side mirror must have "unit magnification," an aspheric mirror in the U.S. would have to combine an aspheric section with a flat section, rather than with a spherical convex section as in Europe. This constraint introduces a manufacturing difficulty (e.g., Erickson & Allen, 1989). It is difficult, using manufacturing processes that are conventional for automotive mirrors, to reliably produce a mirror that combines flat and aspheric sections.

Some manufacturers have suggested exploring whether it is possible to relax the constraint in FMVSS 111 for "unit magnification" on the driver's side, thus easing manufacturing problems, without introducing perceptually significant distortions of speed and distance information. This might be done by allowing spherical convex mirrors with very large radii to be considered effectively "unit magnification." However, such an approach would depend on clear evidence about the perceptual effects of large-radius convex mirrors. The unit-magnification constraint is based primarily on a concern for providing the driver with undistorted information about speed and distance. Planar mirrors (with effectively infinite radius of curvature) do provide such information, but it is unclear what range of radius of curvature might provide virtually the same information. The range of curvature currently permitted on the passenger side in the U.S. (35 to 65 inches [0.89 to 1.65 m]) is too low to include mirrors that might be considered to have virtually unit magnification. New studies of driver perception may be needed to establish how much larger the radius of curvature must be to allay concerns about distortions of speed and distance perception.

This paper reviews what is known about the effects of rearview mirror curvature on drivers' perceptions of distance. This is motivated both by the general importance of understanding those effects, and by the current specific interest in the effects of large-radius mirrors. We first review theoretical considerations, and then past empirical work on the issue.

Theoretical Considerations

The aspects of convex-mirror images that are most relevant to perception of distance are illustrated in Figure 1. In this example, a simple convex rearview mirror with a radius of 1.4 m is installed on the driver side of a right-hand-drive car. The distance from the driver's eyes (assumed here for simplicity to be at single location between what in reality would be two laterally displaced eye positions) to the mirror is 0.6 m. An object with a projected width of 0.5 m is present 50 m behind the mirror location. The driver sees a minified virtual image of the rearward object located 0.7 m behind the mirror. The minification factor is 0.014 and the virtual image therefore has a width of 7 mm.

These circumstances lead to conflicting perceptual distance cues. The distance cue that is probably most important in this situation is the visual angle (as viewed from the driver's eye position) of the virtual image seen in the mirror. If the rearview mirror were planar, that angle would be 0.57 degrees, but in the 1.4-m convex mirror the combination of changes in size and location of the virtual image produces a visual angle of 0.31 degrees. In order to form such an image, assuming that the object did not actually change size, the object would have to be 93.5 m from the driver's eyes. Assuming a simple model in which the driver's perception of distance is based entirely on visual angle (and accurate memory for the actual sizes of familiar objects, such as the widths of vehicles) the object should be perceived as being 93.5 m away.

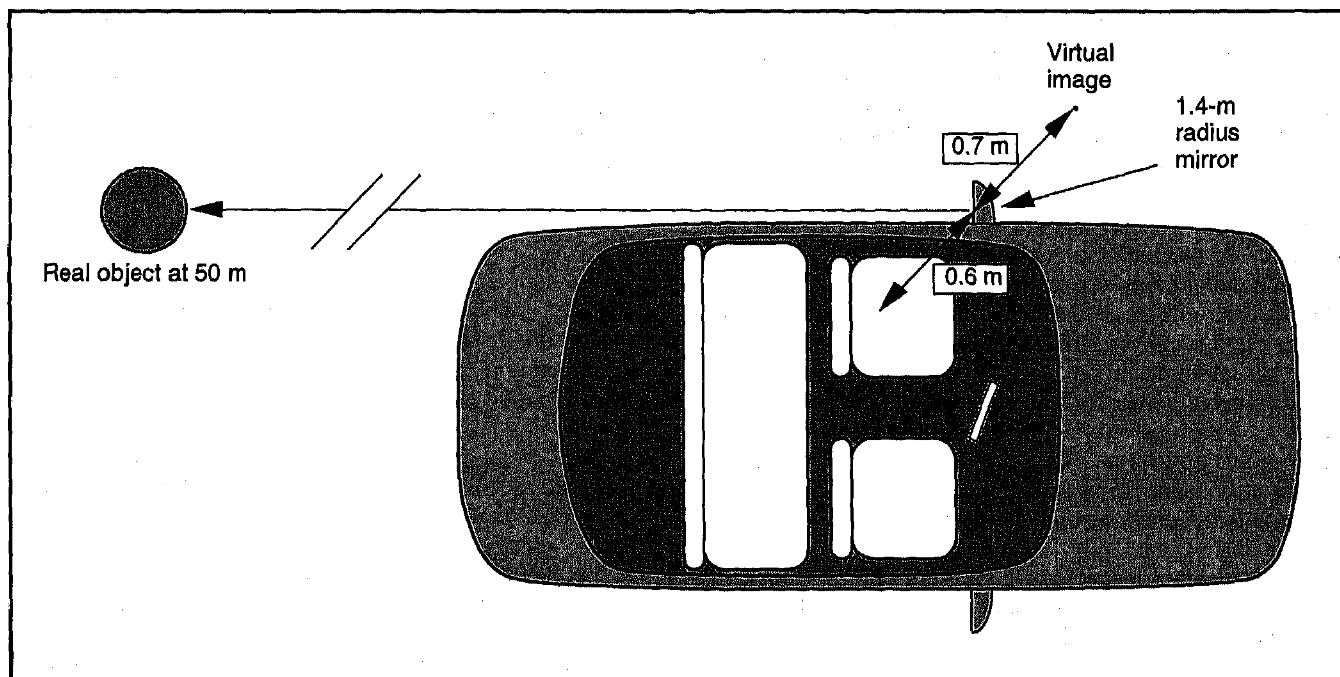


Figure 1. The locations and sizes of an object seen by means of a convex driver-side mirror, and of a virtual image of that object (approximately to scale except for the break indicated in the 50-m distance).

Figure 2 shows the predictions of such a model for a range of mirror radii (0.5 m to 10.0 m) and for three different eye-to-mirror distances (0.55, 0.65, and 0.75 m), for a single object-to-mirror distance (20 m). (Object-to-mirror distance affects predicted distance perception, but to a much smaller degree than the other two variables.) According to this model, even for a mirror with a 10-m radius of curvature the overestimation of distance would be nontrivial (overestimation by a factor of 1.13 for the 0.65-m eye-to-mirror distance).

The visual-angle model thus predicts that convex mirrors will lead to overestimation of distance, even at large radii. However, the distortion of certain other distance cues by the convex mirror might be expected to cause an underestimation of distance.

In the example in Figure 1, the virtual image is located 1.3 m from the driver's eyes. Because of this, in order to see the image as a single, fused image in proper focus the driver's eyes must actually be converged and accommodated at 1.3 m. In many situations, the binocular cues of vergence and binocular disparity, as well as accommodation, have powerful effects on distance perception. In this case, a simple model that predicts distance perception on the basis of either binocular cues or accommodation would predict that the object would be seen as extremely close (and implausibly small). Putting aside for a moment the common experience that objects seen in convex mirrors are indeed closer (at least to some extent) than they appear, it is worth considering how all of these cues might contribute to the driver's actual perception of distance. From a purely theoretical perspective, it is not clear how we should expect the conflicting cues to be resolved.

The actual distance and the two conflicting predictions for the perceived distance of the object are summarized in Table 1. There are at least two reasons that one might expect a rational observer to give greater weight to the visual-angle prediction. First, the prediction based on vergence or accommodation is implausibly short for objects that are normally seen in the driving environment. It is possible to imagine that, through either conscious or unconscious inference, a driver might simply reject the evidence provided by vergence and accommodation. Second, the distances about which it is most important to make judgments while driving are mostly too long (say 10 m or more) to be within the relatively short range in which vergence and accommodation are likely to be useful distance cues. It may be that drivers have learned to generally disregard those cues while driving.

On the other hand, it does not seem safe to assume that vergence and accommodation play no role in distance perception in this context. It may be that a driver's perception of distance in a convex mirror is determined by some combination of these cues. Empirical results, as reviewed in the following section, suggest that something like this may indeed be the case. Although convex mirrors do in fact make distances appear somewhat longer than they actually are, the predictions of the visual-angle model have been violated in every quantitative study of the effects of nonplanar mirrors, and always in the direction that the visual-angle model predicts greater overestimation of distance than actually seems to occur. Although the reasons for this are not fully clear, the possibility of a moderating influence of conflicting distance cues seems worth considering.

Table 1
Actual distance, and two predictions for perceived distance, from the driver's eyes to the rearward object in Figure 1.

	Distance (m)
Actual distance	50.6
Predicted (visual angle)	93.5
Predicted (vergence or accommodation)	1.3

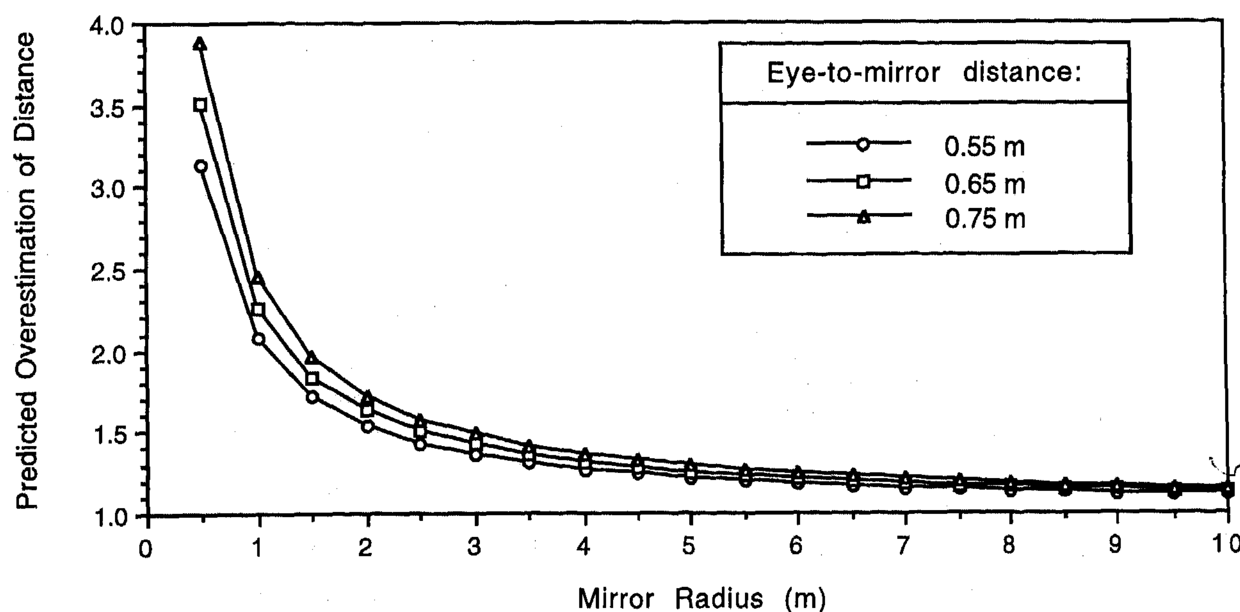


Figure 2. Predicted overestimation of distance (perceived distance divided by actual distance) based on visual angle, as a function of convex mirror radius, for three eye-to-mirror distances. In all cases, the distance from the mirror to the object is 20 m.

Summary of Empirical Work

Table 2 lists 10 studies in which the radii of convex mirrors have been varied, and the effects of that manipulation on distance perception (and sometimes on other aspects of driver performance) have been observed. Following an earlier review by Flannagan (1988), the studies have been divided into two sets on the basis of what is probably the most important distinction among the tasks that they required subjects to perform. One set of studies (labeled "distance-perception" studies in Table 2) involved relatively direct measures of subjects' perceptions of distance. These studies required subjects to make judgments about the distances of vehicles seen in mirrors of various radii, but with no connection to a specific driving task. The second set (labeled "lane-change" studies) required subjects to make judgments about how they would perform a specific maneuver that depends on perception of distance and speed: changing lanes at the last possible safe moment into the path of an overtaking vehicle seen in a rearview mirror. The implementation of the lane-change task varied from laboratory studies using stimuli presented on film, where no actual maneuver was possible (Fisher & Galer, 1984), to field studies in controlled traffic in which subjects actually performed the lane-change maneuver (Mortimer, 1971).

Our purpose here was to extract from as many of these studies as possible a quantitative estimate of the effect of mirror radius on distance perception. The variety of experimental methods, and of data-analysis and reporting styles, that were used in these studies makes it a nontrivial task to assemble results in a unified format. However, the task was made tractable because our goals were somewhat limited. In each case, we tried to select data that would reflect the effect of mirror radius on distance perception in its simplest form, without possible influences from mechanisms such as learning a new calibration between distance and the stimuli seen in the convex mirror, or adopting strategies that reduced reliance on the convex mirror. So, for example, from the study by Flannagan et al. (1996) we selected data that were collected prior to a set of training trials, in which subjects had an opportunity to recalibrate their judgments by comparing views through a convex mirror with direct views to the rear of their vehicle. Similarly, in the study by Mortimer (1971) we selected data from a condition in which subjects had to make judgments using only a convex exterior mirror, rather than another condition in that same study that allowed them to use both a convex exterior mirror and a planar interior mirror.

For each study, we quantified the effect of mirror radius in terms of an index that compared perceived distance while using one or more nonplanar mirrors to perceived distance while using a planar mirror. This index is therefore not strictly a measure of how subjects' performance with nonplanar mirrors compared to a standard of objective correctness, but rather how it compared to their performance of the same task while using a planar mirror. For data from the lane-change studies, we calculated effects on perceived distance by assuming that subjects perceived criterion distances in nonplanar mirrors to be equal to the criterion distances that they selected in planar mirrors.

Table 2
Studies in which the perceptual effects of varying mirror radius have been measured.

Distance-perception studies
Burger, Mulholland, Smith, & Sharkey (1980)[1]
Flannagan, Sivak, & Traube (1996)
Rowland, Silver, Volinsky, Behrman, Nichols, & Clisham (1970)
Smith, Bardales, & Burger (1978)
Sugiura & Kimura (1978)
Lane-change studies
Bowles (1969)
Burger, Mulholland, Smith, & Sharkey (1980)[1]
Fisher & Galer (1984)
Mortimer (1971)
Mortimer & Jorgeson (1974)
Walraven & Michon (1969)

[1] This study involved both types of task.

The following notes discuss each of the studies listed in Table 2, indicating whether parametric data on the effect of mirror radius on distance perception could be extracted, and identifying the specific sources of the data that we did select.

Burger, Mulholland, Smith, & Sharkey (1980) This was a major study that involved several parts, including variations of mirror radius in both a direct assessment of distance perception and in a lane-change task. However, a small number of subjects were given extensive training with a large variety of mirrors, and the data are not reported in a way that would allow separating performance early in training from performance later in training. We have therefore not used data from this study (or from subsequent experiments performed by the same researchers on substantially the same group of subjects when the subjects were even more experienced: Burger, Mulholland, Smith, Sharkey, & Bardales, 1980).

Flannagan, Sivak, & Traube (1996) The task was magnitude estimation of the distance to a vehicle seen in an exterior rearview mirror. Subjects were given moderate amounts of experience with nonplanar mirrors. The data selected here are from prior to the learning phase of the experiment. The data are from their Figure 3, page 9, averaged over five actual distances.

Rowland, Silver, Volinsky, Behrman, Nichols, & Clisham (1970) This report covers a set of 15 experiments on planar and convex mirrors, but the treatment of data is almost entirely in terms of average absolute error. The report therefore cannot be used to estimate whether changes in mirror radius cause perception of distance to be generally longer or shorter.

Smith, Bardales, & Burger (1978) Subjects estimated the distance of a vehicle that was seen approaching from the rear (in a laboratory setting, using filmed stimuli). The presentation of data is primarily in terms of average absolute error, which does not allow for an estimate of a possible general increase or decrease in perceived distance as a function of mirror radius. A limited summary of average signed error is presented in their Figure 20, page 54. The data appear to be noisy, but no inferential statistics are reported. We used data from that figure for session 4 (when variable error was lowest) to estimate an effect on distance perception, assuming that their Figure 20 presents data collapsed over all the actual distances that were involved in the study.

Sugiura & Kimura (1978) Subjects indicated when a vehicle, seen in a planar mirror or one of several nonplanar mirrors, reached a criterion distance that had been shown to them in a planar mirror. The data used here are from the "Day" condition in their Figure 2, page 3.

Bowles (1969) Subjects performed a version of the lane-change task. We used data from the right panel of the report's Figure 1 (no page number). These data are averaged over three closing velocities, for a condition in which the mirrors were mounted on the door. (There were also data from a fender-mount condition, which we have not used because the eye-to-mirror distance involved is not typical of more recent practice.)

Fisher & Galer (1984) Subjects performed the lane-change task. The data used here are from their Experiment 1, the "true minimum safety margin" from their Table 1, page 1217, averaged over three closing speeds.

Mortimer (1971) Subjects actually performed the lane-change task in traffic. The data we used are from his Table 1, page 4, for the "exterior mirror only" condition. His experiment also included a condition in which exterior mirrors (of varying radius) were provided along with an interior mirror that was always planar.

Mortimer & Jorgeson (1974) This was similar to the earlier study by Mortimer (1971). A major difference for our present purposes is that it did not include a condition without a planar mirror. We have therefore not used data from this study.

Walraven & Michon (1969) Subjects performed the lane-change task. The data we have used are from their Figure 8. They are combined over two different groups of subjects and three closing speeds.

Summary across studies

Figure 3 shows a summary of results from the seven studies from which we extracted parametric estimates of the effect of mirror radius on distance perception. In all cases, the overestimation index has been calculated relative to performance with a planar mirror in the same study and in the same circumstances.

There is a strong consensus among these studies that convex mirrors result in overestimation of distance, and that the degree of overestimation increases with decreasing radius of curvature. However, there is a wide range in degree of overestimation across the studies. This is particularly interesting given that the data shown here are in many cases subsets of the data from their respective studies, selected so that the conditions represented are relatively homogeneous across studies.

The distinction that we made earlier between distance-perception and lane-change studies is reflected in Figure 3 by the use of solid lines and symbols (for lane-change studies) and dashed lines and open symbols (for distance-perception studies). There is no obvious difference between these two types of studies in the extent of overestimation that they demonstrate.

These studies provide evidence about the effects of mirror radius up to 2.0 m. Although there is strong evidence for a decrease in overestimation with increasing radius, this change does not seem to be asymptoting within the range covered.

Comparisons of the effects of mirror radius shown in Figure 3 to predictions based on visual angle are shown in Figure 4. The predictions were derived as in our earlier treatment of the situation represented in Figure 1, and as discussed in more detail by Platzer (1995). Although mirror radii were always well specified in the reports summarized here, in many cases the published descriptions of method did not indicate exact values of the other two parameters that are needed to calculate predictions from the visual-angle model: eye-to-mirror distance and object-to-mirror distance. In such cases we used 0.75 m and 20 m as default values for those two parameters, respectively.

In every case, the predictions of the visual-angle model are violated, always in the direction that there is less overestimation than predicted. In most cases the discrepancy is large. As discussed earlier, the visual-angle model does not reflect all the theoretically plausible effects of convex mirrors on distance perception. Accommodation and vergence cues suggest that convex mirrors should cause strong underestimation of distance, and a moderating effect of these cues might be at least a partial explanation for the failure of the visual-angle model.

Summary and Conclusions

Previous data offer empirical evidence about the effects of mirror radius up to 2.0 m. The evidence indicates that distances are overestimated, and that overestimation is still substantial at a radius of 2.0 m.

The simplest, and perhaps to many people the most compelling, model of the effects of mirror radius—the visual-angle model—predicts that convex mirrors will cause overestimation of distance, and that the overestimation will be nontrivial even when the radius is as large as 10 m (for which the overestimation factor is predicted to be about 1.13). However, the visual-angle model is strongly violated in all existing results.

Given the current practical interest in mirrors with radii above 2.0 m, the lack of empirical evidence above 2.0 m, and the theoretical uncertainty caused by the substantial violations of the visual-angle model, it seems worthwhile to conduct further empirical work using mirrors with radii over 2.0 m. At a more theoretical level, the lack of a satisfying account of the effects of mirror radius suggests that more work should be done to evaluate the predictions of the visual-angle model, for example with respect to the relative importance of eye-to-mirror and object-to-mirror distances.

Acknowledgments

We wish to thank Ichikoh Industries, Ltd. for their generous support of this research.

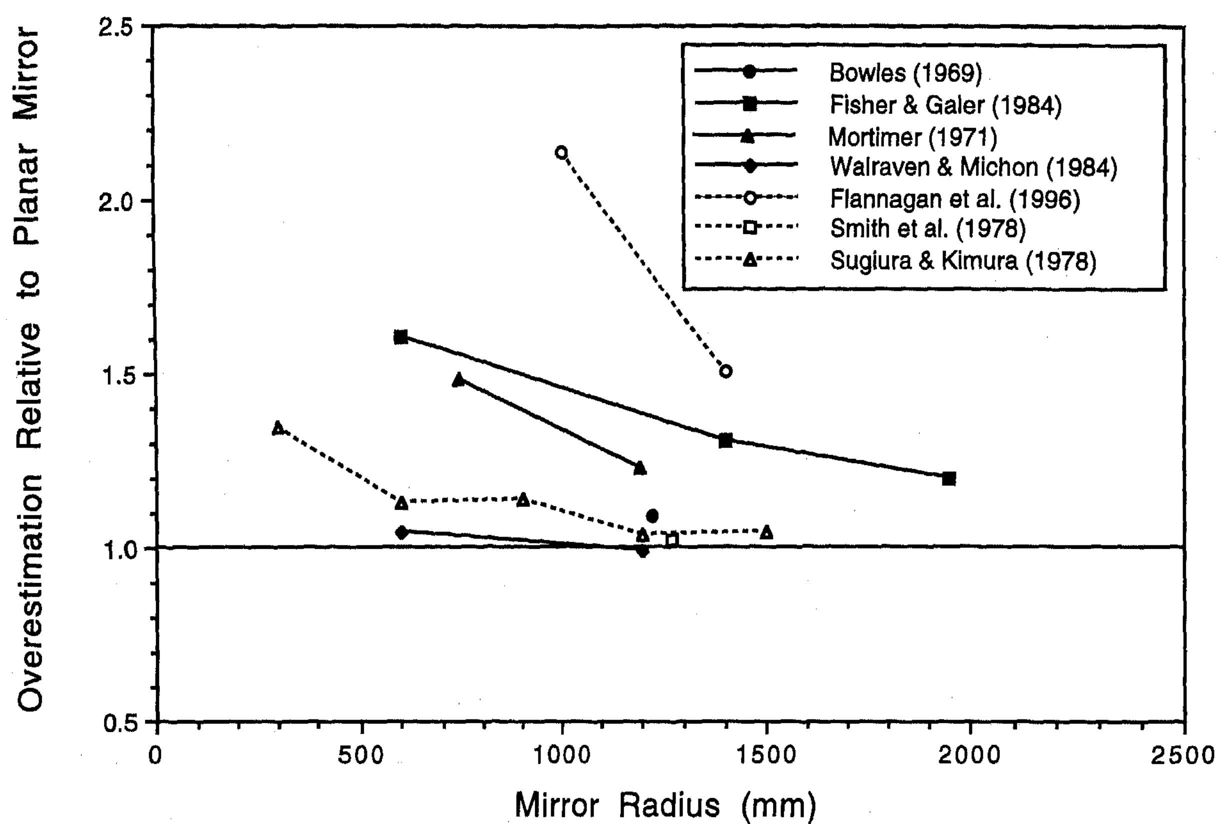


Figure 3. Effect of mirror radius on estimation of distance. Distance estimates in each study are shown relative to performance with a planar mirror in the same study. The horizontal line at 1.0 represents performance that would be the same as with a planar mirror. Data from lane-change studies are shown with solid lines and symbols, and data from distance-perception studies are shown with dashed lines and open symbols.

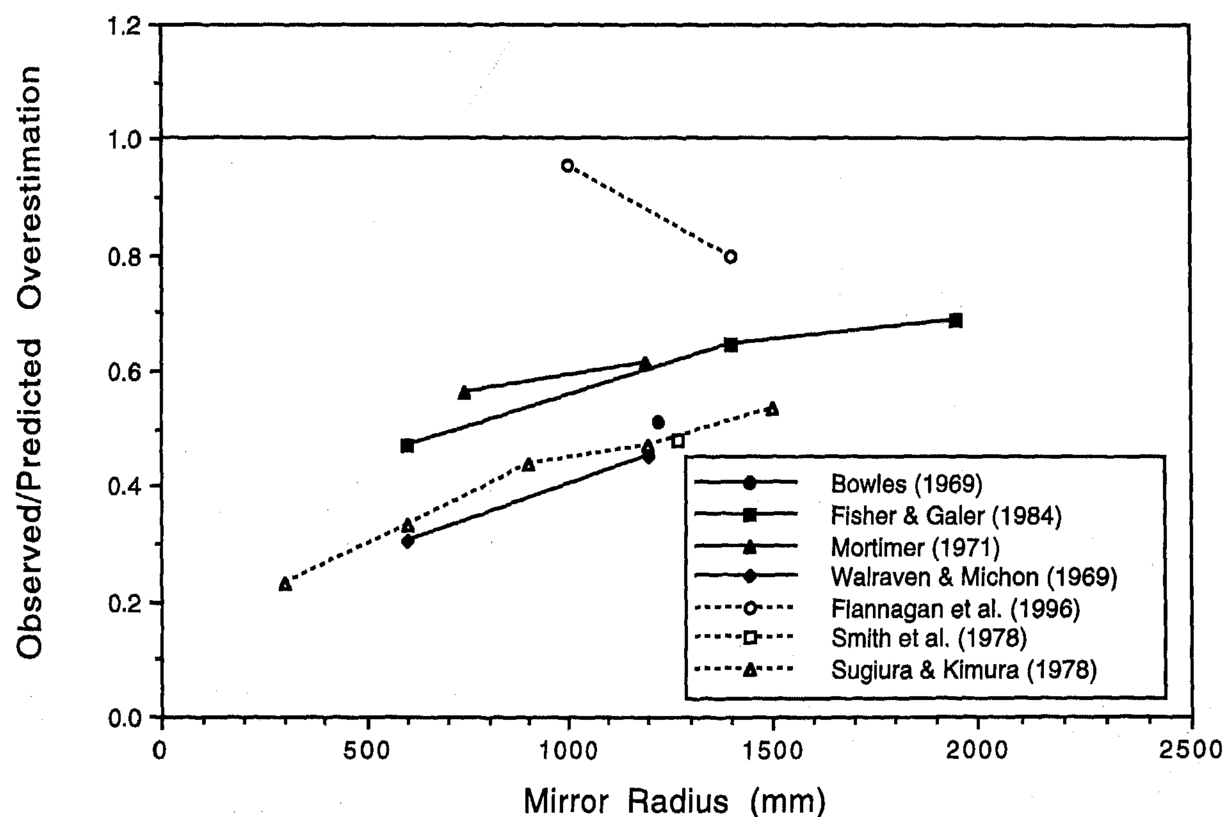


Figure 4. The data from Figure 3 plotted as ratios of observed distance overestimation relative to predicted overestimation based on the visual-angle model. In every study, observed levels of overestimation were lower than predicted, usually by a large margin.

References

Bowles, T. S. (1969). Motorway overtaking with four types of exterior rear view mirror. In *International Symposium on Man-Machine Systems, Volume 2, Transport systems and vehicle control (IEEE Conference Record no. 69C58-MMS)* Institute of Electrical and Electronic Engineers.

Burger, W. J., Mulholland, M. U., Smith, R. L., & Sharkey, T. J. (1980). *Passenger vehicle, light truck and van convex mirror optimization and evaluation studies: Vol. 1: Convex mirror optimization* (DOT HS 805 695). Washington, D.C.: Department of Transportation, National Highway Traffic Safety Administration.

Burger, W. J., Mulholland, M. U., Smith, R. L., Sharkey, T. J., & Bardales, M. C. (1980). *Passenger vehicle, light truck and van convex mirror optimization and evaluation studies: Vol. 2: Evaluation of alternative mirror configurations* (DOT HS 805 778). Washington, D.C.: Department of Transportation, National Highway Traffic Safety Administration.

Erickson, G. A., & Allen, P. J. (1989). *Vision improvement through computerized design and fabrication of controlled curvature exterior mirrors* (SAE Technical Paper Series 891636). Warrendale, Pennsylvania: Society of Automotive Engineers.

Fisher, J. A., & Galer, I. A. R. (1984). The effects of decreasing the radius of curvature of convex external rear view mirrors upon drivers' judgements of vehicles approaching in the rearward visual field. *Ergonomics, 27*, 1209-1224.

Flannagan, M. J. (1988). *Human performance aspects of rearview mirrors: An applied-literature review* (Report No. UMTRI-88-20). Ann Arbor, Michigan: The University of Michigan Transportation Research Institute.

Flannagan, M. J., Sivak, M., & Traube, E. C. (1996). Driver perceptual adaptation to nonplanar rearview mirrors. In C. Serafin & G. Zobel (Eds.), *Automotive design advancements in human factors: Improving drivers' comfort and performance, SP-1155* (pp. 213-220). Warrendale, Pennsylvania: Society of Automotive Engineers.

Mortimer, R. G. (1971). *The effects of convex exterior mirrors on lane-changing and passing performance of drivers* (SAE Technical Paper Series No. 710543). New York: Society of Automotive Engineers.

Mortimer, R. G., & Jorgeson, C. M. (1974). *Drivers' vision and performance with convex exterior rearview mirrors* (SAE Technical Paper Series No. 740961). New York: Society of Automotive Engineers.

Pilhall, S. (1981). *Improved rearward view* (SAE Technical Paper Series No. 810759). Warrendale, Pennsylvania: Society of Automotive Engineers.

Platzer, G. E. (1995). *The geometry of automotive rearview mirrors: Why blind zones exist and strategies to overcome them* (SAE Technical Paper Series No. 950601). Warrendale, Pennsylvania: Society of Automotive Engineers.

Rowland, G. E., Silver, C. A., Volinsky, S. C., Behrman, J. S., Nichols, N. F., & Clisham, W. F., Jr. (1970). *A comparison of plane and convex rearview mirrors for passenger automobiles* (Report No. FH-11-7382). Washington, D.C.: Department of Transportation, National Highway Traffic Safety Administration.

Seeser, J. (1974). *Automotive convex mirrors: Optical properties* (Technical Report 201). Holland, Michigan: Donnelly Mirrors.

Smith, R. L., Bardales, M. C., & Burger, W. J. (1978). *Perceived importance of zones surrounding a vehicle and learning to use a convex mirror effectively* (DOT HS 803 713). Washington, D.C.: Department of Transportation, National Highway Traffic Safety Administration.

Sugiura, S., & Kimura, K. (1978). *Outside rearview mirror requirements for passenger cars: Curvature, size, and location* (SAE Technical Paper Series No. 780339). Warrendale, Pennsylvania: Society of Automotive Engineers.

Walraven, P. L., & Michon, J. A. (1969). *The influence of some side mirror parameters on the decisions of drivers* (SAE Technical Paper Series No. 690270). New York: Society of Automotive Engineers.

970911

Comparison of Models for Detection of Highway Obstacles with Headlamps

Walter J. Kosmatka
GE Lighting

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Models using contrast as the measure of discernibility or detection of roadway obstacle on non-lighted roadways have not recognized the lighting effects caused by the lateral spacing of headlamps on the vehicle. When the only significant sources of roadway light are a vehicle's headlamps, shadows on the roadway behind and to the left and right of the obstacle are viewed simultaneously with the obstacle itself. It can be shown that in certain obstacle-sighting situations, where there is little or no significant background light from sources other than the headlamps themselves, sufficient contrast is provided by the shadow outlines. Shadows created by the lateral spacing of the headlamp light sources and the driver's eyepoint, framing one or both sides of the obstacle, serve to outline the illuminated obstacle on the roadway in front of a vehicle. In the context of detecting a roadway obstacle, the contrast requirements are satisfied (or obviated) and threshold luminance becomes a relevant measure of obstacle discernibility. Comparison of the results of field experiments with the predictions of both *contrast predictive* and *luminance predictive* models shows that, within the limits of the situations tested, the luminance model gives a better measure of detection distance and obstacle luminance at detection.

INTRODUCTION

This paper is composed of two parts. The first part discusses the implications of an obstacle-detection model relying on threshold luminance compared with a model using contrast as the predictor of obstacle detection. The second part compares the results of a contrast-predictive model and a luminance-predictive model with the results of dynamic field test of young and old subjects and varying levels of obstacle reflectance.

PART I - HISTORICAL BACKGROUND

Contrast models are used for a variety of visibility and legibility assessments. The sighted "object" is two-dimensional and is on the surface of a two-dimensional background. Using Blackwell's [1] visual performance curves and given knowledge of the back-ground luminance, the threshold contrast requirement is defined. Comparison of this threshold contrast with the known contrast gives an indication of the likelihood of detection or legibility.

Contrast models employing these visual performance curves for obstacle detection on unlighted roadways commingle the effect of the observer's (driver's) adaptation level with the background luminance that produced the contrast in the basic lab experiments upon which these contrast models are based.

There are situations where the Blackwell data could be applied where there is no fundamental difference between the empirical experiment and the predictive application. If the data is used to predict detection or legibility in a format similar to the experiment, the immediate surround-background of the test object is continuous and extends up to the periphery of the object. This situation exists in many street lighting situations. For instance Janoff [2] uses the performance curves to analyze detection criteria for lighted roadways. In this case, the use of these curves is appropriate because the application is consistent with the experiment itself.

HEADLIGHTING ROAD SCENE FACTORS

In many nighttime automotive driving situations headlights are the only (significant) source of illumination, When a model using contrast as the predictor of detection of a roadway obstacle is applied, subtle differences in the way the scene is illuminated by the headlamps and viewed by the driver, tend to limit the

predictive utility of threshold-contrast models. There are two arguments favoring threshold-luminance-predictive models.

Contrast-Adaptation Dependencies Given the manner in which Blackwell's experiment was performed, it is likely that the contrast-background dependency is partially the result of eye adaptation to the non-zero background luminance. This was probably appropriate for the purposes of the experiment and is valid for some specific situations to which it is applied.

In roadway situations, where the only sources of light are the vehicle headlamps, the illuminated roadway section now represents a relatively small, nonuniform portion of the visual field. It extends visually from approximately 200 feet away (about a degree below the driver's horizon) to about 30 feet away (about 7 degrees below the driver's horizon). On the left, the lighted roadway extends beyond the center of the left headlamp and on the right, to the edge of the roadway or berm. Within this lighted scene, the road luminance is mostly a function of the beam pattern distribution and the distance from the source to the pavement. The pavement luminance, which often is the background for an obstacle, varies over a wide range. It is not uncommon to find 10 or even 20-to-1 ranges in pavement luminance. So it is reasonable to believe that an obstacle on the roadway, which can be located anywhere in this roadway scene, will find itself surrounded by backgrounds varying by 10 or 20-to-1.

However, the driver's adaptation level probably does not respond to such localized variations. Adaptation does not vary instantaneously over a range of 10 or 20-to-1. Normal visual tracking movements will expose the driver's eye, over a short interval of time, to most if not all of the illuminated road scene. Adaptation level will not so much depend on where in the scene an obstacle is located but more on the gross characteristics of the luminant road scene. i.e. the roadway luminance affecting the driver's adaptation level is relatively constant for a specific headlamp system beam pattern and adaptation is not a function of a localized obstacle-background contrast relationship.

This is not consistent with use of Blackwell's curves to define the threshold contrast because, as noted above, the hypothetical driver's adaptation luminance is identical with background luminance. The conventional application of the Blackwell visual performance curves results in commingling the adaptation-induced brightness threshold with the inferred contrast threshold in such a way that their independent effects cannot be separated.

Shadows and Background Discontinuities In roadway situations, the viewed object is not generally two-dimensional, nor does it lie on the background. In fact, the object and background are not coplanar nor even parallel; they intersect in almost all cases. Even more important, the object sits on the plane of the roadway and, as a result, can cast shadows on the illuminated background. As the driver observes it, the surrounding roadway background, illuminated by both headlamps, is not contiguous with the periphery of the obstacle. For this to be the case, the background would have to be either 1) back-lighted by roadway luminaires for instance or 2) illuminated with the source very near the observer's eye location. In reality, the light sources (the headlamps) are laterally spaced away from the observer. And, located where they are, they cast shadows from the perspective of the driver.

The presence and significance of shadows in detecting roadway obstacles is noted by Huculak. [3] He states that the visibility of an object "can be governed" by the luminance difference of the object's shadows. I will attempt to quantify the significance of the contrast and obstacle-framing effect of these roadway obstacle shadows. It can be shown (at least by calculation) that the shadows that serve to outline a roadway obstacle are of sufficient width and provide sufficient contrast by Blackwell's contrast definition, for detection of an obstacle over a fairly wide range of distance. Moreover, we can also show that the gradients in a headlamp beam pattern combined with the geometry of the observer-object-light source, result in a relatively constant contrast level. This makes discrimination of differences in detection distance using contrast thresholds prey to the subtle, almost undetectable changes within the light distribution in the headlamp beam pattern.

OBSTACLE-BACKGROUND CONTRAST REQUIRED

Driver's Eye Location Measurements were made of both the headlamp location on a sample of both small and large passenger vehicles. A second set of measurements was made of the location of the steering column. Given that the steering column is a practical measure of the driver's lateral head location, the difference in centers should be a good indicator of the driver's lateral position with respect to the headlamp and the sighted obstacle.

This informal sampling of vehicles shows that the steering wheel (and the driver-observer) would be located from 9 inches to 15 inches to the right of the left headlamp. In all cases, the steering column was right of the left headlamp. The headlamps were generally spaced from 3 to 5 feet apart. Given this information as a basis, the following analyses were based on lateral headlamp separation distance of 4 feet, with the driver centered 1 foot right of the left headlamp.

Eye height for passenger vehicle drivers is in the range of 40 to 44 inches. A height of 42 inches (3.5 feet) was used for the sake of simplicity and the purposes of the following calculations.

Location of Roadway Obstacle Background Using the locations noted above, and an isocandela plot of a headlamp beam pattern (and making some other general assumptions), it is possible to calculate the luminance in the general location of the shadows from each of the two headlamps as they would be seen by the driver. Assume for the sake of simplicity the following:

- Driver's eye height at 3.5 feet above ground plane
- 2-foot high obstacle centered at a height of one foot
- Headlamp mounted at a height of 2 feet

Then an angle, A to the center of an obstacle 250 feet away with respect to the driver eyepoint is,

A = atan [(3.5-1) / 250] = atan (.01) = .57o (down)

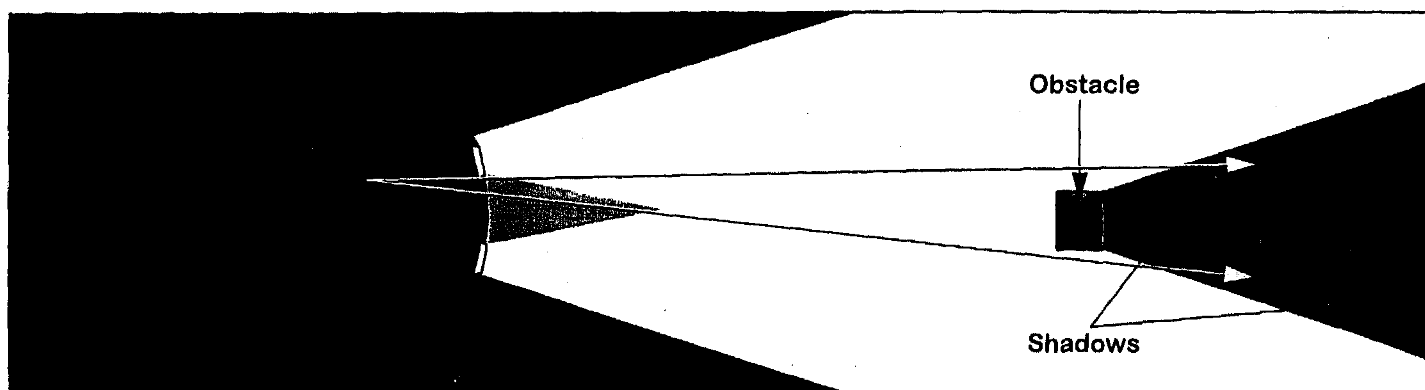


figure 1 - Plan View of Roadway

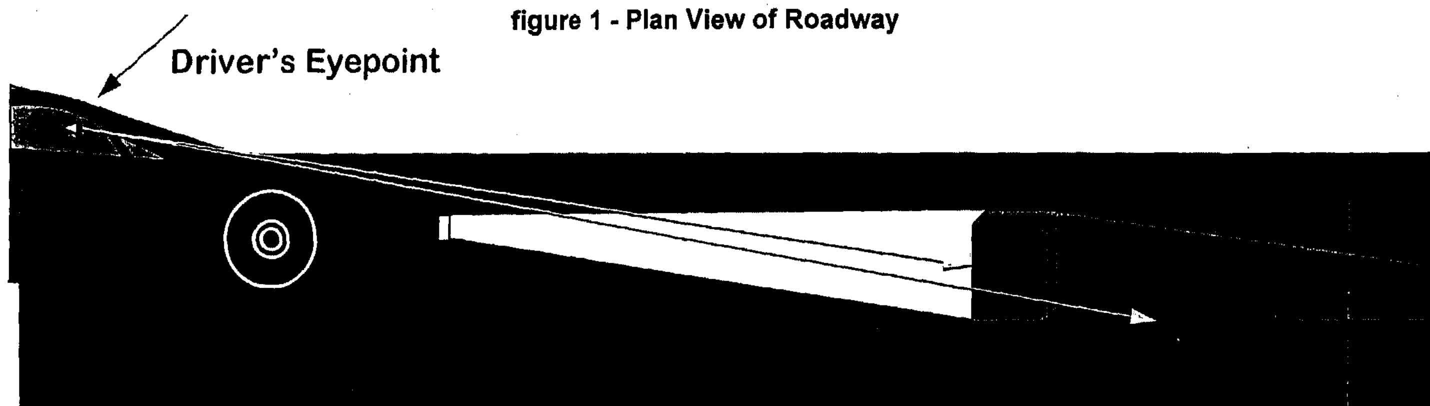


figure 2 - Side View of Roadway Scene

The location (shown schematically in **figure 1**), at distance, S, on the roadway surface which would serve as a background to the obstacle center located one foot up, 250 feet away, is defined by angle, A and the distance to the obstacle.

S = 3.5/tan .57 = 3.5/0.01 = 350 feet (from the driver's eyepoint)

A driver viewing the center of a 2 foot high object from 3.5 feet above the roadway would see, as a background, the roadway surface 350 feet away. This point would be approximately .33 degrees (atan 2/350) below the horizontal plane of a headlamp mounted at a 24 inch height.

Roadway Illumination If the lateral location of the headlamps with respect to the obstacle is given, the illumination of the background which surrounds (and contrasts with) the obstacle can be calculated.

For obstacles located almost anywhere in the lane of travel of a vehicle, the light distribution of low-beam headlamps will cause stronger contrast (lower background luminance) on the right side of the obstacle. As **figure 2** demonstrates, an obstacle located approximately between the headlamps will be to the right of the left headlamp (and to the left of the right headlamp). At this location the light directed on the roadway, past the right edge of the obstacle will be coming from the right headlamp, while the more intense light from the left headlamp is shadowed in that location by the obstacle itself. Because of the beam distribution found in all

headlamps, the left headlamp normally illuminates the roadway (obstacles) more strongly than the right lamp. On the right side of the obstacle, this illumination from the left headlamp beam is shadowed.

[Note that for objects located to the left of the vehicle the driver will see no right-side shadow. But the left-side shadow will be all the more contrasting because now the beams of both headlamps cast shadows to the left of the obstacle.]

The roadway illumination from a representative SAE low beam (shown in isocandela form in **figure 3**) was approximated to be 3000cd on the left side of the beam at .32 degrees down from the lamp's horizontal axis. The background illuminance:

$$I_{background},\ I_b = I_{hdlp}/d^2 = 3000 \times 10.76/350^2 = 0.26 \text{ Lux}$$

Shadow-Outline Contrast for Distant Obstacles At the same time the centerpoint of the obstacle (250 feet away, located 1 foot above the roadway, *i.e.* .23° below the headlamp's horizontal centerline) is receiving illumination from both headlamps. Moreover, the illumination from the left headlamp is much greater owing to the obstacle's location farther right in the headlamp's High-Intensity-Zone (HIZ).

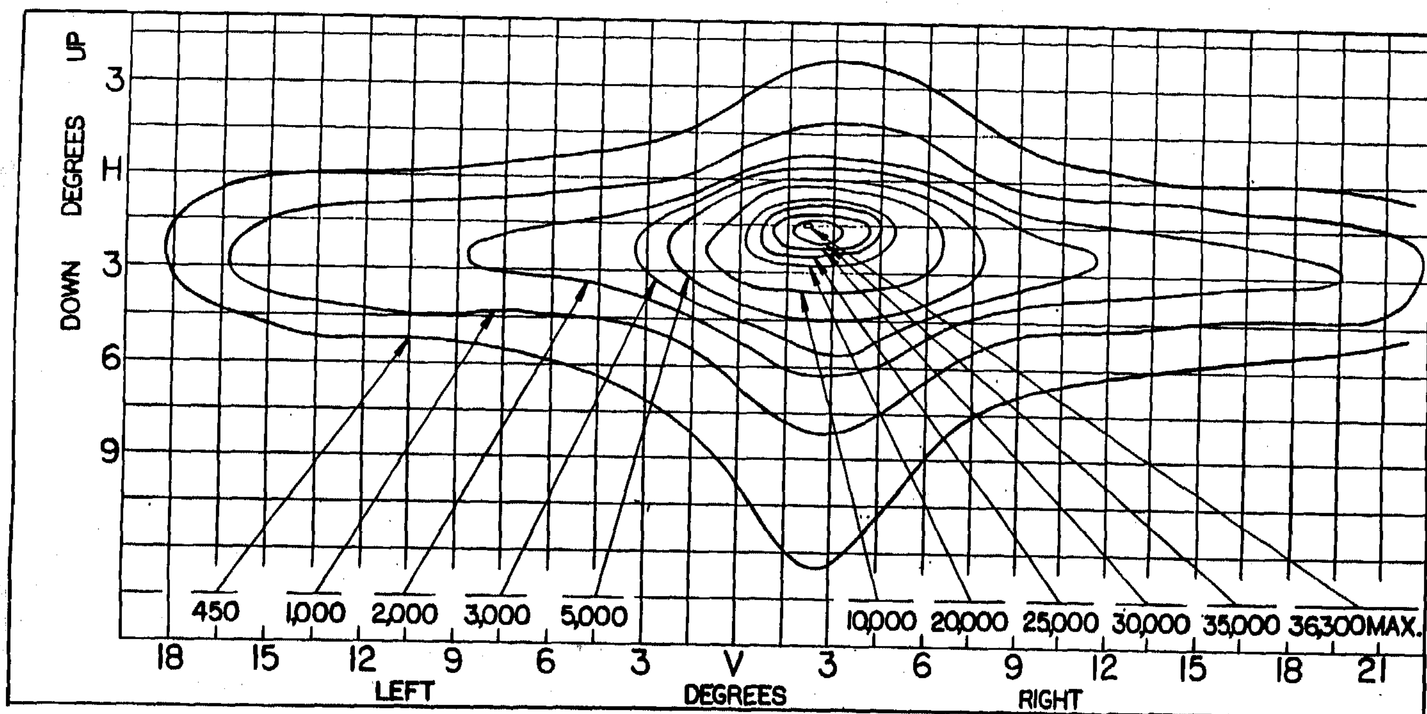


Figure 3 - Isocandela Curve for GE H6054 Headlamp

For this case, the obstacle illumination was interpolated from **figure 3** to be

$$I_{obstacle},\ I_o = (I_{rt} + I_{lt})/d^2 = (2800 + 9000) \times 10.76 /250^2$$

$$I_o = 2.03 \text{ Lux}$$

The contrast for the right edge of the obstacle and the background shadow will be

$$C_{(d=250)} = (I_o - I_b)/I_b = (2.03 \times r_o - 0.26 \times r_b)/(0.26 \times r_b)$$

If a simplifying assumption is made for the purposes of illustration, that the reflectance of the obstacle, r_o, is very close to that of the back-reflectance of the road surface, r_b (.1 for instance), the equation simplifies to

$$C_{(d=250)} = (0.203 - 0.026) / 0.026 = 6.8$$

Interpolation of Blackwell's graphical data (**figure 4**), for t = 1/3 second[a] , a 16-minute target, a 90% performance level, and B (calculated in the graph's foot-lambert units[b]) can be done for:

B = .026 x Pi/10.76 = .0076 f-l; Log B = -2.12.

[a] The target acquisition duration of 1/3 second was chosen based on the authors opinion. Huclack [3] measures rates of 3 to 5 times per second.
[b] The abcissa units on the Blackwell curves are expressed in "foot-lamberts".

We find from the **figure 4** graph, log C = approximately 0, and threshold contrast, C = 1.0.

The obstacle-shadow contrast is well above the threshold (6.8 > 1.0) at 250 feet.

Shadow Outline Size The width of the shadow around the obstacle is a function of both the lateral separation between the observer eye-point and the headlamp position. For the case above, the width is about 4 inches for the right side shadow and about 12 inches for the left side shadow. For a distance of 250 feet to the obstacle, with the shadow reference location at 350 feet, the shadow will subtend 4 to 16 arc-minutes. For shorter obstacle locations, the shadow will subtend a larger arc in the visual field. This width of shadow outline is probably sufficient for visual discrimination. Versace [4] indicates that normal visual acuity is one minute of arc except for very low luminance (.34 cd/m^2, .099f-l) and very low contrast (<.25).

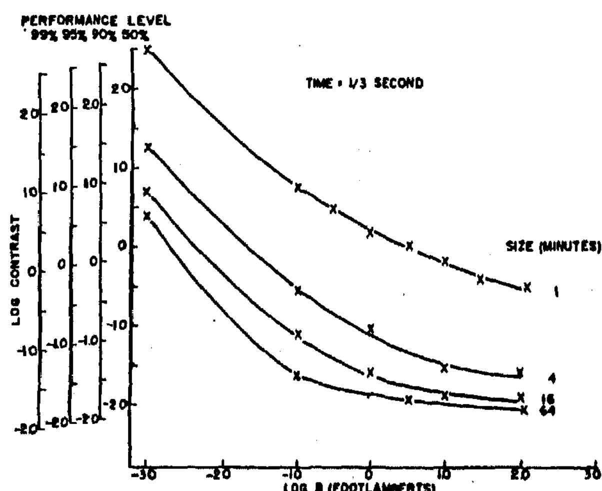


Figure Data obtained with a time of exposure of 1/3 second. Each point represents at least 500 responses by two subjects.

Figure 4 - Blackwell Visual Performance Curves

Shadow-Outline Contrast for Close Obstacles A similar analysis was done for the situation where the object is at a distance of 100 feet. The obstacle that is centered on the vehicle centerline falls approximately .57 degrees (atan 1/100) below the headlamp system's horizontal plane. The headlamps' illumination at that point can be interpolated from **figure 3**,

$$I_o = (3000 + 15000) \text{ cd} \times 10.76/100^2 = 19.4 \text{ lux}$$

The background roadway is located at about 140 feet from the headlamps (.81 deg. below the headlamp horizontal). It is illuminated by approximately 4500cd from the right headlamp,

$$I_b = 4500 \text{ cd} \times 10.76/140^2 = 2.5 \text{ lux, and}$$

$$C_{(d=100)} = (19.4 \times r_o - 2.5 \times r_b)/(2.5 \times r_b)$$

Making again an assumption of equality of the background and obstacle reflectance ($r_b = r_o = .1$), the contrast,

$$C_{(d=100)} = (1.94 - .25)/.25 = 6.8$$

Using Blackwell's curves (**figure 4**) for the now, visually larger target, (for t = 1/3 second; a 64 minute target; 90% performance level), we interpolate for log C at

$$B = .25 \times Pi/10.76 = .073 \text{ f-l} \quad (\log B = -1.14)$$

We find log C equals approximately -1.3 and thus $C_{Threshold} = 0.05$.

The contrast of the obstacle with its shadow is far above the indicated threshold (6.8 >> 0.05).

CONTRAST MODEL ASSUMPTIONS

Contrast-based models typically do not address the shadow-frame in defining the background luminance. The illumination in the immediate area of the obstacle is based on an assumption that the background is illuminated by the light from both headlamps.

Considering the *250 foot case* above, at approximately .32 degrees down from horizontal in the area of the hypothetical obstacle, the light was estimated from the isocandela plot (**figure 3**) to be about 3000cd from the right headlamp and about 7000cd from the left headlamp. Then the background illumination:

$$I_b = (3000 + 7000) \times 10.76/350^2 = 0.88 \text{ lux; and contrast}$$

$$C = (2.03 \times r_o - 0.88 \times r_b)/(0.88 \times r_b) = 1.31$$

(Assuming again, for the sake of simplicity, that the relative reflectance of the background and obstacle are coincidentally equal at 0.1.) In this case the background luminance,

$$B = .1 \times 0.88 \times 0.292 = 0.026 \text{ f-l}, \ \log B = -1.59.$$

Interpolation from the curve for t = 1/3 sec.; a target size of 16-minutes; 90% performance level, log C = -0.5, and the threshold contrast C = 0.31.

The contrast model predicts detection (*i.e.,* 1.31 > 0.31) at 250 feet but with less of a margin than the contrast of the obstacle and the shadow-frame (6.8 > 1.0) would indicate.

For the *100-foot* example used above, the illumination in the region of the obstacle from two headlamps was interpolated from the isocandela curves in **figure 3** to be 20,000 cd for the left headlamp and 4500 cd for the right headlamp. The background illuminance used for contrast-defined detection is then:

$I_b = (4500 + 20000)\text{cd} \times 10.76/140^2 = 13.5\text{lux}$

Contrast, $C = (19.4 \times r_o - 13.5 \times r_b)/(13.5 \times r_b) = .43$

(assuming again that $r_o = r_b$)

then B = .l x 13.5 x .292 = .39 f-l; and Log B = -.4

The value of Log of C, extrapolated from the curve is -1.6. Threshold contrast is 0.025.

The contrast model correctly predicts detection for the contrast threshold (0.43 > 0.025) in the *100-foot* case, but this time with a more significant margin than it predicted in the *250-foot* case.

DISCUSSION

The shadow-contrast calculations show that over the distances considered (and most likely beyond), the contrast between a road obstacle and its shadow is sufficient by Blackwell's measure for detection-discernibility.

A corollary statement could be that the contrasting shadow of an obstacle is present in almost all (or at least many) roadway situations in which the obstacle is illuminated primarily by the headlamps and that the resulting contrast is sufficient over a relatively wide range. If this hypothesis is acceptable, then it might follow that an appropriate determinant for obstacle detection using automobile headlamps is threshold luminance.

One could argue that since threshold luminance is implicit in visual performance curves, it is implicit in the contrast models using these curves. This is true, of course. However, the visual performance curves do not use threshold luminance as the independent variable but commingle threshold-luminance with adaptation level. In contrast-based models, contrast is actually a function of varying background luminance, not varying obstacle luminance. And it is the luminance of the background which induced adaptation level changes in the experiment's subjects at the same time they responded to the varying contrast. However, in actual dark roadway driving situations, adaptation is mostly a function of the interaction of the projected light pattern and roadway reflectance. It is therefore relatively constant.

If one accepts the these arguments, it would follow that the models which use the threshold contrast curves to determine thresholds, infer that observer-driver adaptation levels are essentially defined by obstacle-background contrast, when in reality they are not. Contrast deterministic model(s) used for vehicle headlighting evaluations tend to overstate the luminance of the background (the shadow) which is actually surrounding the obstacle. At the same time they use the background luminance implied by the ratio-difference contrast as a *de facto*, constantly varying, determinant of driver-observer adaptation level.

In the hypothetical examples given, the contrast model demands over 12 times (0.31/0.025) more contrast for the more distant obstacle located at 250 feet than for the same one located at 100 feet. The subtended visual angle alone does not account for the difference. The main reason for the excessive ratio is the high contrast implied for the distant obstacle owing to the lower background (lower observer adaptation) luminance.

This argument having been made, it is true that the contrast-based model would predict detection-disernibility for a 2-foot by 2-foot, 10% reflectance obstacle at 250 feet, as would a luminance-based model. As we shall see however, when it is necessary to predict the point at which a driver test subject is able to first discern an obstacle, or to predict what level of illumination is necessary for detection, some inconsistencies of contrast-based models become evident.

PART II - COMPARISON OF FIELD TEST RESULTS

CONTRAST MODEL PREDICTIONS

While attempting to test the validity of an existing contrast detection model under study by the National Highway Traffic Safety Administration, the University of Michigan Transportation Research Institute (Olson, Aoki, Battle and Flannagan) [5] performed actual dynamic field experiments. The model that UMTRI used relied on calculated contrast between the obstacle and the nearby roadway area. It predicted obstacle luminance levels 16 to 24 times that which Olson's report shows in table 7 for the younger driver test subjects.

For instance, with a small, .06 reflectance (.06 fl/fc[c]) target, a driver traveling at 25mph using low beam headlamps, detected the target with an average luminance of 0.078 cd/m^2 (.023 f-l) at 207 feet. At 207 feet, the contrast model required a contrast of .76 and target luminance of 1.9 cd/m^2 (.556f-l). The luminance required to satisfy the contrast model was about 24 times that found in the dynamic experiment.

---

[c] Reflectance factor units of f-l/fc require a conversion factor of 3.42 to convert to cd/m^2-lux

For a small, 0.12 reflectance-factor target, actual detection occurred at 269 feet on low beams, with a target luminance of .085 cd/m^2 (.025f-l) and contrast ranged from 1.1 to 2.1. However, at 269 feet the model required 1.61 cd/m^2 (.471 f-l of luminance and 2.2 contrast). The ratio of target luminance in the contrast model versus experimental target luminance is 18.8 times.

For a small 0.25 reflectance factor target, actual detection occurred at 387 feet on low beams, with a target luminance of .078 cd/m^2 (.023f-l) and contrast of 1.9. However, the contrast model at 387 feet, required 1.28 cd/m^2 (.376 f-l) of luminance and 4.7 contrast. The ratio of the target luminance in the contrast model versus the experimental target luminance is 16 times.

The conclusion that one might infer from these results and the data in the report is that this contrast model, at the very least, was not a predictor of obstacle detection in the dynamic highway situation where there is no significant overhead illumination.

Other research generally supports the detection distances found by Olson. Meese and Westlake [6] report detection distances of 244 feet to 266 feet for SAE lower-beam headlamps. Mortimer and Olson [7] report detection distances of 130 feet to 170 feet for 6% reflectance objects. They report 180 feet to 230 feet for detection of 12% reflectance obstacles and 240 to 300 feet for detection of 24% reflectance obstacles.

LUMINANCE MODEL PREDICTIONS

A proposed luminance model (Kosmatka) [8] was used to determine the location at which the obstacle luminance was sufficient to allow obstacle detection. This model uses a *base case* luminance (for detection) of 0.130 cd/m^2. The luminance requirement is modified by a size (angular area) factor, F_A, in order to account for the subtended visual angle of the obstacle or target. The 0.130 cd/m^2 (.038f-l) *base case* value was established from tests by Meese and Westlake for a variety of SAE beam patterns.

The luminance requirement divided by the target reflectance gives the illumination requirement ($I_{required}$):

$$I_{required} = L_{required} \times 1/r \times F_A = 0.13 \times 1/r \times F_A$$

The luminance of the test targets resulting from the headlamp illumination was determined by comparison of the target location in the beam pattern of the respective right and left headlamps. The obstacle illumination interpolated for each headlamp from isocandela plots was used in the general equation:

$$I_{obstacle} = I_{headlamps} \times 1/d^2$$

A graphical comparison of illumination required versus illumination provided yields the distance at which the detection requirement is satisfied. When this was done using Olson's lower beam, "small target, young driver tests", the results shown below in table 1 were obtained. (This graphical method is depicted in Appendix *A*. See also Kosmatka [8])

Table 1 - Detection Distances and Obstacle Luminance at Detection for Two Models and Olson's Field Experiment

	Experimental - Olson [5]		Kosmatka [8] Luminance Model		Contrast Model
Obstacle Reflectance	**Detection Luminance**	**Detection Distance**	**Detection Luminance**	**Detection Distance**	**Detection Luminance**
6%	0.078 cd/m^2 (.023 f-l)	207 ft.	0.085 cd/m^2 (.025 f-l)	192 ft.	1.90 cd/m^2 (.556 f-l)
12%	0.085 cd/m^2 (.025 f-l)	269 ft.	0.095 cd/m^2 (.028 f-l)	235 ft.	1.61 cd/ m^2 (.471 f-l)
25%	0.078 cd/m^2 (.023 f-l)	387 ft.	0.133 cd/m^2 (.039 f-l)	290 ft.	1.28 cd/m^2 (.376 f-l)

It should be pointed out here that the beam pattern isocandela diagrams for the lamps used in Olson's test could not be interpolated with sufficient precision to be used in determining the light intensity from the two headlamps directed at specific points. Instead, data from typical 7-inch Par 56 headlamps, the same type as the test headlamps, were used. To the extent the beam patterns could be compared, the light levels used appeared to be consistent with that found on copies of the isocandela plots for those test headlamps.

DISCUSSION OF FIELD TEST COMPARISONS

The results for the 6% and 12% targets are very similar to the luminance-based model predictions, particularly for the obstacle luminance required for detection. In the case of the 25% target, the similarity is not as good. There might be valid reasons for this but they would be difficult to verify. A small, isolated crest on the test track could momentarily direct more intense light at the test target. The observer might then have advanced warning of the target location allowing him or her to concentrate on the immediate vicinity and sight the target earlier than normal. Vehicle load changes can have a similar effect.

There appear to be inconsistencies between the experimental data and the visual-solid-angle-size (spatial summation) effect. The luminance at detection reported by Olson is more-or-less constant between 207 and 387 feet. The luminance model requires some level of increasing luminance to compensate for the decreasing size of the projected retinal image (Hills) [9]. Though it is intuitive (at least to the author) that larger obstacles or targets should be more visually prominent, the relatively constant target luminance over a two-to-one range of detection distances does not show this. Whether this is attributable simply to the probabilistic nature of human responses and experimental error or whether there is a more profound physiological basis is not apparent.

SUMMARY

A luminance-based model appears to be valid based on results of obstacle-detection observations and object luminance-at-detection measurements made in dynamic roadway tests using the vehicle headlamps as the primary light sources. Comparatively, a contrast-based detection model, tested for its ability to predict detection of roadway obstacles with vehicle headlamps, did not predict either detection distance or detection luminance very well. Two reasons are offered for this:

1. There is a framing effect caused by the shadows resulting from the lateral spacing of vehicle headlamps and their relative location with respect to the driver's eyepoint. The use of Blackwell data for obstacle detection on unlighted roadways may not be appropriate because it does not consider the shadow effect, but makes an assumption that the background is contiguous with the periphery of the obstacle.

2. Unstated in the use of Blackwell's visual performance curve is that driver (or observer) adaptation is a function of the contrast by virtue of varying back-ground luminance. In reality, adaptation is relatively static for any particular headlamp-beam distribution and road surface reflectance. Adaptation is not a function of varying obstacle contrast nor does it vary as the location of the obstacle changes over small angles within the driver's view of the road scene.

Comparison of experimental data and predictions shows that the luminance-based model fairly predicts obstacle luminance and detection distance. In one test case, using a higher reflectance object detected at greater distances there is a more significant difference between the detection luminance and the detection distance predicted by the model. The exact reason for the discrepancy cannot be determined from the information available.

REFERENCES

[1] Blackwell, H.R. 1952. Brightness Discrimination Data for the Specification of Quantity of Illumination. *Illuminating Engineering*, 47: 602-609

[2] Janoff, M. 1993 Visibility vs. response distance: a comparison of two experiments and the implications of their results. *J of the IES* Winter, 1993, 3-12.

[3] Huculak, P. 1978. *A Visibility Analysis of Obstacle Detection Experimentation in Opposed Automotive Headlighting*, NRC #16780, May, 1978.

[4] Versace, J. *Driver Vision*, Report # SAE 700391, Warrendale, Pennsylvania: Society of Automotive Engineers.

[5] Olson, P.L., Aoki, T., Battle, D., Flannagan, M. 1990. *Development of a Headlight Performance Evaluation Tool.* Report No. UMTRI-90-4. Ann Arbor, Michigan: University of Michigan Transportation Research Institute.

[6] Meese G.E., Westlake P.E. 1971. *Key Factors in Evaluating Headlighting Systems.* Report no. P_71.23. CIE, 17th Session, December 1971, Barcelona, Spain.

[7] Mortimer, R.G., Olson,P.L. *Development and Use of Driving Tests to Evaluate Headlamp Beams..* Report no. UM-HSRI-HF-74-14. Ann Arbor, Michigan: University of Michigan Transportation Research Institute.

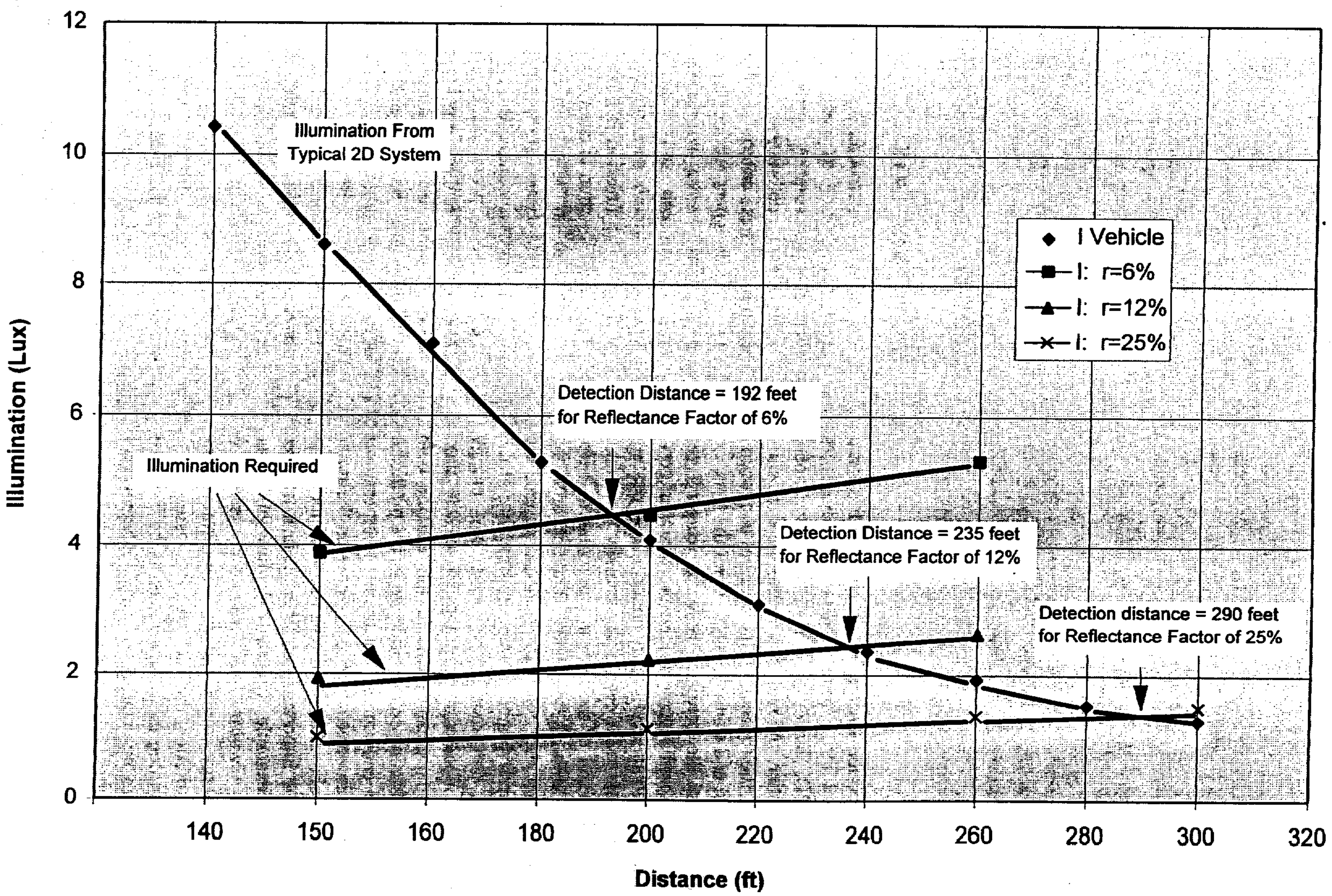
[8] Kosmatka, W.J. 1995. Obstacle detection rationale for vehicle headlamps. *J of the IES*, Winter 1995, 24-1: 36-40.

[9] Hills, B.L. 1976. Visibility under night driving conditions, part 3.derivation of (delta L, A) characteristics and factors in their application. *Ltg Res & Technol* 8:11-26.

ABOUT THE AUTHOR

Walter Kosmatka graduated from the University of Wisconsin with degrees in engineering and business. He is employed by GE and has spent most of his 30 years with the company dealing with the design, manufacture and testing of lighting products. He holds 17 U.S. and foreign patents in the field of automotive lighting. Mr. Kosmatka has chaired several SAE lighting task forces and authored SAE standard SAE J1735 and information report SAE J2338. He is currently vicechair of the SAE Road Illumination Devices Committee. His address is 390 Sandhurst Drive, Highland Heights, Ohio 44143.

Appendix A: Detection Distances
for Younger Drivers



970912

Predicting the Thermal Performance of an Automotive Fog Lamp Reflector

E. W. Liang, R. D. Lillquist, J. P. Gallo, and A. J. Poslinski
GE Corporate Research and Development

A. Grimson
GE Plastics

B. R. Strauss
North American Lighting, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The thermal performance of an automotive fog lamp reflector is predicted with a computational fluid dynamics program. The 2D, steady-state heat transfer model accounts for convection and radiation within the enclosure, conduction through the reflector walls, external convection and radiation losses, and transmission through the lens. A good comparison of the predicted reflector temperatures with experimental thermocouple and infrared data suggests that the specular component of the inner metal coating plays a secondary role in the overall heat transfer and that a detailed thermal model of the bulb is not required. The radiant exchange with the tungsten filament and the conductive energy losses through the bulb connections are accounted for by specifying an appropriate heat flux at the bulb surface, and the transmission of radiant energy through the lens is modeled with an appropriate heat sink. Driven by the thermal expansion of the air near the bulb surface, 4 counter-rotating recirculation zones are predicted within the reflector enclosure. The highest temperatures are predicted at the intersection of these zones on the inner surface of the shelf above the bulb. The thermal model can be used to assess the importance of different heat-transfer mechanisms. For example, sensitivity studies show that thermal radiation within the reflector enclosure redistributes the energy so that the wall temperatures are more uniform. In addition, thermal radiation losses to the environment surrounding the reflector need to be included in the thermal model; otherwise, the predicted wall temperatures are higher than the experimental data.

INTRODUCTION

Two key criteria for part design in automotive lighting applications, such as head lamps and fog lamps, are thermal performance and beam shift caused by thermoelastic deformation of the lens and reflector assembly. These criteria are especially important for head lamps as designers are moving the optics from the lens to the reflector. Lenses that were previously faceted for optics are now smooth and clear, and reflectors that were previously smooth parabolas now have complex faceted surfaces. The situation is further exacerbated as the beam adjustment function is changed from the lens and reflector assembly to an inner movable reflector. To meet the thermal and beam shift requirements, the automotive industry currently uses a thermoset polyester composite for the reflectors, providing high-heat capability and low thermal expansion, and a polycarbonate (PC) thermoplastic resin for the lens. The disadvantages of thermoset reflectors are their higher weight, non-recyclability, secondary trimming operations, and base coating for metallization. Recently, an amorphous polyetherimide (PEI) thermoplastic resin has been introduced for the reflector applications, reducing the part weight, introducing recyclability, and eliminating secondary trimming and basecoat operations. Coincidentally, tests have shown that the PEI reflectors provide improved beam patterns over thermoset reflectors because the faceted corners are not rounded off by the basecoat necessary for thermosets. In order for the PEI resin to be successfully implemented, the thermoplastic reflector must meet the high operating temperatures and tight beam-shift tolerances under which thermoset materials currently perform. The focus of this work was to develop a predictive methodology based on available computational tools that would link the automotive lighting requirements with the material

properties of thermoplastics. A numerical sensitivity analysis could then be used to suggest design improvements and to drive material formulation changes that would help position the PEI resin as a viable material option in the automotive lighting market.

The strategy for predicting the thermal performance and beam shift of automotive lighting applications is illustrated in Figure 1. Whereas sophisticated ray-tracing software is commonly used to optimize the reflector geometry for optics at the outset of the design process, material selection, thermal performance, and thermoelastic deformation are usually addressed later through experimental trial and error. It is proposed that the last 2 steps also be modeled. The thermal analysis can be accomplished with a computational fluid dynamics (CFD) package that is capable of simulating coupled conduction, convection, and thermal radiation inside the reflector enclosure. The required material properties include the thermal conductivity, specific heat of the air, and surface emissivity. Once the temperature distribution of the reflector is calculated, the resulting thermoelastic deformation can be predicted with a structural analysis package. The additional material properties include the coefficient of linear thermal expansion (*CTE*) and the elastic modulus. The deformed surfaces can then be used in a ray tracing analysis to predict the beam pattern and to compare with the results obtained with the undeformed geometry. The maximum operating temperature and beam-shift predictions provide the information needed to change the reflector geometry or to improve the material's thermal performance while maintaining flow, impact resistance, and coating adhesion. Using a fog lamp reflector as an example, this report focuses on the third step in the design process, summarizing the minimum-complexity thermal model developed within the limitations of available computational tools, comparing the predicted temperatures with experimental data, and presenting a sensitivity study on the effects of changing some of the heat-transfer mechanisms, material properties, and operating conditions.

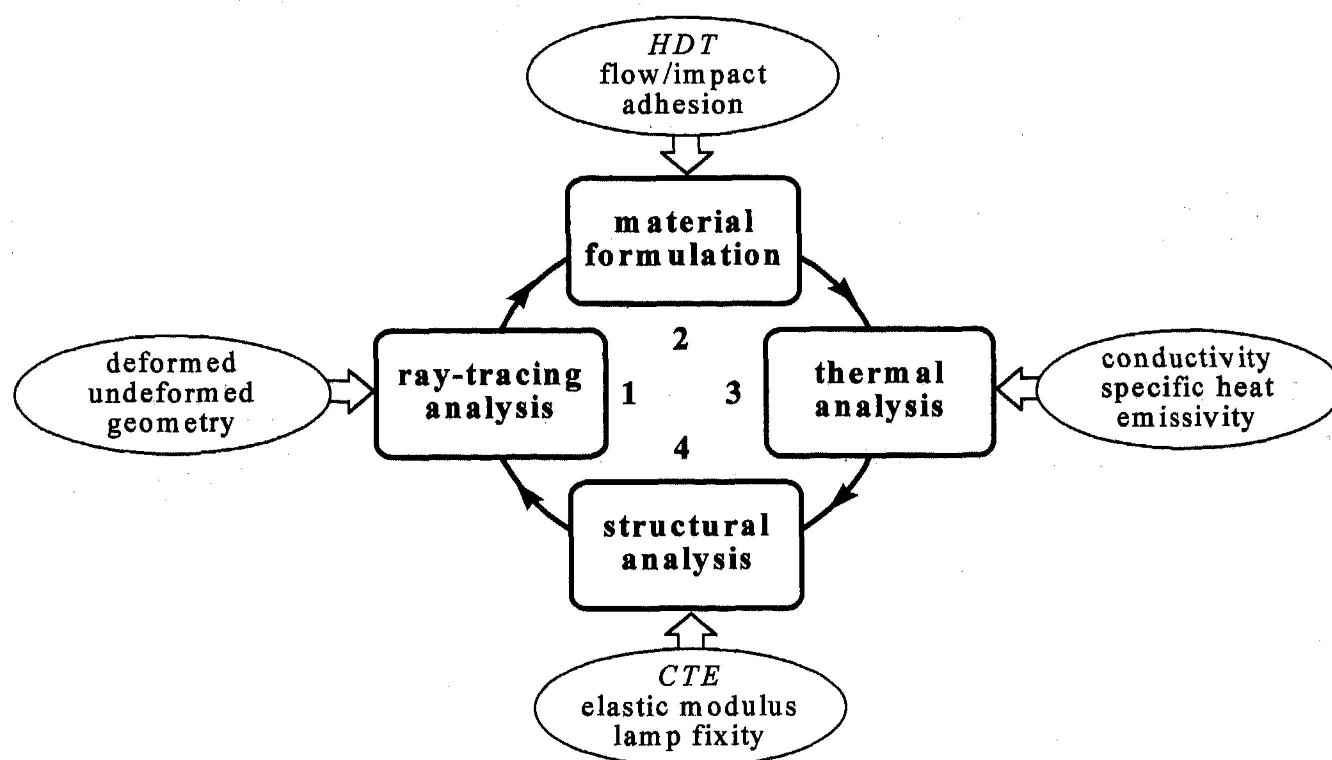


Figure 1: Strategy for predicting the thermal performance and beam shift of automotive lighting applications.

Table I: Radiative properties required for the thermal analysis.

Surface	Emissivity ϵ	Reflectivity ρ	Transmissivity τ
PEI (2-14 μm)	0.96	0.04	opaque
PC (2-14 μm)	0.94	0.06	opaque
Al (2-14 μm)	0.04	0.96	opaque
Quartz (3-14 μm)	0.90	0.10	opaque

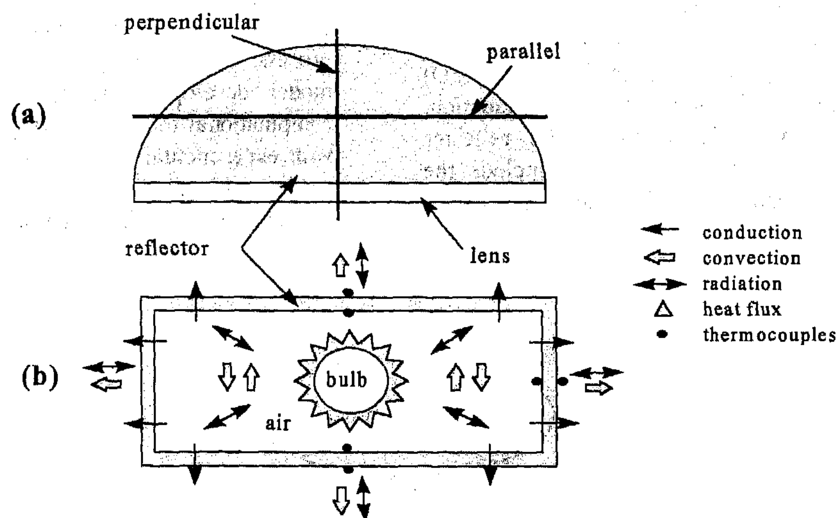


Figure 2: Schematic representation of (a) the top view showing the 2D cross-sections used to set up the thermal model, and (b) the front view showing the thermocouple locations and the heat-transfer mechanisms accounted for in the thermal analysis.

Table II: Experimental thermocouple data under steady-state conditions (°C).

Thermocouple	28 W-25C	28 W-70C	28 W-50C*	37.5 W-50C
Inside top shelf	149	162	156	252
Outside top shelf	140	155	148	190
Inside bottom shelf	137	156	147	222
Outside bottom shelf	—	—	—	173
Inside side wall	66	90	78	116
Outside side wall	64	88	76	89
Inside lens	88	103	96	—
Outside lens	86	103	96	—

**Linearly interpolated with the 25C and 70C data.*

RADIATIVE PROPERTIES

Because the surface temperatures of an illuminated bulb and a tungsten filament are on the order of 500 and 2,700C, thermal radiation plays an important role in the overall heat transfer, and the radiative properties of the bulb, filament, and reflector surfaces are required. These include the absorptivity, reflectivity, and transmissivity, which respectively measure the fraction of absorbed, reflected, and transmitted radiation relative to the incident flux arriving at the surface from other sources. The amount of radiation given off by a surface is characterized by its emissivity, the ratio of emitted radiation relative to the maximum amount possible given off by a black body at the same temperature. In general, the radiative properties exhibit angle, wavelength, and temperature dependence; in the present approach, the radiant exchange is assumed to be gray-diffuse (wavelength and angle independent) and the enclosed air does not participate in the radiation heat transfer. For a gray-diffuse surface of an opaque body, all the incident radiation is either reflected or absorbed by the body. Furthermore, Kirchoff's law states that the surface emissivity $(0 < \epsilon < 1)$ is equal to its absorptivity $(0 < \alpha < 1)$, the sum of the emissivity and surface reflectivity $(0 < \rho < 1)$ equals unity, and the emissivity is thus the only surface property required in the analysis. To further simplify the problem, the emissivity is assumed to be independent of surface temperature.

Pure, smooth metals are known to exhibit low emissivity and absorptivity and therefore high reflectivity; whereas, the opposite is generally observed for dielectric materials such as thermoplastics [1]. Rather than relying on published values and introducing a potential source of error in the numerical calculations, the radiative properties of the thermoplastic resins making up the fog lamp reflector and lens were directly measured. Small plaques were molded out of the PEI and PC resins, and some of the PEI samples were metallized on 1 side with a relatively thick aluminum (Al) coating. To confirm the gray-diffuse assumptions, all 3 properties—emissivity, transmissivity, and reflectivity—were measured using an imaging infrared (IR) radiometer in the wavelength regions where most of the radiant exchange occurs. An inspection of the curves for the emissive power of a black body as a function of wavelength reveals that for temperatures between 200 and 500C significant exchange occurs in the mid-IR region. Consequently, average values were obtained in 2 bands of 2-6 μm and 7-14 μm with 3 simple tests. First, the opacity of each sample was verified by measuring the transmission of radiation through the sample from a black body source. The emissivity was then obtained by simultaneously comparing the IR intensity emitted from the sample and a previously characterized reference, both heated to the desired temperature on a hot plate. Finally, the reflectivity at various angles of an incident IR beam from a blackbody source was determined by comparing the reflected intensity with that of a reference gold mirror.

The IR measurements show that between wavelengths of 2 to 14 μm both the PEI and PC samples are essentially opaque and the radiative properties, summarized in Table I, are approximately constant. Of course, the PC sample is transparent between wavelengths of 0.3 and 2.0 μm [2]. The aluminum coating layer on top of the PEI resin is somewhat specular (reflects in a *mirror-like* manner). When dealing with specular surfaces, it is necessary to account for the directional paths that the reflected radiation follows since multiple reflections are possible. Within enclosures, these directional effects are usually small because of the many reflections taking place between the surfaces [1]. In the fog lamp reflector, multiple reflections are not likely to occur because most of the radiant energy emanating from the tungsten filament and passing through the bulb is reflected by the aluminum coating toward the unmetallized portions of the PEI reflector and the PC lens, where it is either absorbed or transmitted to the external environment. In an effort to develop a minimum-complexity model, the specular component of the metal coating is neglected; the error associated with this approximation is quantified by comparison of the predicted temperatures with experimental data. The additional surface that needs to be considered is that of the quartz bulb. Published radiative properties [3] show that quartz exhibits wavelength-dependent properties in the range of interest. Typically, quartz becomes highly absorbant above 3.0 μm; however, it is transparent between wavelengths of 0.3 and 3.0 μm. Within this relatively narrow band, the quartz exposes the lens and reflector surfaces to the hot tungsten filament. Consistent with the assumptions made for the thermoplastic resins and metal coating, the bulb surface is considered to be gray-diffuse and the filament radiation is accounted for by specifying an appropriate heat flux at the bulb surface; these additional sources of error are quantified by comparison with experimental data.

TEMPERATURE MEASUREMENTS

THERMOCOUPLE DATA -- Steady operational temperatures of a fog lamp reflector were measured in a controlled oven environment [4]. Black pigmented PEI reflectors were partially metallized with an aluminum coating without the use of a basecoat. Based on the extinction coefficient of aluminum [1], the mean penetrating distance of incident radiation is about 0.1 μm so that the coating in the present study can be assumed opaque. Only the parabolic portion of the reflector was metallized; the top and bottom shelves were masked during the metallization process. The parabolic portion of the

reflector extends forward and forms the side walls. Six thermocouples were mounted on the inner and outer surfaces of the reflector with a glass-reinforced PEI tape as shown in Figure 2. Subsequently, a PC lens was bonded to the reflector using an adhesive, a light bulb rated at 37.5 W was inserted into the socket, and the reflector assembly was placed in a non-circulating oven at approximately 50C. The bulb was illuminated at an abusive condition of 14 V (12.8 V standard) for approximately 1 hr, with the temperatures being recorded via computer every 5 sec. A steady state was reached within 10 to 15 min; the final temperatures at each of the 6 locations are recorded in Table II. Additional tests with a lower power, 28 W bulb and different oven temperatures are also shown in the table for comparison. The hottest temperature is observed on the inside surface of the top shelf as hot air is being convected upward from the bulb. In contrast, the air circulation cools the inside surfaces of the bottom shelf and side walls. The lowest temperatures are recorded on the sides where the aluminum coating reflects most of the radiant energy back into the enclosure. The high reflectivity and inner convective cooling also cause the temperature drop through the side walls to be lower than the drop through the top and bottom shelves. Changes in the oven temperature have a less significant effect on the reflector's thermal performance in comparison to changes in the input power. For example, the data show that a 45C increase in the oven temperature results in a 15-20C increase on the reflector temperatures. In contrast, a 10 W increase in the input power can cause an increase as high as 100C.

INFRARED DATA -- The imaging IR radiometer used to measure the radiative properties of the thermoplastic materials was also used to image the steady operational temperatures of the fog lamp reflector and the bulb. An example of the IR data for the outside surfaces of the top shelf and the lens, obtained by illuminating the bulb inside the reflector at the standard operating condition of 12.8 V in an ambient environment of approximately 25C is shown in Figure 3. The internal reflector temperatures are accessible only if the PC lens is replaced by a material transparent in the IR wavelength range. In contrast to the single-point thermocouple data, a contour map is easily produced, pinpointing the actual location of the maximum values. A circular pattern of isotherms is observed with the hot spot centrally positioned on the top shelf of the PEI reflector. The darker band toward the front of the reflector in Figure 3(a) is an extended rib around the periphery attaching the lens, a portion of which can be seen in the top view. The PC lens temperature is significantly lower than the top shelf with the hottest temperature shifted toward the upper half. Although the experiment was carried out at room temperature and at a lower voltage, the overall temperatures are not much different from the thermocouple data in Table II. Table III lists the quartz surface temperatures measured with the infrared radiometer by illuminating the bulb outside the reflector at 3 voltage levels in an ambient environment of approximately 25C. The electric current was also obtained to compute the power input to the bulb. The results show that the top surface is hotter by about 60 to 70C than the lower surface due to air convection inside and outside the bulb. Both surfaces exhibit a wide but similar temperature spread. The highest temperatures are recorded on the blackened surface in front of the hot tungsten filament.

Table III: Quartz surface temperatures as a function of the applied voltage and power.

SURFACE	VOLTAGE (V)	POWER (W)	MINIMUM (^{O}C)	MAXIMUM (^{O}C)	AVERAGE (^{O}C)
Front	12.7	38.0	447	484	469
	13.7	42.5	480	513	500
	14.7	47.5	510	544	531
Top	12.7	38.0	287	502	410
	13.7	42.5	300	514	424
	14.7	47.5	320	536	442
Bottom	12.7	38.0	260	424	342
	13.7	42.5	273	449	356
	14.7	47.5	289	474	378

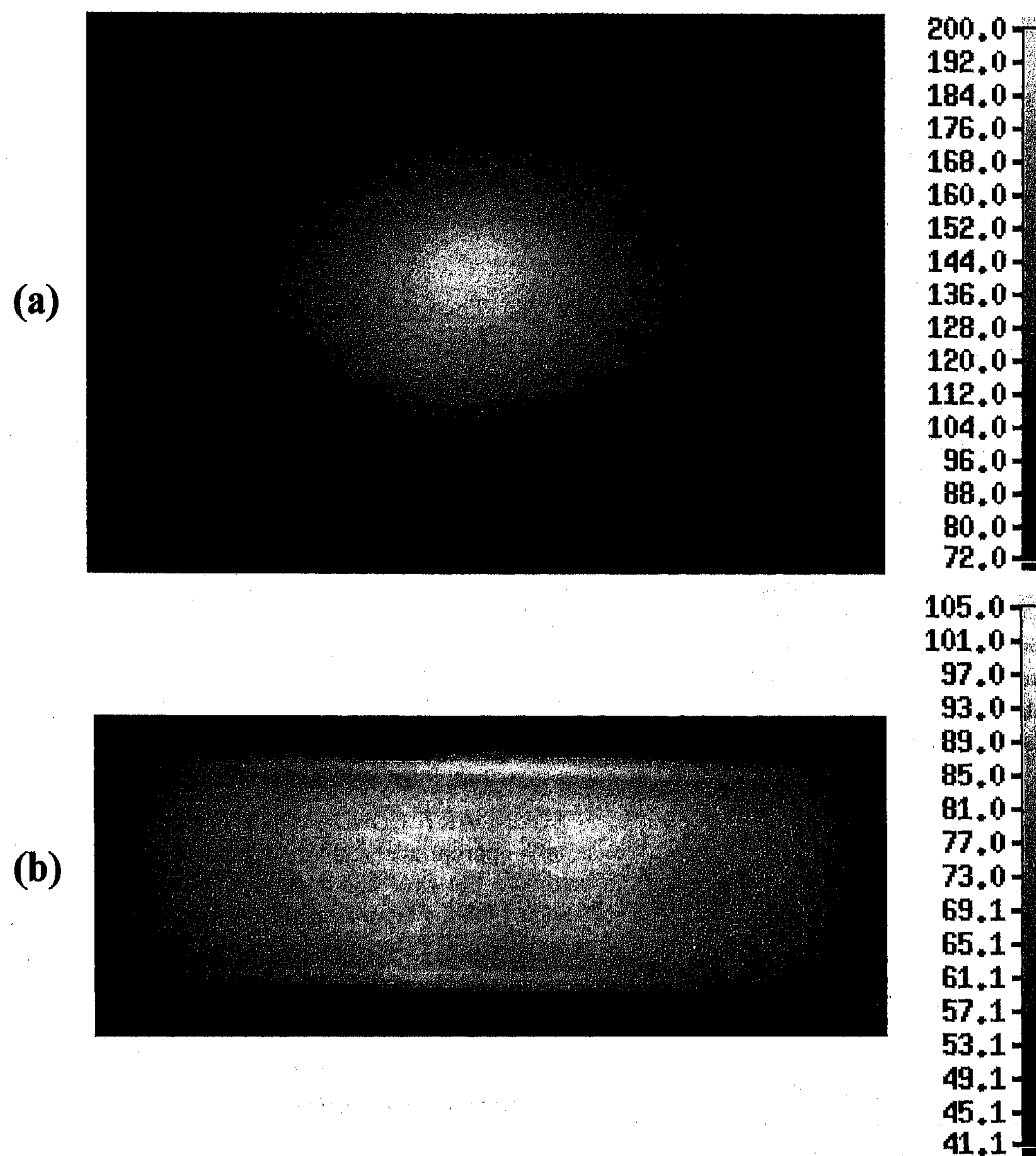


Figure 3: IR images showing the temperature contours on the external surfaces of (a) the top shelf and (b) the lens.

NUMERICAL MODEL

As illustrated in Figure 2, the thermal performance of the fog lamp reflector is determined by convective and radiative heat transfer in the enclosure, by conduction through the reflector walls, and by convection and radiation to the external environment. The FIDAP® CFD package is used to develop a 2D heat transfer model. In this section, a brief overview of the FIDAP features relevant to the present problem is first discussed, followed by a description of the governing equations, the boundary conditions, and the radiation modeling. Several numerical solution schemes and their convergence performance are also described.

® FIDAP is a Registered Trademark of Fluid Dynamics International, Inc.

FIDAP SOFTWARE – The FIDAP package is a general-purpose computer program for simulating, among other phenomena, fluid flow and heat transfer in either 2- or 3D domains. Based on the finite element technique, the domain of interest is divided into small regions (finite elements) and the governing equations are then replaced by a set of algebraic equations in each region. The algebraic system is then solved numerically for the dependent variables such as velocity, temperature, and pressure. This software is capable of solving steady-state or transient, laminar or turbulent, and incompressible or compressible flows. The program can simulate heat transfer in the fluid and solid regions, coupled with heat convection in the fluid and gray-diffuse radiant-heat exchange between the solid surfaces. Although extensions to a non-gray analysis are possible, transmission and specular behavior cannot be

modeled in the current version of the software. For further information, consult the *FIDAP Theory Manual* [5].

GOVERNING EQUATIONS -- In the 2D thermal model, the thermoplastic reflector and lens are modeled as solids that transport energy by conduction. The partial transparency of the PC material in the visible and near-IR wavelength range cannot be modeled. In turn, the air within the enclosure is described as an incompressible viscous fluid, which undergoes laminar flow, transports energy by conduction and convection, and does not participate in the radiation heat transfer. The radiant heat exchange between the bulb and inner reflector surfaces is coupled with the energy balance equation for the air through appropriate boundary conditions. The light bulb is implicitly modeled by replacing its volume in the finite-element mesh with a heat flux boundary condition along the bulb surface as described in ***Boundary Conditions***.

The equations governing the flow of a non-isothermal single-phase viscous fluid are mathematical representations of the conservation of linear momentum, mass, and energy. The momentum balance equation for a steady flow of an incompressible fluid with the buoyancy body force added to account for the thermal expansion of the air is given by:

Momentum Balance (1)

$$\rho_0 \mathbf{u} \bullet \nabla \mathbf{u} = -\nabla p + \mu \nabla^2 \mathbf{u} - \rho_0 [\beta_T (T - T_0)] \mathbf{g}$$

where $\mathbf{u}$ is the velocity vector, p is the pressure, μ is the dynamic viscosity and ρ_0 is the fluid density at the reference temperature T_0. The last term in Eq. (1) represents the Boussinesq approximation for the density-driven buoyancy force resulting from a non-uniform temperature distribution, with β_T representing the thermal expansion coefficient, T, the local temperature, and $\mathbf{g}$, the acceleration vector due to gravity. The momentum equation has to be solved in conjunction with the following mass balance equation:

Mass Balance (2)

$$\nabla \bullet \mathbf{u} = 0$$

For incompressible fluids, the conservation of energy under steady-state conditions with negligible viscous heating effects and no other heat sources is expressed by:

Energy Balance (3)

$$\rho_0 c_p \mathbf{u} \bullet \nabla T = k \nabla^2 T$$

where c_p is the specific heat, and k is the thermal conductivity. In the solid region, the convective term on the left-hand side does not apply, and Eq. (3) reduces to the Laplace equation for the temperature. The material properties appearing in Eqs. (1) to (3) are listed in Table IV; they are taken from the Engineering Design Database [6] produced by GE Plastics and literature (e.g., [7]).

Table IV: Thermophysical properties required for the thermal analysis.

Properties	PEI	PC	Air @ 200C
ρ_0 (kg/m^3)	1270	1200	0.746
k (W/m/K)	0.22	0.19	0.0378
c_p (J/kg/K)	1100	1250	1025
μ (kg/m/s)	—	—	0.0000258
β_T (K^{-1})	—	—	0.00211

BOUNDARY CONDITIONS -- Along the domain boundary, appropriate boundary conditions must be applied to complete the flow and heat transfer model. The air enclosed by the fog lamp assembly is the only fluid medium to be analyzed, and a no-slip boundary condition is imposed on the bulb surface and along all the inside surfaces of the reflector:

Flow Boundary Condition (4)

$$\mathbf{u} = \mathbf{0}$$

In addition to specifying the ambient temperature, the following heat flux condition is prescribed along the normal to each of the domain boundaries $\mathbf{n}$:

Heat – Transfer Boundary Condition (5)

$$-k \nabla T \bullet \mathbf{n} = q_a + q_r + q_c$$

accounting for any applied q_a, radiative q_r, and convective (or conductive) q_c contributions. The energy balance at the surface of the light bulb is modeled with an applied and radiative heat flux ($q_a > 0$, $q_r > 0$, $q_c = 0$), and the bulb volume is eliminated from the finite-element mesh. The convection of heat away from the bulb surface is accounted for by the energy balance in Eq. (3).

Table V: Heat transfer coefficients on the outer reflector surfaces.

Surface	h_c (W/m^2)
Top shelf	9.75
Bottom shelf	4.73
Side walls	8.26

On the exterior surfaces of the reflector, heat is both radiated and convected to the external surroundings so that both the radiative and convective heat-flux boundary conditions are included in the heat transfer calculations ($q_a = 0, q_r > 0, q_c > 0$). Newton's law of cooling:

External Convection Condition

$$q_c = h_c(T - T_\infty) \quad (6)$$

is substituted for the convective part, and the ambient temperature is used as input for the reference temperature T_∞. Assuming that the heat transfer from the fog lamp to the external surroundings is laminar, the heat transfer coefficients h_c can be estimated from generally accepted empirical correlation for free convection past the outer surfaces of the reflector [7]. The results calculated at an ambient temperature of 50C are listed in Table V. Along the boundaries dividing the reflector and the enclosed air, heat is conducted into the walls or radiated back to the enclosure ($q_a = 0,\ q_r > 0,\ q_c > 0$). Fourier's law is substituted for q_c so that the temperature gradients on both sides of the solid-fluid interface are offset by the ratio of solid to fluid thermal conductivities and the amount of energy radiated back to the enclosure. Heat convection along the inner reflector walls is accounted for by the energy balance in Eq. (3).

RADIATION MODELING -- The calculation of radiant exchange between a pair of surfaces requires the determination of a geometrical configuration factor, or view factor, between the 2 surfaces. The view factor ϕ_{ij} is defined as the ratio of the radiant energy arriving at surface j relative to the total radiant energy leaving surface i. By definition, the net radiative heat flux leaving from surface i due to emission and reflection can be written as:

Flux Leaving Surface i

$$q_r^i = se_i T_i^4 - \sum_{j=1}^{N} \phi_{ij} \left[se_j T_j^4 - (1 - e_j) q_r^j \right] \quad (7)$$

View Factor

$$\phi_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos\beta_i \cos\beta_j \, dS_i dS_j}{\pi r^2} \quad (8)$$

where $s = 5.67 \times 10^{-8}$ W/m^2 K^4 is the Stefan-Boltzmann constant, and β_i , β_j are the angles formed by the intersection of the surface normal vectors $\mathbf{n}_i$, $\mathbf{n}_j$ with the line of length r connecting the 2 points i , j. Each finite element on the bulb and reflector surfaces radiates thermal energy, and the view factor between any 2 elements must be calculated as part of the numerical solution. Because the light bulb obstructs the radiant exchange between the reflector surfaces, it is considered as a blocking surface in the view factor calculation. The external reflector surfaces also lose heat by radiation. Thus, Eq. (7) reduces to the simpler form of the Stefan-Boltzman law:

External Radiation Condition

$$q_r = se(T^4 - T_\infty^4) \quad (9)$$

NUMERICAL RESULTS

2D DISCRETIZATION -- The present numerical calculations are based on the 2D cross-sections shown in Figure 2: perpendicular and parallel to the lens and passing through the bulb center. A finite-element model of the cross-section perpendicular to the lens is shown in Fig. 4(a), where the wall on the right-hand side is the PC lens and the wall on the left-hand side is the metallized portion of the PEI reflector. The thin metal coating is modeled with surface elements in the analysis, and the low emissivity of the aluminum is used to simulate its high reflectivity. The finite-element model for the parallel configuration is symmetric with respect to the center of the light bulb, and it is shown in Figure 4(b). The upper and lower walls are the uncoated top and bottom shelves; whereas, the right and left walls are the metallized portion of the PEI reflector.

Table VI: Comparison between predicted and experimental temperatures (°C).

DATA TYPE	TOP/IN	TOP/OUT	BOTTOM/IN	BOTTOM/OUT	WALL/IN	WALL/OUT
Perpendicular (8,000 W/m^2)	255	215	240	205	—	—
Parallel (17,500 W/m^2)	248	216	219	197	122	112
Thermocouples	252	190	222	173	116	89
IR measurements	—	205	—	150	—	80

EXPERIMENTAL COMPARISON -- Numerical calculations were performed for the fog lamp reflector using the gray-diffuse approximation for internal radiation, and accounting for both convection and radiation on the exterior surfaces of the reflector. The applied heat fluxes of $q_a = 8{,}000$ W/m^2 and $q_a = 17{,}500$ W/m^2 were estimated for the perpendicular and parallel cross-sections based on the reflector geometry and the total power input. The comparisons between the simulated results and the thermocouple and IR measurements for the 2 cross-sections are summarized in Table VI. Note that the infrared temperatures were obtained at the same voltage as the thermocouple data (14 V), but at a room temperature of 25C, and have been raised by approximately 10C to compensate for the different oven condition of 50C. This correction factor is based on the change in overall temperatures calculated with the heat transfer model as discussed in **Sensitivity Studies**. The predictions for the top and bottom shelves in both cross-sections are in reasonably good agreement with the experimental data, given the model assumptions and the discrepancies in the measurements. One possible improvement, especially in the perpendicular is the application of a smaller heat flux on the bottom surface of the bulb in accordance with the lower surface temperatures in Table III. The smaller heat flux may bring the predicted temperatures on the bottom shelf of the reflector closer to the actual data. The higher temperature drop across the bottom shelf in the perpendicular case is due to the thicker wall. In fact, the temperature gradient across both sections is about 14 C/mm.

Figure 5 shows the predicted temperature distribution in the perpendicular and parallel cross-sections. For the perpendicular cross-section, the contours are asymmetric due to the different radiation properties of the metallized reflector on the left-hand side and the PC lens on the right-hand side. Because more energy is reflected from the metallized reflector onto the bulb, the peak bulb temperature of about 400C is located towards the rear on the top surface. Directly above, the heat is convected upward creating a hot spot on the top shelf. The experimental IR contours in Figure 3 show that the hot spot is located more toward the center of the reflector. In reality, the reflected energy from the side walls to the bulb probably shift the hot zone forward; this 3D effect cannot be captured with the present 2D model. The temperatures on the front, bottom, and back surfaces of the bulb are similar and, in an average sense, close to the measured data listed in Table III for the bottom surface. The front of the actual bulb is covered with a black coating to prevent the light from traveling directly to the lens. Because the emissivity of the other bulb surfaces is close to unity (ϵ = 0.9), using a black-body approximation (ϵ = 1) for the front surface does not sufficiently increase the temperature to match the high values measured with the imaging IR radiometer. Application of a higher heat flux on the front surface of the bulb may bring the predicted bulb temperatures closer to the actual data. For the parallel cross-section, the temperature distribution is symmetric. The peak bulb temperature is approximately 480C on the top surface in agreement with the averaged measured values, and the hot spot is located directly above on the upper shelf.

Figure 6 shows the temperature distribution along the centerline from the bottom to the top shelf in the perpendicular and parallel cross-sections. The results in this figure are used to determine the temperature data of the top and bottom shelves listed in Table VI. The curves for both cross-sections are similar, showing that the air above the light bulb is close to 300C, about 100C hotter than below. Figure 7 shows the temperature distribution along the centerline of each cross-section in the horizontal direction. This figure is used to obtain the wall temperature data listed in Table VI. The air temperature on the sides of the bulb is about 200C, rapidly dropping near the reflector walls. Consistent with the results in Figure 5(b), the temperature distribution in Figure 7(b) for the parallel cross-section is symmetric, and the predicted temperatures for the metallized walls at x/a = 0, 1 are similar to those in the perpendicular cross-section at x/c = 0, where a, b, c are the depth, height, and width of the reflector. The calculated velocity vectors for the air flow

inside the housing are shown in Figure 8. Two recirculation zones are observed due to the expansion and buoyancy of the air near the hot bulb and the contraction near the cool walls. The recirculation zones are better observed in the streamline contours shown in Figure 9. The strongest air flow occurs near the light bulb and along the inner surfaces of the reflector, where the streamlines are closest to each other. The location of the hot spot corresponds to the point at which the recirculation zones meet. Based on the calculated velocity field, the highest Reynolds number, Re—which measures the ratio of inertial to viscous forces—found in all the simulations is less than 20; therefore, the laminar flow assumption is justified (i.e., Re < 2100 [8]).

LENS TRANSMISSIVITY APPROXIMATION -- Examining Figures 5(a) and 7(a) reveals that the temperature of the PC lens is around 180C, considerably higher than the measured 115 to 120C values. As discussed in **FIDAP Software** and **Governing Equations,** the present model does not account for the radiant energy that is transmitted through the PC between wavelengths of 0.3 and 2.0 μm. One way to account for this phenomenon is to impose a heat sink along the interface separating the lens and the enclosed air. The heat sink can be modeled by approximating the radiant energy coming from the other reflector surfaces between wavelengths 0.3 and 2.0 μm with a constant heat flux. Additionally, 3D effects may need to be included in the heat sink approximation. The heat sink can also be equally applied to the external lens surface, or throughout the lens volume. Alternatively, a more accurate heat sink distribution can be imposed if the light intensity hitting the PC lens can be calculated with a ray-tracing analysis.

The effect of the lens heat sink on the reflector temperatures is shown in Table VII. The heat sink is applied on the inner surface, outer surface, and throughout the volume of the PC lens. All 3 methods yield similar results; the smoothest transition of the temperature through the lens is obtained with the volume-based heat sink. A surface heat sink of 1600 W/m^2, or equivalently a volume heat sink of 640,000 W/m^3 for a 2.5-mm thick lens, is required to bring the lens temperature predictions closer to the experimental data. The remaining reflector surfaces are minimally affected by the addition of the heat sink in the model.

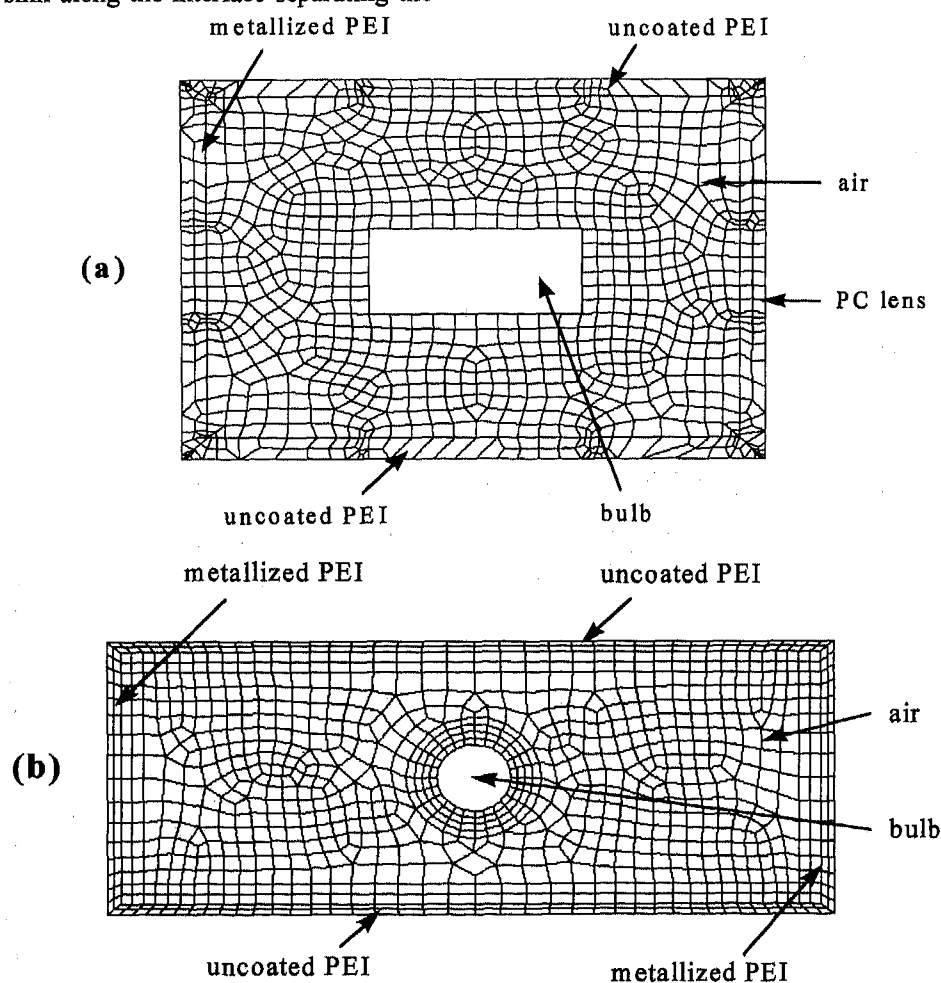


Figure 4: Finite-element mesh for the cross-sections (a) perpendicular to the lens and (b) parallel to the lens.

Table VII: Effect of the lens heat sink on the reflector temperatures (°C).

PERPENDICULAR SECTION	TOP/IN	TOP/OUT	BOTTOM/IN	BOTTOM/OUT	LENS/IN	LENS/OUT
No heat sink	255	215	240	205	200	180
Inner 1,600 W/m^2	244	207	222	190	159	135
Outer 1,600 W/m^2	245	207	224	192	168	126
Volume 640,000 W/m^2	245	207	223	191	164	131

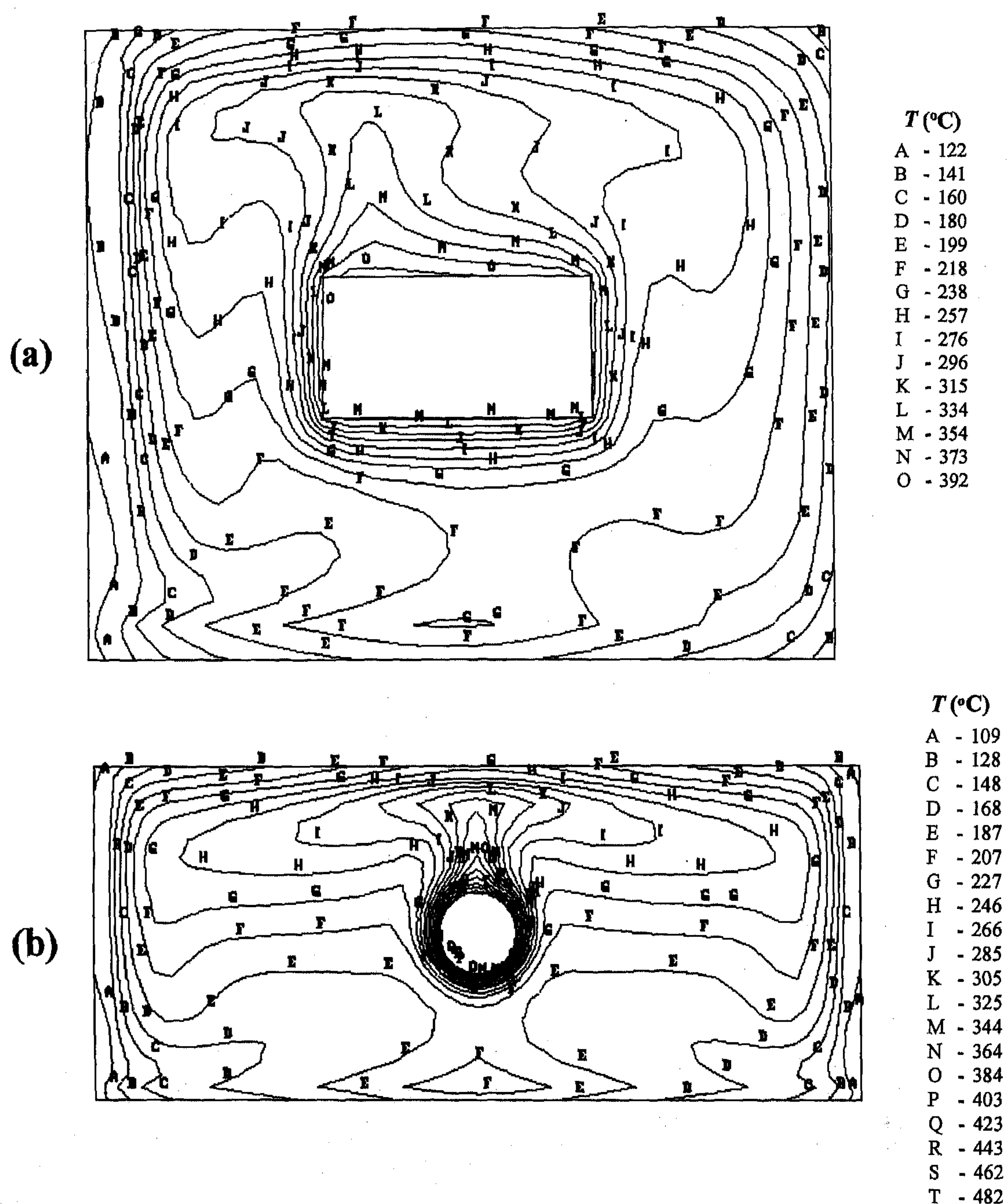


Figure 5: Temperature distribution in the cross-sections (a) perpendicular to the lens and (b) parallel to the lens.

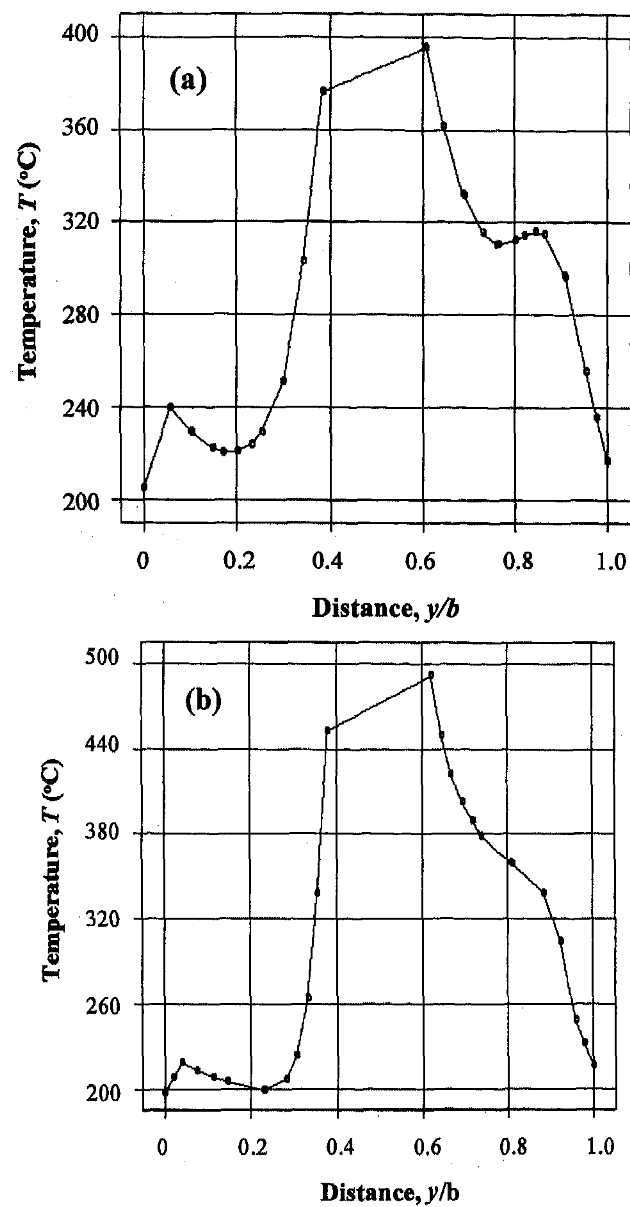


Figure 6. Temperature distribution along the vertical centerline in the cross-sections (a) perpendicular to the lens and (b) parallel to the lens.

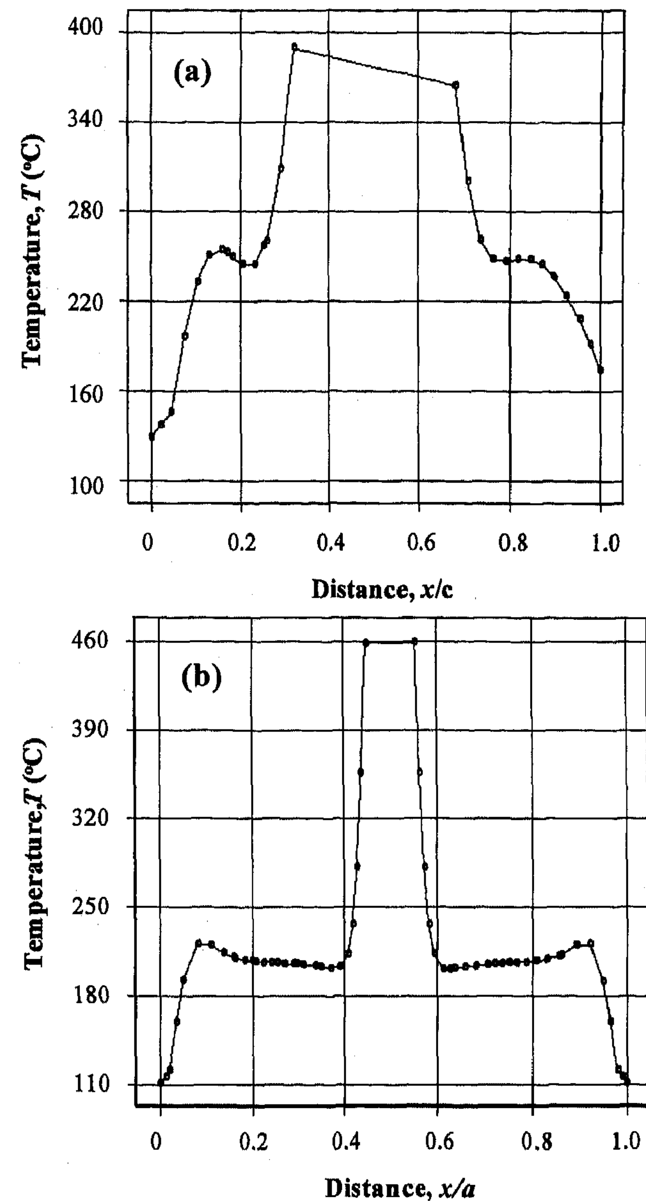


Figure 7: Temperature distribution along the horizontal centerline in the cross-sections (a) perpendicular to the lens and (b) parallel to the lens.

Table VIII: Comparison of sensitivity analysis predictions with baseline case results (°C).

MODEL	TOP/INSIDE	TOP/OUTSIDE	BOTTOM/INSIDE	BOTTOM/OUTSIDE
Base (Perpendicular)	255	215	240	205
Conduction/Convection	700	650	220	210
External Convection	440	410	460	435
Higher Conductivity	220	210	210	200
Higher Applied Flux	277	233	291	241
Lower Ambient Temp	244	207	227	193

SENSITIVITY STUDIES -- As a demonstration of using the thermal model for material formulation changes and reflector design, the effects of changing some of the heat transfer mechanisms, material properties, and operating conditions are investigated next. The baseline case selected for the sensitivity studies is the gray-diffuse analysis of the cross-section perpendicular to the lens described in **Domain Discretization**. The predictions of the various heat transfer models are compared with the baseline temperatures in Table VIII.

CONDUCTION/CONVECTION MODEL -- Radiative heat transfer is a highly nonlinear phenomenon; therefore numerical simulations exhibit convergence difficulties associated with the dependence of the radiative heat flux on T^4, where T is the absolute surface temperature. Because of this complexity, the first simulation only considered heat conduction and convection, and its solution was used as an initial condition for the other analyses with radiation. The contribution of radiation heat transfer to the thermal distribution is examined by comparing the results obtained with and without radiation. Comparison of the conduction/convection model with the base case in Table VIII shows that the radiative heat transfer tends to render the temperature distribution inside the enclosure more uniform, significantly lowering the upper shelf temperatures by 445C and slightly raising the bottom shelf temperature by 20C. Therefore, radiation must be included in the heat transfer model to accurately predict the thermal performance of any lamp reflector.

EXTERIOR CONVECTION MODEL -- Initially, the boundary condition on the exterior of the housing was assumed to be natural convection only. However, the combination of high temperatures along the top surface, the high emissivity of PEI, and the low convective heat-transfer coefficients in the non-circulating oven, suggested that radiative heat transfer might be an important energy transfer mechanism. Hand calculations confirmed the relative importance of the external radiation to the ambient, which was then included in the numerical simulation. The results obtained without external radiation are shown in Table VIII for comparison. Both the top and bottom shelves in the external convection model experience dramatic temperature increases of about 185C and 195C, respectively. Another indication of the importance of external radiation is the resulting increase in temperature gradient through the shelves, which is closer to the experimental data.

HIGHER CONDUCTIVITY MODEL -- To simulate a different reflector material, the thermal conductivity of the solid walls is increased by a factor of 5 to 1.0 W/m K. Compared with the baseline, the temperature difference across the thickness of the shelf is reduced by about 30C on the top and 25C on the bottom shelves as shown in Table VIII. The temperature sensitivity with respect to the material thermal conductivity is only modest because the heat removal from the enclosure is limited by convection and radiation on the outside surface, not by conduction through the plastic.

HIGHER APPLIED HEAT-FLUX MODEL -- The effect of a higher heat flux was investigated by increasing the heat flux in the perpendicular cross-section from 8,000 to 10,000 W/m^2. Table VIII shows that the temperatures in the perpendicular cross-section are very sensitive to changes in the heat flux in agreement with the experimental observations. The temperatures rise by about 20C and 40C for the top and bottom shelves. In contrast, Table IX shows that the parallel cross-section is not as sensitive to a similar change in the applied heat flux. A 1,500 W/m^2 decrease changes the temperatures by less than 20C. Most likely, the decreased sensitivity in the parallel case is due to the larger cross-sectional dimensions.

LOWER AMBIENT TEMPERATURE MODEL -- To gauge the effect of external conditions on the temperature predictions, an analysis was performed with a lower ambient temperature, $T_\infty = 25$C. The results show that the thermal model is not very sensitive to the external conditions; Table VIII shows that a decrease of about 10C is observed in the overall temperature distribution in agreement with the experimental observations. As the external temperature is raised closer to the temperature of the heated reflector walls, it is expected that the cooling ability of the external environment decreases and the temperatures will be much more strongly affected.

CONCLUSIONS

A thermal analysis of an automotive fog lamp reflector that includes conduction, convection, and radiation heat transfer within the lens and reflector assembly has been developed based on the FIDAP CFD analysis package. Comparison of the temperature predictions with and without thermal radiation included in the model confirms that radiative heat transfer has a significant effect, decreasing the temperature of the top shelf and increasing the temperature of the bottom shelf. Radiation loss to the external environment also plays an important role, decreasing the overall temperature distribution by as much as 200C. In general, the temperature predicted by the 2D models is in good agreement with the experiments. However, the predicted temperatures are not able to match the measurements for all the reflector surfaces simultaneously. One reason for the temperature being too high at the top shelf and too low at the bottom shelf could be the uniform heat flux boundary condition applied at the bulb surface. In reality, the heat flux from the light bulb is higher at the top and front bulb surfaces than at the bottom surface. Another contributing factor is the deficiency of using a 2D model to predict 3D

fields. For example, the location of the hot spot on the top shelf may be incorrectly predicted because of missing radiant exchange between the bulb and reflector in the third direction. The transparency of the PC lens is approximated in the present model with a heat sink applied throughout the lens volume, which is required to properly predict the temperature distribution in the lens. At the material selection and design stage of a new lighting application, there will be no temperature data for the light bulb, and it is necessary to establish a systematic methodology for estimating the appropriate 2D heat flux and heat sink. For the analyses performed in this study, the value of the heat flux emitted from the light bulb has been estimated based on the reflector geometry and the total power input. The heat sink is based on the radiant energy emitted from the tungsten filament, reflected with the metalized surfaces and transmitted by the PC lens between wavelengths of 0.3 and 3.0 mm, where PC is transparent.

This work has also demonstrated that the thermal model can be used to guide the material selection and design of lamp reflectors. Sensitivity studies show that increasing the thermal conductivity of the reflector does not appreciably change the overall temperature distribution. The predicted sensitivity of the reflector temperatures to normal power fluctuations or uncertainties in the 2D thermal model suggest that the softening temperature of the material should be at least 50C higher than the predicted maximum values at standard operating conditions. The results also show that the thermal performance of the fog lamp reflector is not very sensitive to changes in the external conditions; however, it is expected that the cooling ability of the external environment will decrease as the ambient temperature is raised closer to the temperature of the heated reflector.

# # # # #

ACKNOWLEDGMENTS

This program was supported by GE Plastics. Extensive technical discussions related to the fog lamp application with Ken Sherman, Brian Ahrens, and Keith DuPont from GE Plastics in Southfield, MI, and with Bob Shay and Wit Bushko from GE Corporate Research and Development are also greatly appreciated.

REFERENCES

1. Siegel, R. and Howell, J.R., *Thermal Radiation Heat Transfer*, 3rd ed., Hemisphere, London (1992).
2. "Optical Properties," in *Encyclopedia of Polymer Science and Engineering*, **V10**, John-Wiley & Sons, (1987).
3. *Fused Quartz Properties and Usage Guide*, General Electric Company (1995).
4. Sherman, K.C., "Infrared Transparent Reflector Materials," *SAE Technical Paper Series* **950556**: *Advances in Automotive Plastic Components and Technology (SP-1099)*, International Congress and Exposition, Detroit, Michigan (1995).
5. *FIDAP Theory Manual*, Fluid Dynamics International, Inc., Evanston, IL (1993).
6. *GE Plastics Engineering Design Database and Design Guide*, GE Plastics, Pittsfield, MA (1989).
7. Chapman, A. J., *Heat Transfer*, 4th ed., Macmillan, New York (1984).
8. Schlichting, H., *Boundary Layer Theory*, 4th ed., McGraw-Hill, New York (1960).

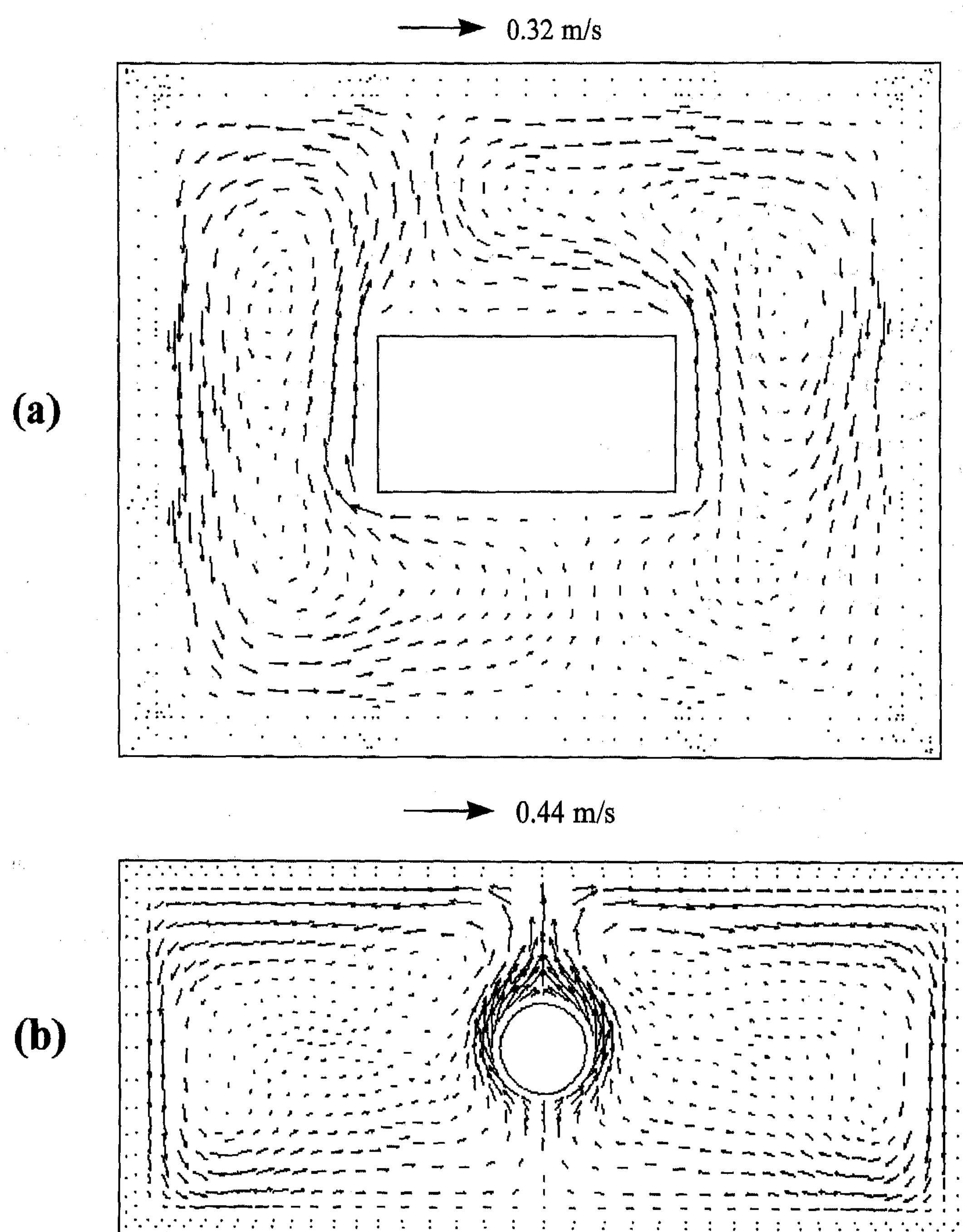


Figure 8: Velocity vectors for air flow in the cross-sections (a) perpendicular to the lens and (b) parallel to the lens.

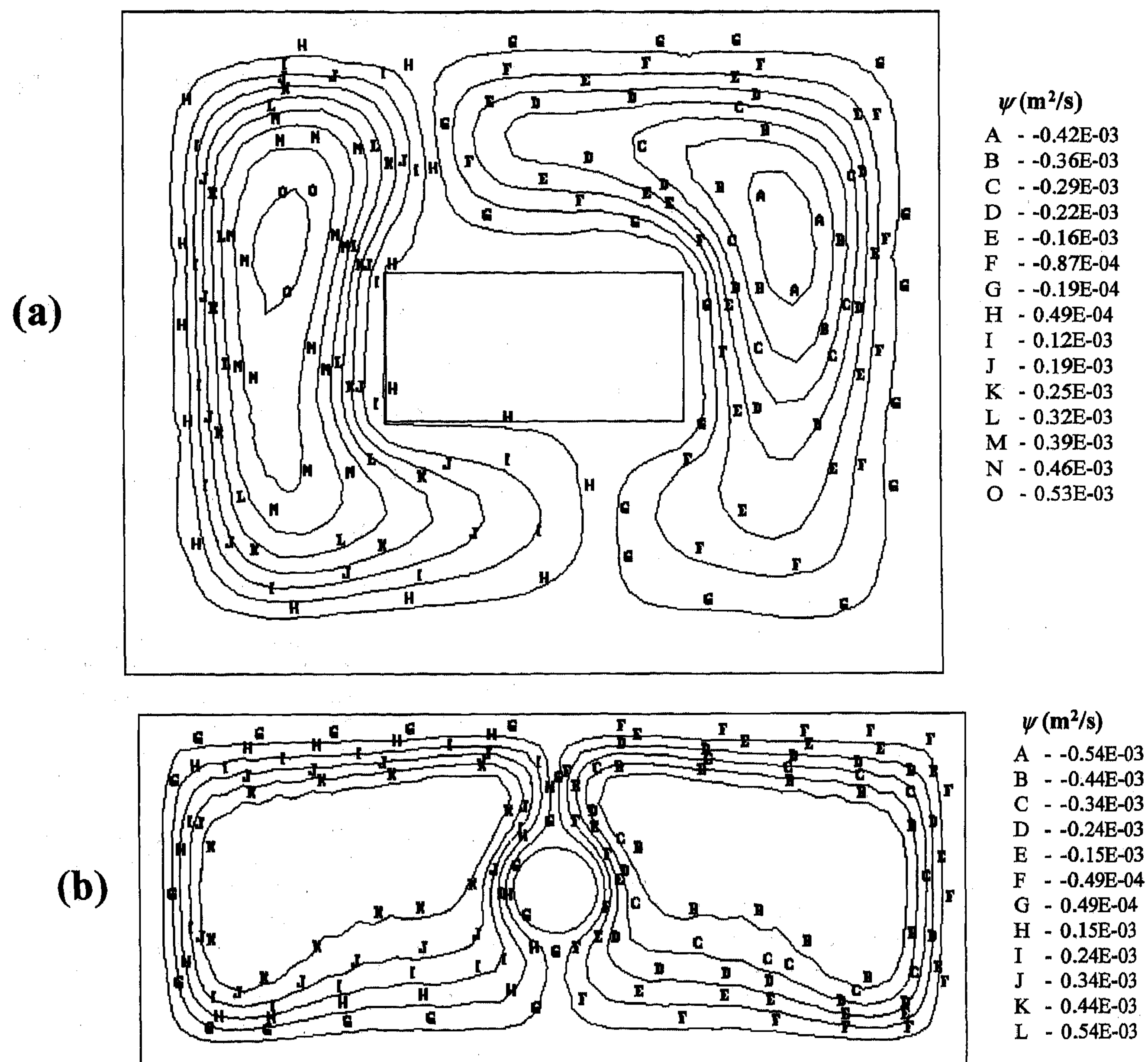


Figure 9. Streamlines for air flow in the cross-sections (a) perpendicular to the lens and (b) parallel to the lens.

Table IX. Reflector temperatures in the parallel cross-section for 2 values of the applied heat flux.

Q_A (W/m^2)	TOP/IN	TOP/OUT	BOTTOM/IN	BOTTOM/OUT	WALL/IN	WALL/OUT
17,500	248	216	219	197	122	112
16,000	236	207	206	187	117	108

970913

World Harmonization and Procedures for Lighting and Signaling Products

Guy Dorleans
Valeo Vision

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Lighting and signaling regulations vary considerably around the world. There are 3 main regulatory regions: the United States, Europe and Japan. Rules vary in scope as well as in strictly technical requirements. The ECE has published the most comprehensive regulations, covering even headlamp leveling and cleaning devices. Japan has, for the time being, the least extensive ones.

Standard lamp Regulations are pretty much alike, with the center high mounted stop lamp being the most notable exception. The most pronounced differences are for headlamps: all values above the horizon line are lower in Europe where being dazzled is not accepted. The cutoff line is also sharper in Europe, where visual aim has been common practice for 70 years.

Europe specifies amber rear side markers and amber rear turn signals but prohibits red contour markings as well as rear turn signals and red rear side markers.

Differences are more acute in the field of procedures. In the US, components are not subject to previous approval by the authorities as in Europe. The NHTSA checks *a posteriori* the compliance of those products which have caused complaint or damage, but the general rule consists of trusting the automobile manufacturer as long as no adverse event has occurred.

Those manufacturers who need to export have to pay high costs for tool duplication and administration. The rationale for this is not quite convincing since one has to admit that each country is consistent on its own territory.

INTRODUCTION

Laws relative to lighting and signaling have been developed over time, country by country. The need for harmonization has not historically been an important criterion. Trade resulting from the import and export of automobiles, or from tourism, was relatively insignificant until the middle of the 1950s. The first technical regulations regarding lighting and signaling appeared in Europe and in North America between 1920 and 1930. Traffic conditions, traditions and chance resulted naturally in laws which were framed differently, though each system had its own specific coherency.

Geography explains certain differences. The vast American spaces led to the creation of wider roads, with larger areas between lanes. This is often thought to be the explanation for the higher lighting levels permitted in the central zone above the low-beam cutoff. The federal structure of the United States and Canada left low-beam regulation to the legislature of each state, which explains the diversity of adjustment methods. Yellow headlamps in France can be explained by the now obsolete need to be able to identify the nationality of a car at night. And daylight running lamps appeared naturally in countries closer to the pole, where twilight periods are longer.

Pure chance, or history, are the reasons for driving on the right or left, the source of so much duplication of effort and investment for the industry. But nobody knows why certain functions such as identification lamps were invented in the USA or Canada or speed indicator lights in Japan, while they remain unknown elsewhere.

Legislative customs leave their mark, and sometimes their weight, on the law. Two doctrines exist: : auto-certification in Anglo-Saxon countries and type approval in Europe and now in Japan - with a corollary fraught with consequences for technical progress: the prohibition by omission which is the consequence of the latter system is an *a priori* obstruction to the introduction of new techniques which are not the subject of any law or regulation.

European texts are thus much more voluminous than their equivalents in the USA or Canada. Take the example of position lamps and parking lamps (European denominations) and compare this to American *parking lamps* or Japanese *clearance lamps.*

ECE: R7 and 77 respectively total 34 and 21 pages, while the 3 pages of R48 specify the installation of these functions.

USA: Regulation FMVSS 108 deals with the question in 14 lines, but incorporates by reference SAE J 222 and J 585e.

Japan: The Blue Book accords 49 lines in its September 1995 issue.

The incorporation of regulations by reference alleviates official texts by dispensing governments from such technico-legislative tasks.

GENERAL PRINCIPLES

a) Prohibition by omission in Europe and Japan has been elevated to a legal principle. The result is a considerable hindrance to progress, since everything which is not permitted by the law is deemed forbidden. This was for example the case for gas discharge light sources and the headlights that incorporate them. In the USA and Canada, the opposite principle applies; a function is permitted as long as its utilization is not shown to be harmful.

b) Each country has regulations which establish the minimum and maximum values for each low beam or high beam measurement point. Certain headlights fitted with very small parabola and HB2 or H4 bulbs (100X165 mm) are legal, but are close to the legal minimum. The legislation guarantees however that their utilization a) is technically feasible and b) ensures minimum but adequate safety.

c) In Europe, the performance of bulbs is dissociated from the lighting functions that use them. Bulbs are the subject of separate legislation, while headlamps and signal lamps are measured under standard laboratory conditions which are a long way from the normal usage of the vehicle.

d) US legislation is applicable to all vehicles or vehicle parts in the state where they are sold. In Europe, less demanding regulations, known as conformity of production, "COP", are applied to traded products.

Comparing statutory specifications is difficult in absolute terms, since the manufacturer has considerable leeway between required minimum and maximum values. In this investigation we have chosen to compare products which are comparable, by selecting headlamps designed for vehicles which are identical or very similar in dimension. We have excluded three sources of divergence associated with the utilization of the vehicle :

i) The voltage available at the bulb terminals, which varies considerably:

- new / used, cold / warm alternator, connectors, battery,
- engine idling or running at maximum revs, or not running,
- few or many electricity consumers in use simultaneously,
- left or right, rear lamp or headlamp.

It is not unusual to find voltages ranging between extremes of 15 to 9 volts, and in normal practice from 14 to 11 volts on a single vehicle. This is a factor greater than 2 for all photometric values.

ii) The maintenance of the vehicle. Dirt can cause a loss of up to 80% in the effectiveness of lighting systems. The age of the vehicle too is a decisive factor, as wind pollution affects all reflective surfaces over time however carefully the lighting units are sealed. The slow degradation of the light sources also is a contributory factor, and for all these reasons it is not unusual to observe a loss of 50% in the efficacy of lighting and signaling units after six years of use.

iii) Vibration often causes the vertical misaim of low beams. In this case any comparison of regulations becomes meaningless.

In the remainder of this investigation, we have therefore chosen to examine standard headlamps or signal lamps which are new, clean, powered at identical voltages, well-aimed and incorporating standard light sources as defined in ECE R37. Figures 1 thru 4 concern a single side of the automobile.

Low Beams

Figures 1 thru 4 were generated by computer from light sources powered at 12.8 volts. These curves represent the mean isocandelas of three identical headlamps, selected from a production run, fitted with the same bulb, used successively in each lamp. Relative to European laboratory practice, the 12.8 volt voltage increases the photometric values by approximately 25%, but it was necessary to choose identical voltages in order to compare products.

The left-hand cutoff was situated at 0.4° below the horizon line HH for all three products, in compliance with US legislation. This setting is higher than that specified by ECE R48 Rev1 by approximately 0.3°, which came to artificially increasing the glare levels tolerated in Europe. Here too the need to compare like to like took precedence.

Figure 1a is the mean isocandela curve of a European complex shape headlamp. Its opening was 159X93 mm and it was fitted with an H7 bulb producing a flux of 1100 lumens as defined by regulation ECE R37. This curve represents the average of three same side headlamps. Figure 1b also represents the average of three headlamps on the same side of the same automobile, but on the version sold in the US.

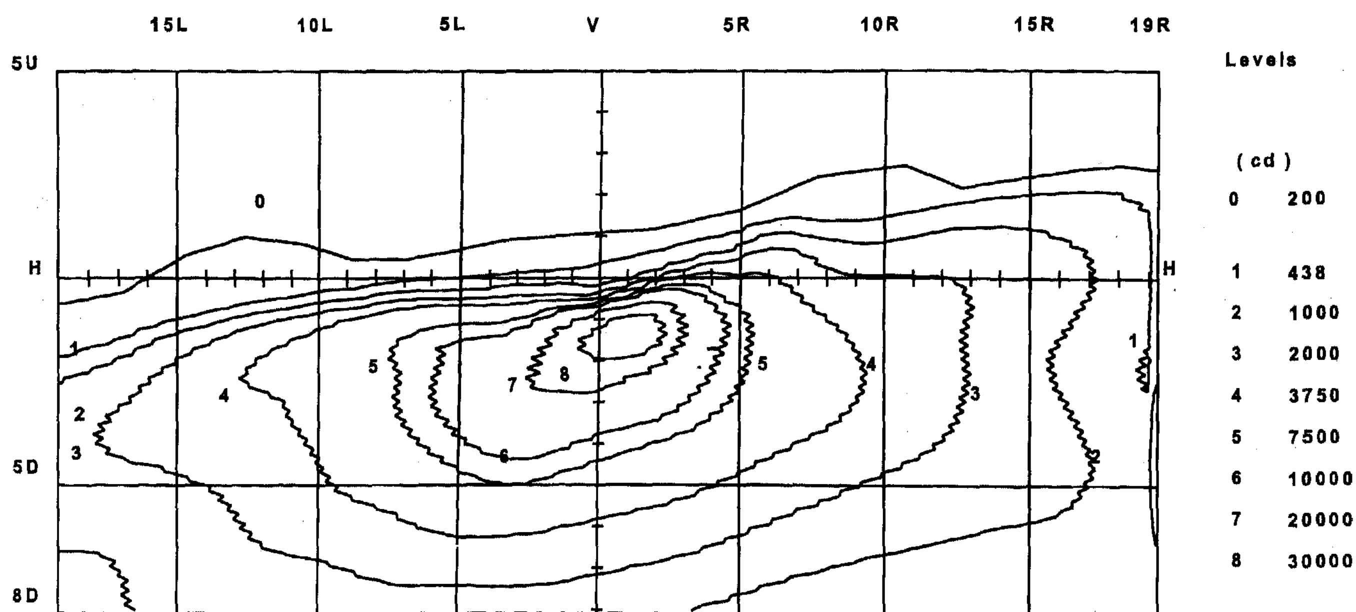


Fig 1a - Low Beam ECE MAX = 38808 cd

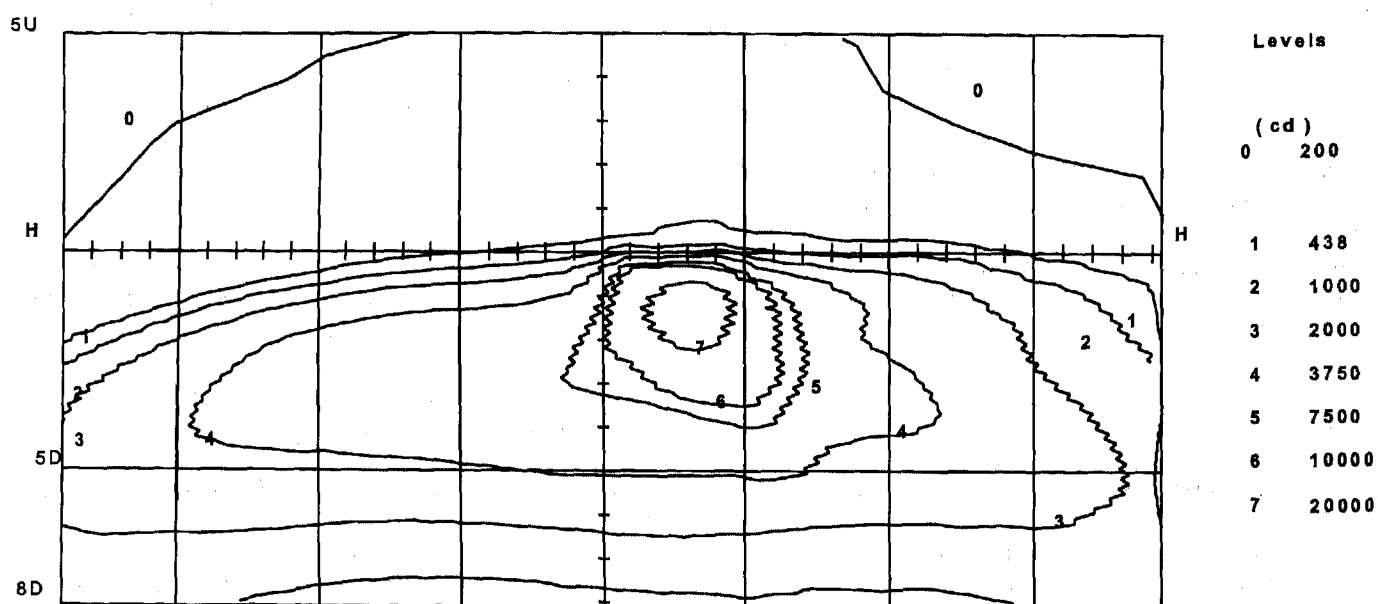


Fig 1b - Low Beam US MAX = 28474 cd

Figure 1

Headlamp 1a satisfies the "COP" photometric prescriptions of ECE R8 Rev. 4. Although "mis-aimed" upward, its maximum intensity reaches 38,800 cd, and is located at 1.3D-1R. The 10,000 cd contour extends from 6L to 5R, and the 2,000 cd contour from 18L to 13R.

Headlamp 1b satisfies the photometric requirements of US federal regulation 108 table 17a. Its hot spot reaches 28500 cd and is located at 1.8D-3R. The 10,000 cd contour extends from 0.5L to 6R, and the 2,000 cd contour from 23L to 18R.

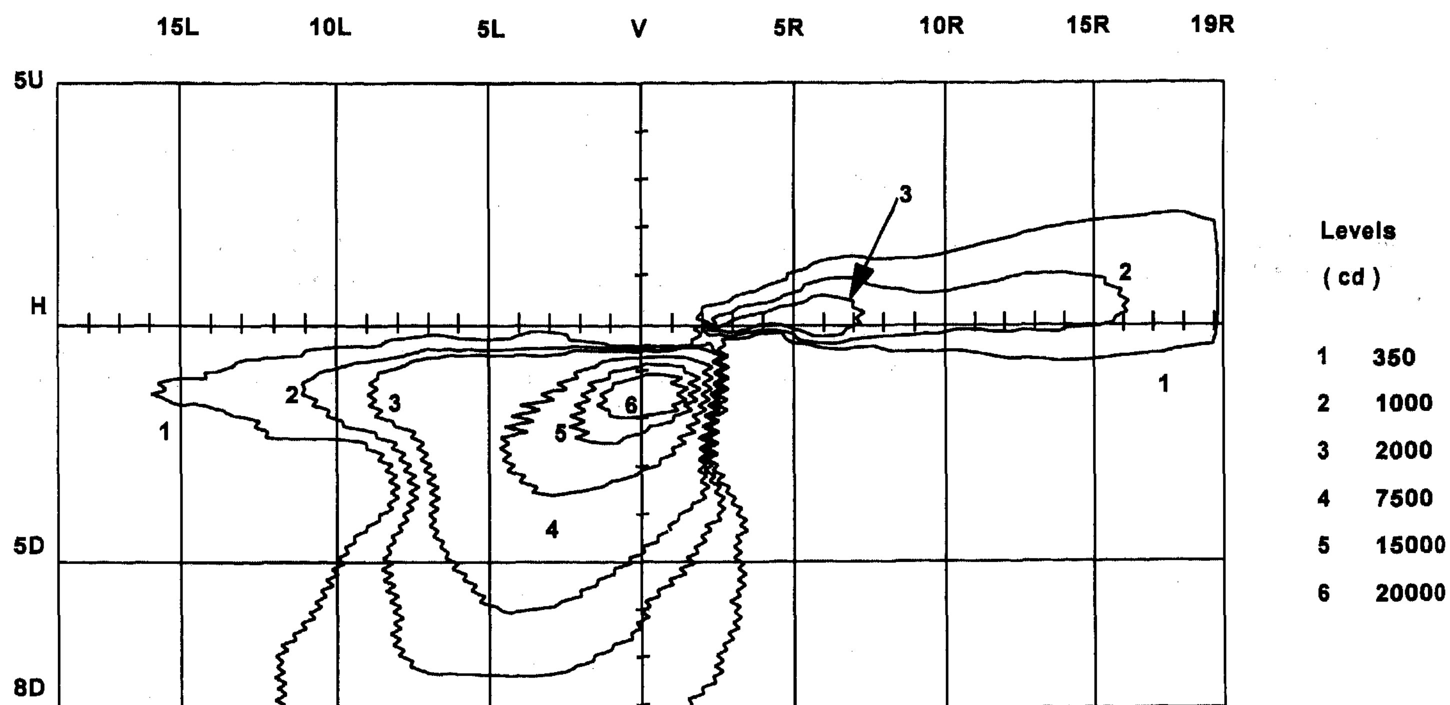


Fig 2a Low Beam (ECE - US) MAX = 26274 cd

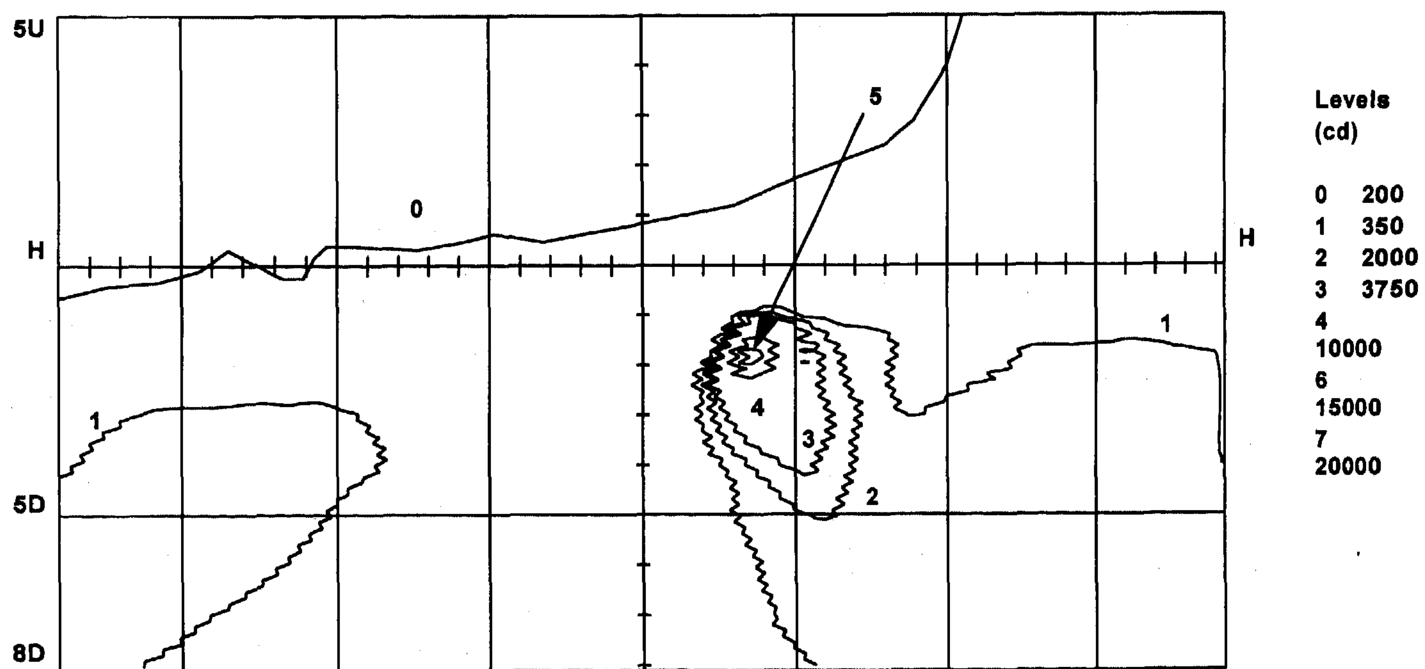


Fig 2b : Low Beam (US - ECE) MAX = 13699 cd

Figure 2

Figure 2a, which is computer-generated, is the differential beam obtained by subtracting the values of 1b from 1a. It shows the advantage of the European beam over the US equivalent. The two products correspond to the regulations in force in each zone, and also to the choices made within the limit of those regulations by the engineers of the different manufacturers. As this is the same vehicle, tested and evaluated by the same persons with, it can be presumed, the same degree of objectivity, it can also be presumed that the essential differences between the two are the result of differences in legislation. The European beam shows the following advantages :

- in the upper right quarter, contours 2 and 3 give a far better legibility of road signs located on the shoulder. The difference can be explained by the absence of a high cutoff line at 15° in US beams, and a minimum of only 200 cd on the segment 1.5U/1R-3R.

- straight ahead of the vehicle, beyond 20m, the European beam is significantly more intense. The difference of 26,300 cd corresponds almost to the hot spot of headlamp 1b, which in driving terms means a 40% increase in the obstacle detection distance.

Figure 2b is the differential beam obtained by subtracting the values of 1a from those of 1b. It represents the advantages of the US beam over the European equivalent. The American beam shows the following advantages:

- the shoulder is much better illuminated to a distance of about 20 meters. Contour 5 reaches 12,500 cd and contains the hot spot of 13,700 cd. The American beam is generally wider than its European equivalent. The maximum limit at 20,000 cd at 0.5D/1.5R gives greater width to the flux .

- above the horizon line, contour 0 shows a better illumination of overhead road signs. The part of this line to the left of 5L is not generated by a legal requirement but is useful to read road signs installed on the left side, a situation which is quite common on country roads in the US.

The oncoming vehicle glare criterion is more important for Europeans. The point representing the statistically average position of the oncoming drivers' eyes is limited to 460 cd. American beams allow this zone to reach 1,000 cd along the line 0.5U/1.5L, which is twice the European values. This is the greatest legal difference between the two sides of the Atlantic. In 1998 Japan will legislate for an ECE beam for left-hand traffic, but without a cutoff.

Because such fundamental differences make special tooling inevitable, low beams are the most costly for the automobile industry. A world car today requires three separate low-beam tooling systems: ECE for left hand traffic, ECE for right-hand traffic (+Commonwealth, Japan), and USA. Depending on the techniques used, either the lenses, or the reflectors, or both, need to be duplicated to ensure worldwide market coverage.

To this we need to add the cost of specific aiming devices in the USA where non-mechanical aiming is not yet authorized.

High Beams

Figures 3a and 3b show the isocandela curves of a high beam fitted with an H7 bulb for the American and European versions of the same headlamp studied in the previous paragraph. The opening of the high beam reflector is 127 X 110 mm. The unit was initially aimed using the low beam, the lamps then being displaced horizontally by a distance equal to that separating the two H7 bulbs. There was no separate adjustment of the high beam. The two reflectors were mounted in the same dimensionally stable part.

See Figure 3

US legislation limits high beam intensity to 75,000 cd at HV. The European limit is higher - 112,500 cd per lamp. The differential beams in 3d naturally show the greater width of the beam designed for the USA, which shows two zones of greater illuminance located at approximately 5° on the HH line. This facilitates driving on winding roads.

Conversely, curve 3c confirms the stronger concentration of the European high beam, which corresponds perhaps to the habit of driving faster at night.

Even so the legislative differences concerning high beams are much less problematical than those of low beams. The window of compatibility is sufficiently large that no specific study is necessary. In economic terms that means identical tools.

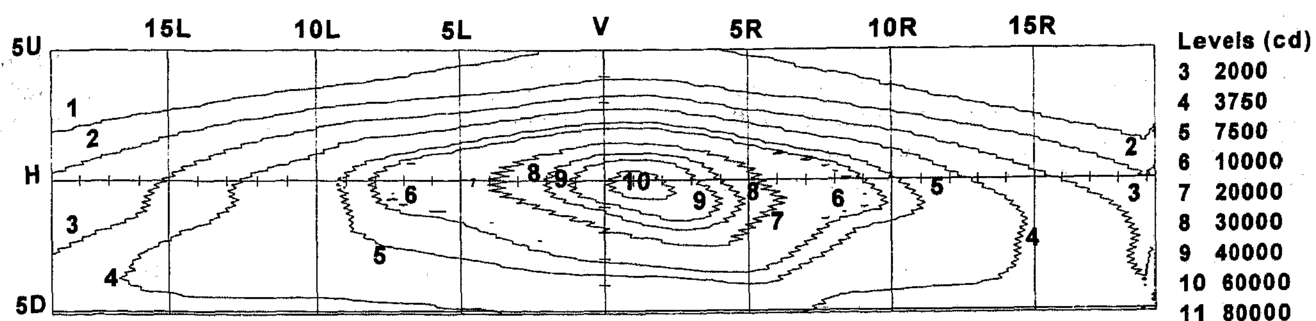


Fig 3a High beam US. MAX = 69976 cd

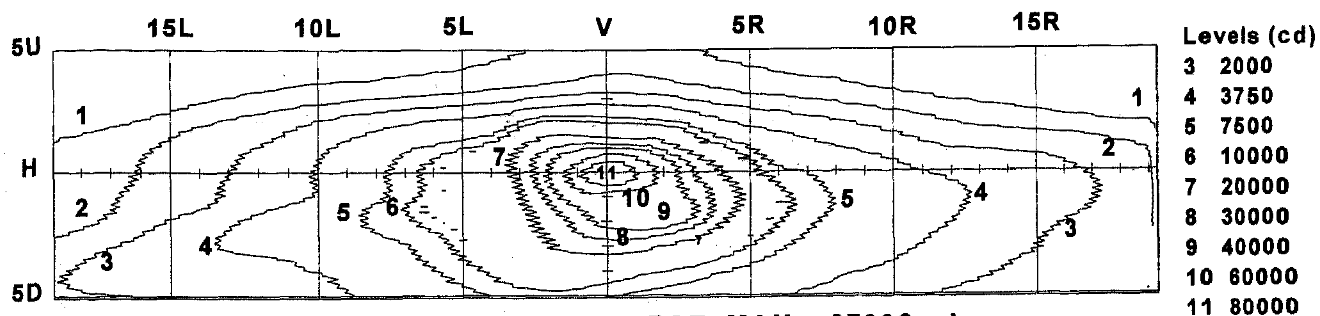


Fig 3b High beam ECE. MAX = 97336 cd

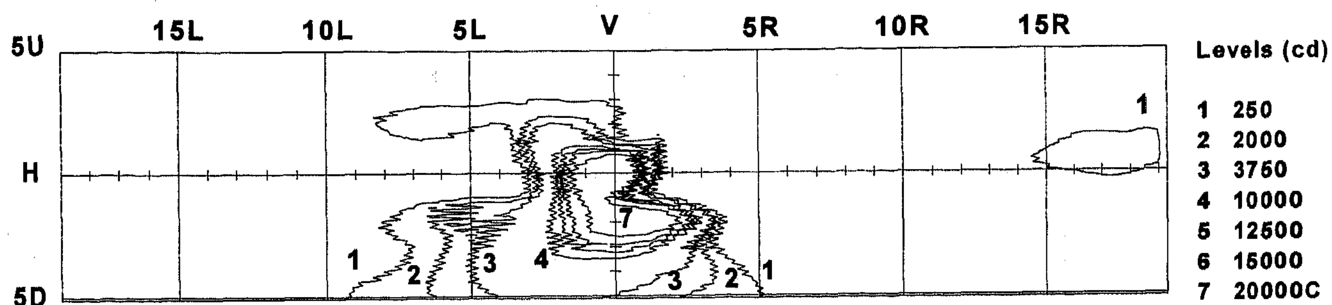


Fig 3c High beam (ECE - US). MAX = 42276cd

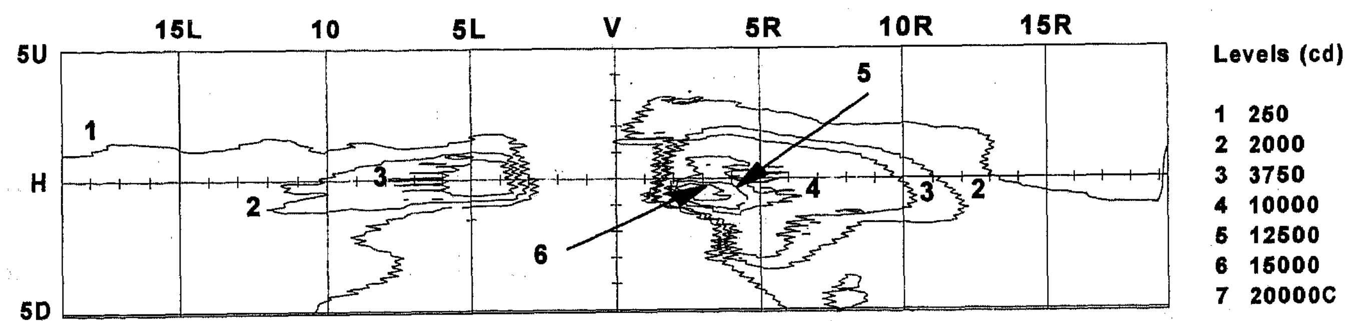


Fig 3d High beam (US - ECE). MAX = 18359cd

Figure 3

Non-regulatory Divergences

Laws and regulations can only be effectively compared if we eliminate the choices imposed by the customer, and which are the consequence of his own subjective experience or of his specific idea of night-time driving. Figures 4a and 4b show the isocandelas of the low beams of two products which comply with the regulations but which are very different in their execution.

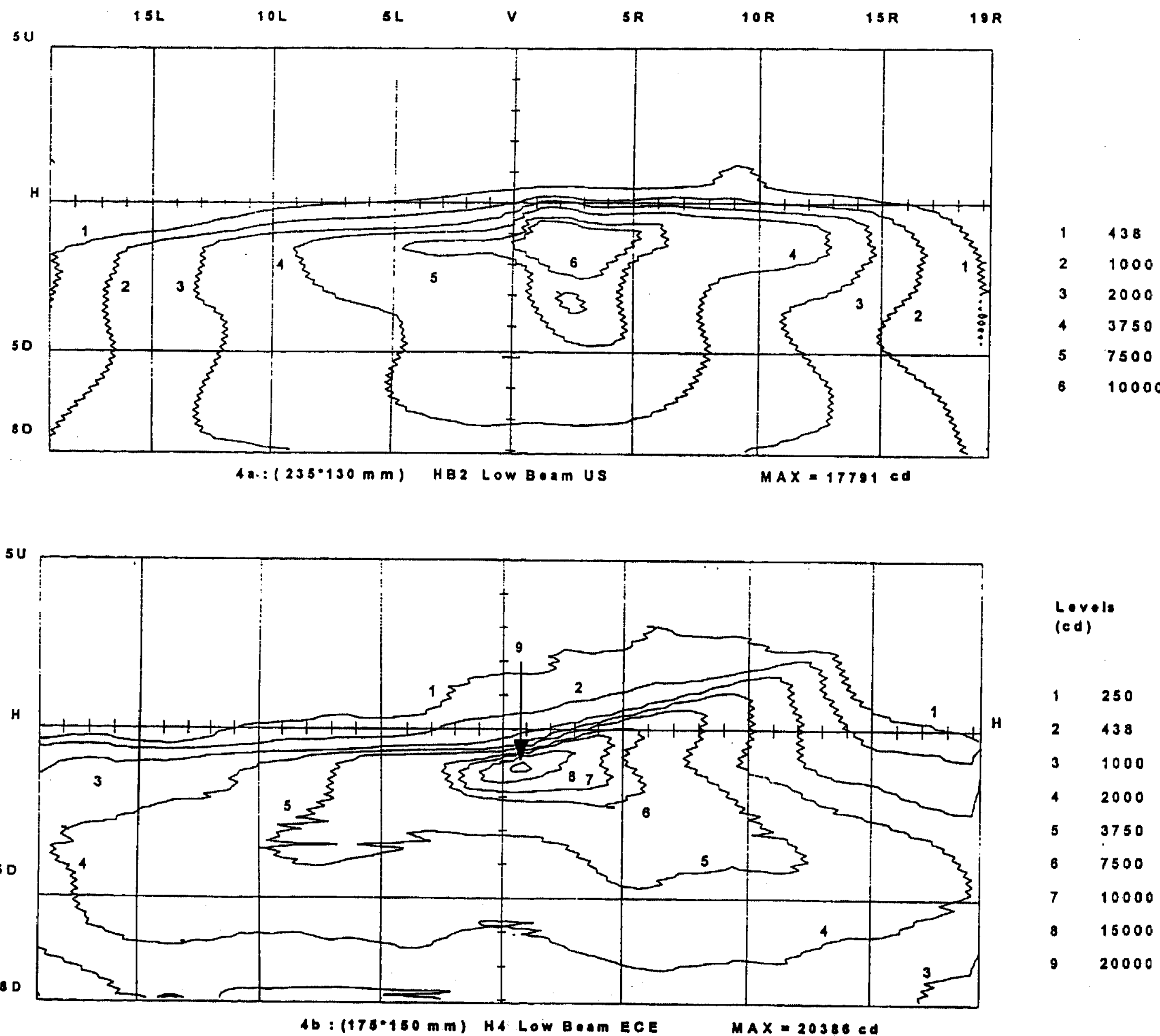


Figure 4

The lamp in 4a is a parabolic bifocal fitted with an HB2 bulb for a light truck sold in the USA. The technique is traditional, comprising a glass lens with optics, inclined at 27° in the horizontal section and at 11° in the vertical section. For this utility vehicle, the customer has opted for beam width and homogeneity, and Figure 4a shows regularly shaped isocandelas with a modest maximum of 17,800 cd.

Lamp 4b benefits from the most recent complex surface technology. It is designed for Europe, for a passenger car and is fitted with an H4 bulb with a completely smooth plastic lens. Its inclination is 40° and 17° respectively in the two sections. The reflector has two slight vertical ribs. The performance and the styling were the two main criteria of choice. The isocandela curves in Figure 4b are wider than those in 4a, and the maximum is also greater at 20,400 cd.

The H4 and HB2 bulbs which are fitted to these products differ only in terms of their mechanical characteristics. Their filaments are identical. Because of the angle of its lens, lamp 4b would be significantly less effective than lamp 4a without the complex surface technique. The deliberate choice of the designers has reversed the classification of the beams while nevertheless remaining within the legislative framework.

Such non-statutory divergences too often confuse the debate on international low beam harmonization, since they introduce subjective, or even emotional criteria. The choice which determined the structure of our investigation in paragraphs 3 thru 5 eliminates this variability.

Signal Lamps

Signaling functions are particular in that they are only partly uniform even as regards their definition.

Table 5 shows these divergences:

X: means "no requirement"

Functions		Europe	USA Canada	Japan
Turn signal	Condition front	mandatory : 2	mandatory : 2	mandatory : 2
	Condition side	mandatory : 1L + 1R	x	mandatory : 1L + 1R
	Condition rear	mandatory : 2	mandatory : 2	mandatory : 2
	Height	350-1500	381-2108	2300 maxi
	Width	L/R >600 ext<400	x	ext<400
	Surface front & rear	x	50 sqcm min	> 20 sqcm
	Surface side	x	x	> 10 sqcm
	Color front	amber	amber	amber
	Color side	amber	x	amber
	Color rear	amber	red or amber	amber
	Photometry front	Cat 1 : 175-700	200-500 cd	x
	Photometry front	Cat 1a : 250-800	200-500 cd	x
	Photometry front	Cat 1b : 400-860	200-500 cd	x
			if more than 100 mm from LB or HB or Fog : > 500 cd	x
	Photometry rear	50-350	130-700 cd	x
	Visibility V	45°int, 80° ext	45°int, 80° ext	x
	Visibility H	15U, 15D	10U, 10D	x
Stops	Condition	mandatory : 2	mandatory : 2	mandatory : 2
	Height	350-1500	381-1829	2000 maxi
	Width	L/R >600 ext<400	x	ext<400
	Surface	x	50 sqcm mini	20 sqcm mini
	Color	red	red	red
	Photometry	60-185	80-300 cd	x
	Visibility V optical	15° int, 15°ext		x
	Visibility V surfacic		15°int, 15°ext	x
	Visibility H optical	45°int, 80° ext		x
	Visibility H surfacic		45° int, 80° ext	x

Table 5

	Condition	mandatory : 1	mandatory : 1	optional : 1
	Height	>850 mm	>76 mm above window	< 200 cm
	Width	center	center	center
	Surface	x	29 sqcm mini	20 sqcm mini
CHMSL	Color	red	red	red
or	Photometry	25-80 (external)	25-160 cd (in situ)	25-160 (external)
S3	Visibility V optical	10U-5D	x	x
	Visibility V surfacic	x	x	x
	Visibility H optical	10R-10L	x	x
	Visibility H surfacic	x	45°int, 15°ext	x

	Condition	mandatory : 2	mandatory : 2	mandatory : 2
ECE :	Color front	white	white or amber	white to amber
Position	Color rear	red	red	red
lamps	Height	350-1500	381-1829	<2000
USA-Can :	Width	L/R >600 ext<400	x	ext<400
Parking	Surface	x	x	> 15 sqcm
lamps	photometry front	4-60 or 4-100	4-125	x
Japan :	Photometry rear	4-12	2-18 or 3,5-20 or 5-25	x
Clearance	Visibility H optical	45°int, 80° ext	x	x
lamps	Visibility V optical	15U-15D	x	x
	Visibility H surfacic	x	12,5 sqcm at 45¡ext	x

	Condition	optional 2+2	mandatory : 2+2	optional 2+2
	Color front	amber	amber	amber
	Color rear	red or amber	red	red or amber
	Height	250-1500	381-1524	<2000
Sidemarker	Length	1/3 median	x	dist front or rear>400 mm
lamps	Surface	x	x	> 10 sqcm
	Photometry	SM1 H0V0 =4	amber>0,62, red >0,25	x
	Photometry	SM2 = 0,62		x
	Visibility V optical	30° front, 30° rear	45° front & rear, X° int	x
	Visibility H optical	10U-10D	10U, 10D	x

	Condition	mandatory : 1 or 2	x	Optional : 1 or 2
	Height	250-1000	x	< 100 cm
Rear	Width	opposite to traffic, or =	x	> 100 mm apart from stop
Foglamp	Surface	< 140 sqcm	x	x
	Color	red	x	red
	Photometry	150-300	x	x
	Visibility V optical	5U-5D	x	x
	Visibility H optical	25L-25R	x	x

Function		Europe	USA Canada	Japan
	Condition	mandatory : 1 or 2	1	1
	Height	250-1200	x	> 1200 mm
	Width	x	x	x
Back-up	Surface	x	x	x
lamps	Color	white	white	white to yellow
	Photometry	80-300	80-300	5000 cd maxi
	H Visibility 2 lamps	30° int, 45° ext	x	ground impact at < 75 m
	H Visibility 1 lamp	45L-45R	x	x
	Visibility V	15U-15D	x	x

Table 5 (continued)

	Condition front	optional : 2	not regulated	not regulated
	Condition side	optional : 2	mandatory : 2	mandatory : 2
	Condition rear	mandatory : 2	mandatory : 2	mandatory : 2
	Height	250-900	381-1524	< 1500 mm
	Length, for side RR	L/3 maxi from the ends	x	dist front or rear>400 mm
	Width front or rear	L/R >600 ext<400	x	ext<400
	Surface	x	x	> 10 sqcm
	Color front	White	x	White or amber
Reflex Reflectors	Color rear	red	red	red
	Color side front	amber	amber	amber
	Color side rear	amber or red	red	red
	Photometry mcd/lux	>300(dia 200,100 sqcm)	>420(dia 254,77,5 sqcm)	>300(dia 120, 75 sqcm)
	Visibility rear & front	H = 30° int, 30°ext	x	x
	Visibility rear & front	V = 15U, 15D	x	x
	Visibility side	H = 45°front, 45°rear	x	x
	Visibility side	V = 15U, 15D	x	x

	Condition	optional : 2	optional : 2	optional : 2
	Height if high beam	250-1500 mm	< 864 mm	?
	Height if - headlamp	250-1500 mm		?
Daytime Running lamp	Width	L/R >600 ext<400	x	?
	Area	> 40 sqcm	x	?
	Color	White	White to yellow	White to yellow
	Photometry if high beam	400-800	< 7000 cd	?
	Photometry	400-800	500-3000	< 300 cd
	Visibility V optical	10U, 10D	x	?
	Visibility H optical	20L, 20R	x	?

	Condition	optional : 2	not regulated	optional : 2
	Height	250-1500 mm	x	?
	Width	L/R >600 ext<400	x	?
Front Foglamp	Color	White or yellow	x	White to yellow
	Photometry	ECE R 19	x	?
	Photometry	400-800	x	?
	Visibility V optical	10U, 10D	x	?

Table 5 (continued)

Japanese law has no mention of daylight running lamps, rear fog lights and front reflectors. US legislation nowhere refers to front and rear fog lights. ECE regulations fail to mention truck identification lamps while directives of the European Union, bogged down by translation into 11 different languages, continue officially to prohibit daylight running lamps, central high mounted stop lights and gas discharge light sources "by omission"...

It is not however the function of signal lamps that most hampers designers. In fact points of fundamental divergence between countries or groups of countries are relatively rare. Legislators could however simplify the task of exporters by concentrating on the following issues:

a Eliminate the minimum surface areas required for all functions in the USA and Japan,

b Authorize amber as the color of rear side position lamps in the USA and Canada,

c Harmonize the terminology for all functions,

d Authorize amber front position lamps in Europe,

e Eliminate all requirements regarding simultaneous use of the low beam with front fog lights, and of daylight running lamps with rear position lamps,

f Require the presence of rear window glass during photometry tests of CHMSLs in Europe and Japan.

g Eliminate the restrictive character of ECE R 37 regarding bulbs in Europe,

h Harmonize the two photometric variants of side position lamps in Europe.

The "Harmonization" group of the ECE working on angles of visibility should resolve certain of these contradictions.

CONCLUSIONS

None of the legal or regulatory differences of the different regions of the world seems to have a decisive influence on the safety of night-time traffic. It seems that the range of tolerance or adaptability of the manufacturers is at least as wide as the legal differences. Each country or group of countries has its own legislative coherency.

But these legal or regulatory differences are strong enough to necessitate the duplication of production tools, thus creating a financial and technical burden which is not justified by night-time driving. Therefore, it is necessary to harmonize legislation in such a way that we can focus on safe customer satisfaction, based on continuous technological innovation and standardization at most competitive cost.

REFERENCES

ECE R 48-02 Agreement concerning the adoption of uniform conditions of approval and reciprocal recognition of approval for motor vehicle equipment and parts, 1993.

US Code of Federal Regulations. Title 49 parts 571.108 & 564. Revised as of October 1st 1996.

Official Journal of European Communities. Council Directive n°92/53/EEC of 18 June 1992 amending Directive 70/156.

Automotive type approval handbook for Japanese certification. July 1994, revised as of September 1995.

Lichttechnische Entwicklung von Anforderungen an Kraftfahrzeugscheinwerfer. Dipl.-Ing. Joachim Damasky. June 1995.

Effects of the lateral position of low-beam headlamps on the perceived distances of vehicle. UMTRI

970935

Numerical Optimization for Design of an Automotive Cooling Fan

Jae-Ho Choi and Kwang-Yong Kim
INHA Univ.

Duck-Soo Chung
Doowon Climate Control Co., Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The numerical optimization techniques coupled with a three-dimensional Navier-Stokes solver have been developed to design an automotive cooling fan. The conjugate gradient method is used to look for the search direction and the golden section method is used for the one dimensional search. Concerning the constraints, exterior penalty functions are employed. In the applications of this numerical optimization technique to the design of an automotive cooling fan, using the object functions defined as an increase in the pressure coefficients and a ratio of production rate of turbulent kinetic energy to the pressure head, the final shapes are obtained after sixth and seventh changes of shape function, respectively.

INTRODUCTION

Recent developments in computer hardware have increased the utility of the numerical optimization techniques in aerodynamic design. While the inverse design methods[1] are efficient for the design of simple aerodynamic geometries of which ideal flowfields are prescribable, the numerical optimization methods[2] can be general design tools because of many advantages, such as the automated design capability, varieties of constraints, and multidisciplinary applications. This numerical optimization can obtain optimal design by minimization or maximization of object function, even without any information of ideal flowfields.

Item and Baysal[3] performed a supersonic wing-section optimization considering the viscous effects of the flow to increase the lift-to-drag ratio using two-dimensional analysis. And, Huan and Modi[4] reported on design of minimum drag bodies in incompressible laminar flow in order to obtain maximum static pressure rise with two-dimensional Navier-Stokes solver. Massardo and Satta[5] carried out a design optimization of axial flow compressor with a pitchline analysis. Most of the optimization methods reported so far, have used inviscid flow analysis or two-dimensional viscous flow analysis. However, the object function evaluated with three-dimensional viscous flow analysis is desirable for the numerical optimization of complicated shapes.

The present work reports on a numerical optimization of an automotive cooling fan, for which object function is evaluated with three-dimensional Navier-Stokes solver[6]. The exterior penalty method is used with the constraints for geometric design variables related to a radial distribution of sweep angle.

DESIGN FORMULATION

The optimization problem requires the definitions of design variables, object function and constraints. General optimization problems can be expressed as follows.

$$\text{minimize} \quad f(x)$$

$$\text{subject to} \quad \begin{array}{ll} x_i^l \le x_i \le x_i^u & i = 1,n \\ g_j(x) \le 0 & j = 1,m \\ h_k(x) = 0 & k = 1,l \end{array}$$

where, $x = \{x_1, x_2, x_3, x_4, \cdots, x_n\}^T$.

In the optimization algorithm, conjugate gradient method[7] which uses the gradient vector and the information on previous search direction is used to find the search direction. In order to minimize the object function along the search direction, the golden section method is employed in this work.

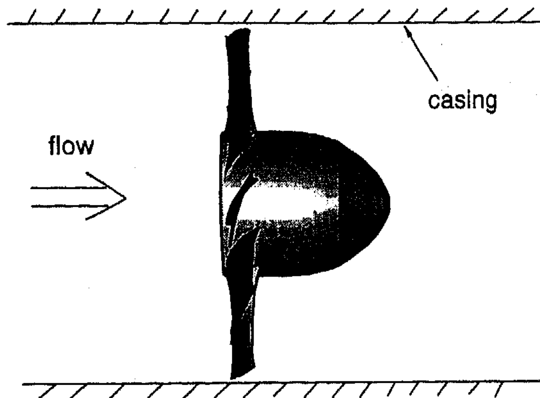


Figure 1. Automotive cooling fan model

In the design of an automotive cooling fan shown in Fig. 1, there are many geometric design parameters. The initial shape of the automotive cooling fan considered here has its own radial sweep angle distribution. The amount of sweep angle changed is defined by the following quadratic distribution function. Fig. 2 shows the design variables defined by sweep angles.

$$\gamma = a \cdot Rn^2 + b \cdot Rn + c$$

$$Rn = 0, \quad \gamma = 0$$
$$Rn = 0.5, \quad \gamma = \gamma_m$$
$$Rn = 1, \quad \gamma = \gamma_t$$

where, $Rn = \dfrac{(R - R_{HUB})}{(R_{TIP} - R_{HUB})}$.

So, in this work, two variables related to radial distribution of sweep angle are chosen as design variables; one(γ_m) is a sweep angle at the mean blade height, and the other(γ_t) is that at the tip.

In this work, the following two kinds of object function are applied to the optimization.

CASE 1 :

In case 1, the object function is chosen as an increase in the pressure coefficients between the leading edge of the fan and the outlet of duct. And, side constraints are imposed to the design variables. Thus, the numerical optimization problem of the case 1 can be stated as follows

maximize Δc_p

subject to $-0.5 \le \gamma_m \le 0.5$, $-1.0 \le \gamma_t \le 1.0$

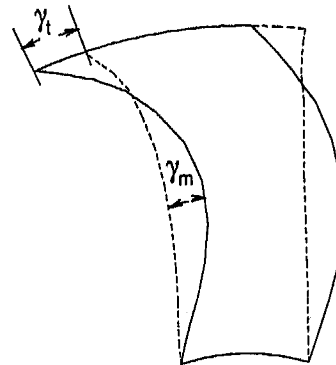


Figure 2. Definition of design variables

The pressure coefficient, c_p is defined by $(p - p_0)/\rho U_0^2$, where p_0 is the reference pressure and U_0 is the flow inlet velocity.

CASE 2 :

The object function of case 2, is chosen as a ratio of the production rate of turbulent kinetic energy(G) to the pressure head(H). And, no constraint is imposed to the design variables. The production rate of turbulent kinetic energy is considered to be related to the noise level.

minimize $\dfrac{c \cdot G}{H}$

where c is a weighting factor.

FLOW ANALYSIS

In the present optimization method, the flow analysis is repeatedly applied to the calculations of the object function and its derivatives, for the three-dimensional flow through a rotating axial-flow fan. The governing equations in Cartesian coordinates are transformed to a non-orthogonal curvilinear coordinate system. The governing transport equations are discretized by the finite volume approach[8] where the equations are integrated over a finite number of control volumes. As a numerical scheme for the convection terms, the linear upwind difference scheme is employed. The system of linear algebraic equations are solved by strongly implicit procedure(SIP)[9]. And SIMPLEC algorithm[10] as a velocity-pressure correction procedure is used in this work. In conjunction with this procedure, we employ the grid system which is not staggered. For the computation of turbulence, the present work employs k-ε model[11]. And, the empirical wall functions are used to provide the near wall boundary conditions for the momentum equations and the k-ε model.

The governing equations of elliptic type require boundary conditions at all surfaces of computational domain which consist of inlet, wall and periodic surfaces. At the inlet boundary, the axial velocities are assumed to be uniform, and the uniform turbulent kinetic energy and its dissipation rate are assumed. At the outlet boundaries, all axial gradients of the dependent variables are neglected. At the wall boundaries, no-slip condition and empirical turbulent wall law are employed. And, the periodic conditions are applied to all variables at the periodic boundaries.

RESULTS AND DISCUSSION

Fig. 3 shows the computational domain which is one ninth of whole domain. There is a motor cap in the rear of the fan. The hub-tip ratio is 0.418. The number of grid is 106×16×28. The length of the computational domain from inlet to leading edge of the blade is 2.5R, and that from trailing edge to outlet is 5R, where R is the radius of fan. And, the Reynolds number of the flow is 136,000. The effects of blade thickness are neglected. In the computation, the motor cap is treated as an obstacle, in which very large viscosity is given. The average axial velocity and the radius of casing are taken as velocity and length scales of the flow, respectively.

For the noise test, we also have carried out simple experiment. The experimental system is composed of microphone, sound level meter, FFT, and a personal computer. The measurements are made 3.3R upstream of the fan which is installed in the automobile.

The results of the case 1 are shown in Fig. 4 - 8. In Fig. 4 which shows the history of object function, it is found that the function increases steadily during the iterations. The design variables which are modified during the optimization procedure is shown in Fig. 5. It is very rare that the step size for the search along the direction is fixed in the optimization algorithm. Thus, the efficient algorithm to find the moving distance has a great role in the cost of optimization. Fig. 6 shows the history of scalar multiplier, that is the moving distance along the search direction.

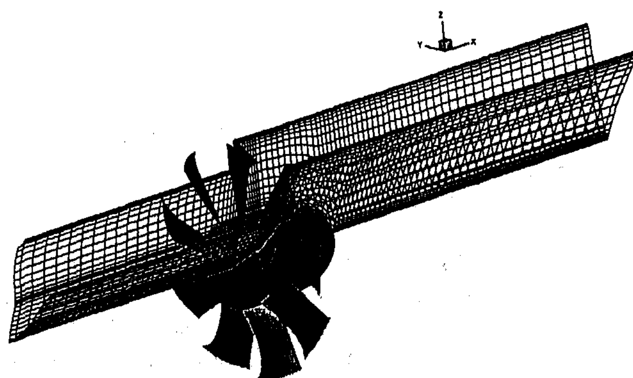


Figure 3. Computational domain with fan

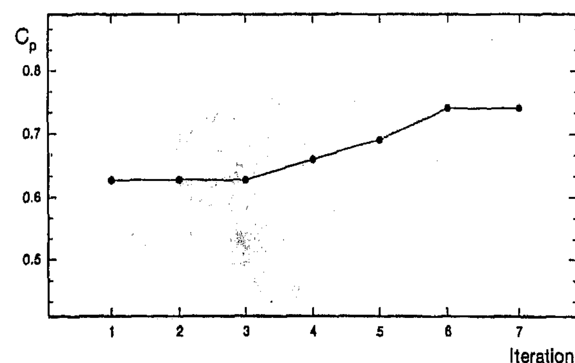


Figure 4. History of object function in the case 1

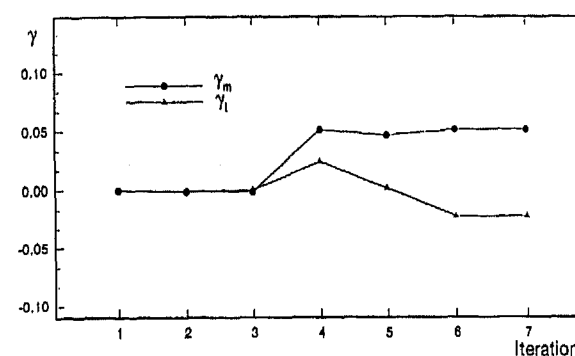


Figure 5. History of design variables in the case 1

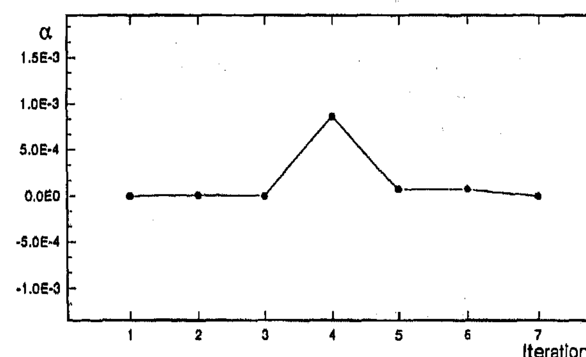


Figure 6. History of scalar multiplier in the case 1

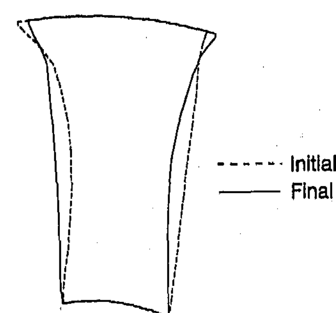


Figure 7. Initial and final shape of the blade in the case 1

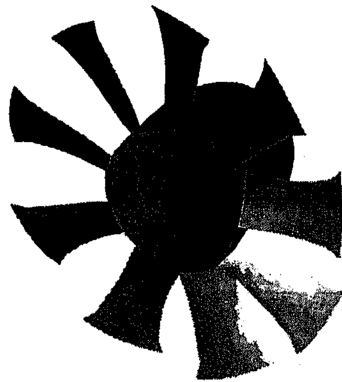


Figure 8. Final shape of the fan in the case 1

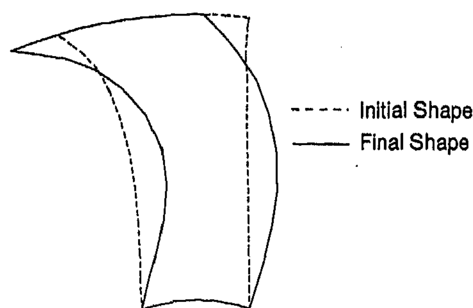


Figure 9. Initial and final shape of the blade in the case 2

After six design modifications, final shape of the blade is obtained. Fig. 7 compares initial and final shape of the blade. Fig. 8 shows the final shape of the fan in the case 1.

In the case 2, the object function is reduced from 107.9 to 89.6 after seventh update of design variables. Fig. 9 shows the initial and final shape of the blade, and the final shape of the fan is shown in Fig. 10.

Fig. 11 shows the noise spectrums of initial and final fans in the case 2, where the object function is defined as a function of the production rate of turbulent kinetic energy and the pressure head. In these results, the final fan gives the lower turbulent boundary noises and the lower discrete peaks. It seems that the noise reduction results from the suppression of relative turbulent level by minimizing the object function. The noise level of the final fan is about 4.5dB lower than that of initial fan in the region of flow rate between 1,550 m^3/h and 2,150 m^3/h. However, in the case 1, we could not obtain the better noise characteristics.

In the present work, only two design variables are used for this numerical optimization method. But, if we use more design variables related to the shape of the blade, this method can be more practical and efficient design tool for the automotive cooling fan.

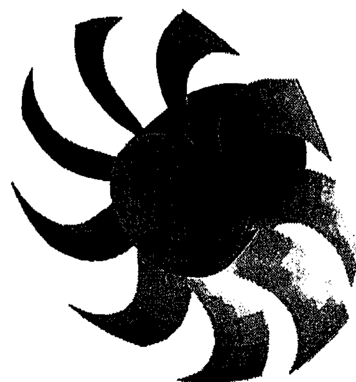


Figure 10. Final shape of the fan in the case 2

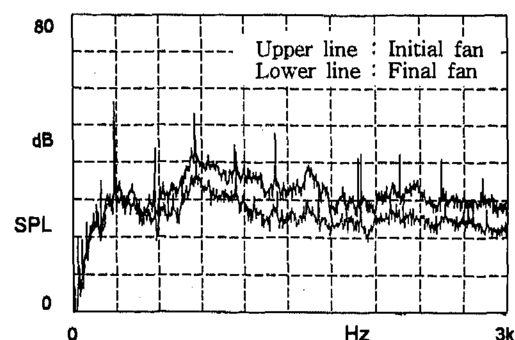


Figure 11. Noise spectrum

CONCLUSION

In the present work, the numerical optimization technique incorporated with three-dimensional Navier-Stokes solver is applied to the design of automotive cooling fan with two different object functions. The present optimization method with the conjugate gradient method obtains the final shapes after sixth and seventh updates of design variables, using the object functions defined as the increase in the pressure coefficient and the ratio of the production rate of turbulent kinetic energy to the pressure head, respectively. In the later case, we could obtain the final design with remarkable noise reduction.

REFERENCES

[1] Malone, J. B., Vadyak, J., and Sankar, L. N., Inverse Aerodynamic Design Method for Aircraft component, J. of Aircraft, 24(1)(1987), 8-9.

[2] Savu, G. and Trifu, O., 1991, On A Global Aerodynamic Optimization of A Civil Transport Aircraft, Proceeding of 3rd Inverse Design Concepts and Optimization in Engineering Sciences, Washington, D. C., 45-54.

Optimization in Engineering Sciences, Washington, D. C., 45-54.

[3] Item, C. C. and Baysal O., Wing-section Optimization for Supersonic Viscous Flow, Proceeding of the 1995 ASME International Mechanical Engineering Congress and Exposition, ASME FED-Vol.232(1995), San Francisco, 29-36.

[4] Huan, J. and Modi, V., Design of Minimum Drag Bodies in Incompressible Laminar Flow, Proceeding of the 1995 ASME International Mechanical Engineering Congress and Exposition, ASME FED-Vol.232(1995), San Francisco, 37-44.

[5] A. Massardo and A. Satta, Axial Flow Compressor Design Optimization: Part I -Pitchline Analysis and Multivariable Objective Function Influence, J. Of Turbomahinery, 112(1990), 399-404.

[6] Kim, K. Y., et al., Computational and Experimental Investigations on The Flow through An Axial Flow Fan, Proceedings of the 6th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery, 2(1996), Honolulu,190-198.

[7] Arora, J. S., Introduction to Optimum Design, McGraw-Hill(1989), 316.

[8] Peric, M., A Finite Volume Method for the Prediction of Three-Dimensional Fluid Flow in Complex Ducts, PhD Thesis, Imperial College, London(1982).

[9] Stone, H. L., Iterative Solution of Implicit Approximations of Multi-dimensional Partial Differential Equations, SIAM J. Num. Anal., 5(1968), 530.

[10] Van Doormal J. P. and Raithby, G. D., Enhancement of the SIMPLE Method for Predicting Incompressible Fluid Flows, Numerical Heat Transfer, 7 (1984), 147-163.

[11] Launder, B. E., and Spalding, D. B., The Numerical Computation of Turbulent Flows, Computer Methods in Applied Mechanics and Engineering, 3 (1974), 269-289.

970938

In-Vehicle Engine Coolant Void Fraction and De-aeration Monitoring Using a Computerized Electrical Conductivity Method

Ashok Garg, Richard S. Marano, Alex D. Colvin, and John M. Jakupco
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Instruments and analytical techniques are described for in-vehicle monitoring of amounts of air (void fraction) in engine coolant systems and for evaluating the performance of degas reservoirs. This method, based on electrical conductivity measurements of flowing air / coolant mixture, provides measurement, acquisition and display of coolant system temperature, pressure, flow rate, instantaneous void fraction and rate of air removal by degas bottle. Embedded temperature compensation equations are used for essentially real time display of the void fraction.

INTRODUCTION

In recent years, changes in engine materials and construction have resulted in new concerns regarding air entrapped in cooling systems caused by improper fill procedures. For example, air captured in aluminum heads can cause local "hot spots" under certain operating conditions. These hot spots may cause thermal stress fracture due to the high thermal expansion characteristics of aluminum. Many engine coolant system designs now include degas reservoir (also known as quiescent bottle or Q-bottle) to afford the removal of entrapped air.

The degas bottle is a conical plastic bottle with internal baffles to increase the residence time of flowing coolant. The coolant is forced in tangentially and returned back from the bottom with a residence time of about 30 seconds. Most of the air remains in the air space above the coolant - air interface, thus facilitating continuous removal of trapped air from the coolant. These degas bottles have been is use for last few years but no comprehensive test procedure for evaluating their performance has been available. The proposed method is an attempt to develop a comprehensive test procedure for evaluating the performance and efficiency of the degas bottles in laboratory as well as in actual vehicles.

The proposed method is based on electrical conductivity measurements of the mixture of the flowing coolant - air mixture with real time temperature compensation of conductivity. The conductivity sensor is a plastic tube with four electrodes connected to a conductivity meter for signal conditioning and display. Conventionally, two electrode systems have been in use but the major disadvantage is that over a period of time, the electrodes may get coated with non-conductive contaminants. These coatings cause large changes in the apparent conductivity of the solution, making tru solution conductivity measurements inaccurate and non-reproducible. With a pair each of current and voltage electrodes as used in four electrode systems, the conductivitiy of the solution alone is measured and is unaffected by electrode fouling. A detailed description of conductivity meter used for monitoring the liquid conductivity for electrocoat paint system has been described and successfully used in electrocoat paint systems for ensuring a uniform coating by monitoring and controlling the conductivity of the paint[a]. Conductivity measurements have also been used in the two phase flow in nuclear reactors[b].

The conductivity of air is essentially zero and if it is present and well dispersed in the liquid (coolant), the conductivity of mixture should be proportionally reduced.

---

[a] Colvin, A.D. and Butler, J.W., U.S. Patent No. 4,751,466
Colvin, A.D., Graham, D.J. and Cassatta, J.C., U.S. Patent No. 5,025,220

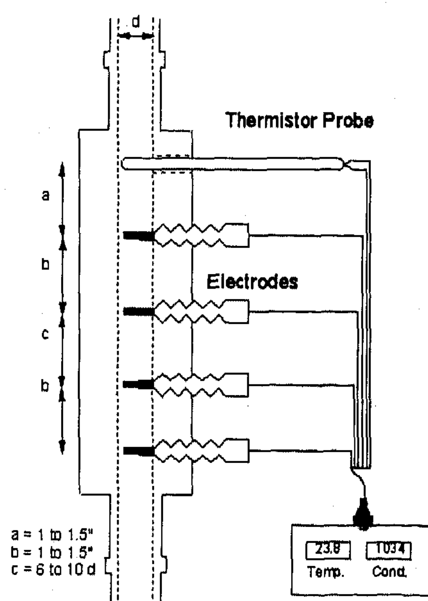
[b] Lafferty, J. F and Hammitt, F. G., "A Conductivity Probe for Measuring Local Void Fractions in Two-Phase Flow", *Nuclear Applications"*, **3**, 317 (1967)

ıe drop in the conductivity due to presence of air can be ectly correlated to the void fraction (ratio of volume of air STP to the total liquid volume). This would normally ld true for constant temperature systems. But the nductivity rises with increase in temperature and ıpirical temperature compensation equations are veloped and used to convert the conductivity value to a iform base temperature of 20 °C.

After completing the calibration process, initial ts were done by installing a degas bottle in the test rig; ɔsequently, tests on three different vehicles - a pickup ck, a heavy diesel truck and a passenger car, were npleted. The data for the two trucks are presented e.

PERIMENTAL SET UP

DESCRIPTION OF CONDUCTIVITY METER - e conductivity sensors are fabricated of nylon, starting n a single solid rod and machining as required for ling a hole, preparing the ends and screwing in the ctrodes and fittings for thermistor probe. Sensors for , 1 and 2 inch coolant hose have been used. The ısor sizes were chosen so as to coincide with the typical t(s) and outlet diameters of the degas bottle for light ks and heavy diesel trucks. A sketch of the ıductivity sensor is given in Figure 1.



1: Sketch of Conductivity Sensor

As mentioned above, the conductivity rises with ɔerature and the relationship is used to compensate effect of temperature on the conductivity value. The sure and flow rate of the circulating liquid have no t on the conductivity in the absence of any air. ever, if air is present, the volumetric expansion of d and the air pressure have to be taken into account, cially for closed systems. Air bubbles of varying size ong as the volume of each bubble is much smaller than the cylindrical volume between the two central electrodes) do not affect conductivity values.

The temperature is measured using a linearized thermistor probe (1μA/°C) installed within the body of the sensor so that the temperature can be measured at a location very close to the conductivity measurement. The signal conditioning for the temperature readout and digital display is also built in the conductivity meter. The back of the conductivity display box accepts nine pin connectors, and each pin is specifically assigned to electrodes on the conductivity sensor and the thermistor probe.

DESCRIPTION OF CALIBRATION RIG - The calibration rig has been installed in the modular, fully portable fashion and stands on a 1 m^2 chassis with wheels. All the pipelines, pumps, motor and coolant reservoir rest on the self supporting structures. In addition, shelves have been provided for all display instruments and the notebook computer used for data acquisition and analysis. The schematic of the calibration rig is shown in Figure 2, including the connections for degas bottle, which are described later.

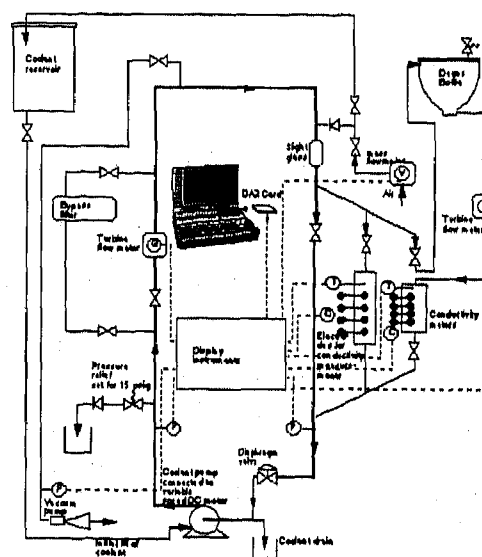


Fig 2: Schematic of Calibration Rig

A 5.5 liter truck pump for coolant circulation is rested on the base frame and connected to 0.5 hp, 3450 rpm dc motor. The pump has been completely sealed off to prevent any leakage. The speed of the motor is varied using a dc controller connected to the motor and is mounted on the instrument panel. The two discharge ends of the pump are merged together and the coolant is circulated through 1¼" copper pipe. The circulation loop forms a vertical rectangle, roughly 0.6 m x 1.8 m high. A turbine flow meter with oscillator / amplifier and digital panel meter is installed in the line to measure the circulation flow rate. Pressure transducers for 0-25 psig (0 - 175000 N/m^2) have been provided at the pump suction and discharge. Occasionally, the coolant may

need to be filtered, and thus, a bypass filter with isolation valves is provided midway between the pump discharge and the highest point in the system. The highest point is connected to a vacuum pump, through a liquid trap, so that the pipelines can be evacuated before the initial fill and subsequent topping up of the system.

The 1¼" copper pipe branches into three lines before the pump suction to accommodate installation of three conductivity sensors of sizes up to ½, 1 and 1½ inches. The diameters and angles of elbows of the network have been selected so as to distribute flow among different branches depending upon the cross sectional area of each branch. The flow of liquid through any or all branches can be controlled by partially closing the butterfly valves provided upstream of each conductivity sensor location. The three branches are merged together and connected to the pump suction through a diaphragm valve. By partially closing the diaphragm valve, it is possible to generate additional pressure inside the system, especially for the situation when no air is present so that the liquid compression boosts up the pressure. Air can be introduced at a controlled rate through a mass flow meter, capable of handling up to 1 liter / minute of air at STP conditions. A sight glass has been provided for visual observation of the air - coolant mixture.

DATA ACQUISITION / ANALYSIS - The data acquisition, on-line display and data analysis are made with IBM thinkpad 750C notebook computer using LabVIEW® for Windows version 3.0.1 software supplied by National Instruments. LabVIEW is an integrated graphical programming system for data acquisition, control, analysis and presentation. Instead of writing computer programs for data acquisition, virtual instruments (VIs) are built on the computer screen using graphical icons. VIs, with their pictorial representation, are easy to modify and maintain, and are completely self - documenting. Before building a VI, a front panel is created with all required knobs, slides, switches, graphs, strip charts and so on. The front panel serves as an interactive interface for supplying inputs to, and observing outputs from the instrumentation system. Once the VI is completed, the front panel alone is used to control the system - even while it is running - by clicking a switch, moving a slide, tweaking a knob, selecting items from menu, or entering values from the keyboard. The front panel immediately starts displaying the data, providing real time feedback from the system.

The data were acquired using a PCMCIA (Personal Computer Memory Card International Association) data acquisition card with IBM thinkpad notebook computer. The DAQCard-700, supplied by National Instruments is a type II PCMCIA I/O board, requiring very little power (120 mA at 5 volts DC) so that it can be used continuously with the battery powered PCs. The card has a 12-bit analog to digital conversions with 16 single ended analog inputs, an 8-bit TTL-compatible digital input port, an 8-bit TTL-compatible digital output port, and two user - available 16-bit counter / timer channels. All the instruments are wired using regular screw terminals to a 50 point connector block assembly, which, in turn, is connected to the data acquisition card through ribbon cable.

The combination of DAQCard 700 and LabVIEW software is capable of sampling up to 16 channels at a rate of up to several thousand data points per second. With the additional constraints of real time graphic display and use of temperature compensation equations and other correction factors, it was found that up to about 200 points can be sampled every second. For a typical configuration of about 10 channels in use at a time, each graph could be updated about 20 times per second. But in order to minimize the noise in signals, usually 10 to 30 point averages are used and thus, screen updating every one or two seconds is most practical.

CALIBRATION DATA

TEMPERATURE COMPENSATION OF CONDUCTIVITY - The most important and crucial calibration data is the effect of temperature on conductivity of the coolant (ethylene glycol - water mixture). The conductivity has been found to increase roughly 4% for every degree Celsius rise in temperature at ambient conditions, with the slope tapering down to about 1% at actual operating temperatures of 100° C or more. Because of a large temperature range to be considered (20° - 100° C), extreme care is required in temperature compensation of conductivity so as not to cause any great errors at high temperature range. Since each thermistor may have slightly varying temperature slopes, each sensor was individually calibrated using respective thermistor and conductivity meter. Most of the data were obtained by simultaneously using one 3/8 inch sensor and one 1 inch sensor installed in the two smaller branches of the calibration rig, while the largest diameter branch was used for bypass.

In a typical calibration run, the rig is evacuated using the vacuum pump to the best of its capability of about 25 mm Hg (3000 N/m^2) and filled with coolant from the reservoir. Some air is still trapped in the pipelines and is sequentially removed by using either or both of the following procedures, one or more times as necessary. Some liquid is drained from the rig, leaving extra air space at top, at which point vacuum pump is started again. Since there is lesser air to be removed (about 0.5 - 1 liter as against 7.2 liters at the beginning), an absolute pressure of less than 25 mm Hg (3000 N/m^2) is reached quickly, leaving very little residual air. The procedure may be repeated again, if necessary. Alternatively, the diaphragm valve near pump suction may be closed to compress the liquid, with all air accumulating at top, which can be discharged to atmosphere. The diaphragm valve is then opened, and more coolant allowed to flow from reservoir.

By switching on the motor controller, coolant is re-circulated at desired speed and flow rate. For pump rpm of about 3000, a discharge flow rate of up to 60 gallons / minute (0.0038 m^3/sec) can be obtained. Most of the experiments are conducted at 40 - 50 gpm (0.0025 - 0.0031 m^3/sec). These flow rates were chosen so as to be compatible with the actual water pump performance.

Three heating tapes are wrapped around different locations in the calibration rig. To get the conductivity data over a wide temperature range without concentration of data over any specific range, only one heater is started at the beginning, and the data acquisition is started. When the temperature slope tapers off, second heater, and then the third, as required, are started. The heat input through some of these heaters can also be controlled by adjusting the voltage in the variac. Most of the no void calibration data have been obtained for temperature range of ambient to 100° C. At the end of the run, heating tapes are switched off, and circulation continued for some time to avoid overheating of the pipelines.

During the entire period, pressure relief valve, set for 15 psig (10^5 N/m^2), is kept open to avoid excessive build up of pressure in the system. The liquid expands with the rise in temperature, and since there is no air space for expansion, the coolant starts dripping out through the pressure relief valve. The amount of liquid drained through the pressure relief valve should be the same as the increase in liquid volume from the onset of activation of pressure relief valve to the end of the run.

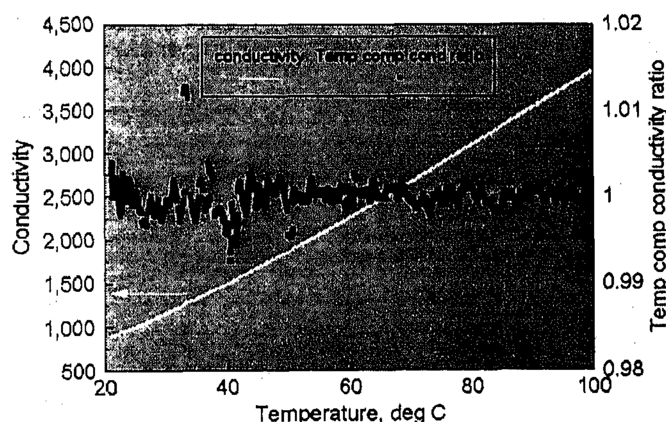


Fig 3: Conductivity - Temperature Graph for no-Void Calibration of 1 inch diameter sensor

The conductivity - temperature graph for a no void basis of a typical run is shown in Figure 3 for a 1" diameter sensor. Very similar graphs were obtained for other sensors also. The conductivity (in arbitrary units - only the change in conductivity is important rather than its absolute value), is found to depend on temperature according to quadratic equations. These equations are expected to be valid over the entire temperature range of 20 to 100° C for which data were taken at the calibration rig and can be extended up to 120° C, the highest system temperatures, expected in a vehicle cooling system. However, at lower temperatures cubic equations were found more appropriate, as described below:

<u>Sensor A (3/8 inch diameter):</u>

(T>91 °C):

$$C = 820.16 + 32.089(T-20) + 0.09781(T-20)^2 \quad (1)$$

(T≤91 °C):

$$C = 850.34 + 27.464(T-20) + 0.24308(T-20)^2 - 0.00121143(T-20)^3 \quad (2)$$

<u>Sensor C (1 inch diameter):</u>

(T>91 °C):

$$C = 838.77 + 31.618(T-20) + 0.09715(T-20)^2 \quad (3)$$

(T≤91 °C):

$$C = 864.15 + 27.824(T-20) + 0.21485(T-20)^2 - 0.00097178(T-20)^3 \quad (4)$$

For normalizing the conductivity to a base temperature, the temperature compensated conductivity ratio (TCCR) is defined according to equation (5).

$$TCCR = \frac{C_i}{\alpha + \beta(T_i - 20) + \gamma(T_i - 20)^2 + \delta(T_i - 20)^3} \quad (5)$$

In this equation, C_i and T_i refer to the instantaneous conductivity and temperature of a sensor. α, β, γ and δ are the coefficients of the conductivity - temperature relationships given in equations (1) to (4). Figure 3 also shows the temperature compensated conductivity ratio for sensor C over a temperature range of 20 - 100° C. The numerical value is reasonably constant and is found to be 1.0±0.005 over this range. The nominal noise is due to the limitations of the instruments and the data acquisition system.

CALIBRATION WITH KNOWN AMOUNTS OF AIR - To check the void fraction dependency of conductivity, several runs with known amounts of air were made. In a typical experiment, the calibration rig is completely filled with coolant as described earlier. An initial check of the complete fill (no voids) is made by circulating the coolant and checking the instantaneous temperature and conductivity for each of the two sensors at ambient temperature by running the data acquisition software and watching the temperature compensated conductivity ratio on computer screen as calculated using equation (5). If TCCR is 1.0±0.005 (allowing for experimental errors), the experiment is continued; otherwise, the fill of the calibration rig may not be completely void free, or another problem exists, and the whole process of filling needs to be repeated.

Once the system is completely filled, a known amount of air is introduced through the mass flow meter. Using a small air flow rate of 100-200 ml/minute at STP (1.7×10^{-6} - 3.3×10^{-6} m^3/sec), equivalent amount of liquid is continuously discharged from the bottom of the pump. Once the desired amount of air has been injected (as indicated by time elapsed, as well as the volume of the

liquid expelled), air line is shut off, immediately after which, the drain valve is closed. The collected volume of the liquid is weighed, and thus, the volume discharged is quantified. The system volume is known to be 7.2 liters and so the void fraction artificially introduced is determined. The pump is started and the homogeneity of the mixture is monitored through the sight glass. When the system is homogeneous (indicated by visual observation through the sight glass as well as constant conductivity value), 'START' button on the computer screen is pressed to start the data acquisition and display the results on the computer screen.

The temperature of the system is raised in the same manner as for no void basis, by sequentially starting the heating tapes one by one. With enough air in the system, the pressure relief valve should not be opened, and the pressure is allowed to rise till the upper limit of transducers (25 psig). With small amounts of air (void fraction < 0.04), the system pressure reaches 25 psig (175,000 N/m^2) before the 100° C temperature, at which point the system is shut down. For higher void fractions, data acquisition is continued till the temperature reaches 100° C. Results from several runs with varying amounts of air are shown in Figure 4.

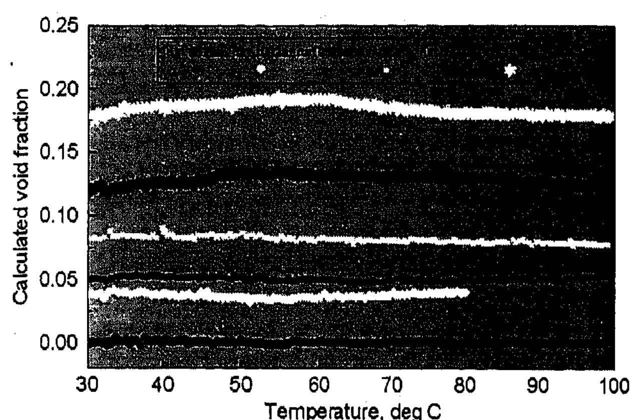


Fig 4: Calibration Data with Varying Amounts of Air

As expected, the void fraction based on these calculations remains constant (same as the initially created void fraction) throughout the experimental temperature range, subject to the experimental error. In the temperature compensation of conductivity, an error of up to ±0.5% can be expected. While accounting for liquid expansion and other unforeseen circumstances, the method is found accurate up to about ±0.75% over the entire temperature range from 20 - 100° C and initial air content of 2 - 16 % for which the experiments were conducted.

DEGAS BOTTLE TESTS IN LABORATORY

Initial tests for evaluating the performance of degas bottles have been conducted in the laboratory set up. A fixture for mounting the degas bottle was built and attached to the chassis on which other equipment rest. The height of degas bottle is adjustable, and is kept at a location so that the maximum liquid level in the degas bottle is at the same vertical level as the highest point in rest of the system. This arrangement ensures that overflow of degas bottle would be avoided at all times.

The pipelines upstream of one of the branches (3/8 inch) of the discharge end of the circulation loop were modified to include connections with the degas bottle. Isolation valves have been provided in the main line so that degas bottle can be brought into service whenever required; at other times, the main loop (without one conductivity sensor) can be operated in the usual manner. The complete schematic of the calibration rig with degas bottle attachments has been shown in Figure 2.

The laboratory tests have been conducted at ambient and near - ambient temperatures (<50° C). In a typical experiment, the degas bottle loop is isolated and the main system is filled with coolant completely, without any voids or air bubbles. The coolant is circulated and the no-void conductivity is established using the first sequence of the VI on the computer screen. As outlined in previous section, a known amount of liquid is drained while simultaneously substituting with ambient air. The heating tapes may be switched on, if desired, to get the data at elevated temperatures. The mixture in main loop is allowed to homogenize by re-circulation and the 'START' button on the computer screen is pressed to start data acquisition. The void fraction graph on computer screen, of course, indicates a constant void fraction equivalent to the artificially created void fraction in the main system.

To bring degas bottle in the service, it should be pre-filled with liquid level up to the 'maximum level', and with no air bubbles or voids in rest of the pipelines. Both the valves on the main system - inlet to the degas bottle, and the outlet from the degas bottle into main loop, are simultaneously opened. The liquid level in the inlet line momentarily drops, after which, the coolant - air mixture enters the degas bottle. The flow pattern of the coolant - air mixture can be clearly seen inside the degas bottle, especially if the mixture is very milky with air content exceeding 10% of the total system volume.

A very steep drop in the void fraction is observed for several seconds, as displayed in the graph on computer screen. This is followed by tapering of void fraction, ending with asymptotic behavior with nearly zero void fraction. Typical operating conditions consist of coolant flow rate of 40 gpm (0.0025 m^3/sec) with 2-5% flow through the degas bottle. The void fraction reaches zero in about 20 - 30 minutes. A typical graph, with an initial void fraction of about 6% is shown in Figure 5 and another with 14% initial void fraction is given in Figure 6. The horizontal portion of the graph shows the void fraction before connecting degas bottle to the system, showing a constant void fraction. As soon as the degas

bottle is connected, it starts removing the air from coolant. The slope of void fraction vs time is exceptionally high during the first 30-60 seconds, indicating a very high rate of air removal. The tapering down zone, followed by asymptotic behavior follows a pattern similar to an exponential decay curve. The void fraction reaches 0.5% slightly faster in case of lower air content but the rate of air removal is not remarkably different for the two cases.

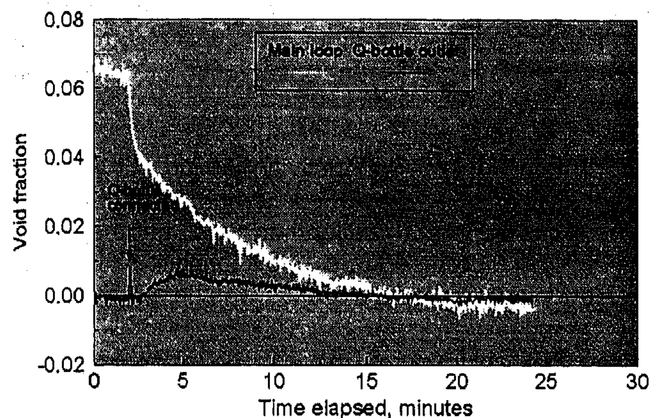


Fig 5: Air Removal from Degas Bottle for 6% Initial Void Fraction

The laboratory tests were conducted with the degas bottles of pick up truck, heavy diesel truck and a passenger car. The shape, size and geometry of baffles of these bottles are different but the working principle is same in all cases, with a residence time of 20 - 30 seconds. In each case, air in the system drops down to below 1% within 15-20 minutes.

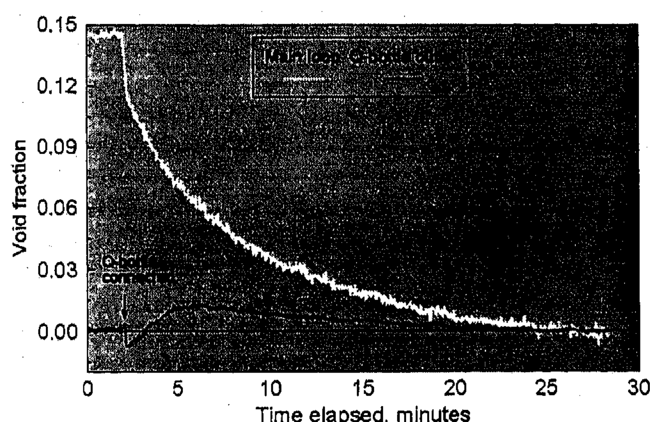


Fig 6: Air Removal from Degas Bottle for 14% Initial Void Fraction

After the coolant was left in the experimental rig for several weeks, the degas bottle test was repeated a few more times at varying flow rates. It was found that when the degas bottle flow rate was well below 1 gpm ($6.3x10^{-5}$ m^3/sec), the de-aeration took place in the normal way but when flow through the inlet nozzle was increased to greater than 1 gpm ($6.3x10^{-5}$ m^3/sec), degas bottle stopped functioning and in fact, started putting air back into the system. These characteristics are shown in Figure 7 which shows four clear zones of flow rates through the degas bottle. Initially, the flow was maintained at 1.1 gpm ($7x10^{-5}$ m^3/sec) for which the void fraction could not be brought below 0.06. When the flow was dropped to 0.55 gpm ($3.5x10^{-5}$ m^3/sec), the de-aeration followed the regular exponential decay trend and the complete removal of air was observed in 25 minutes. The flow rate was increased back to 1.1 gpm ($7x10^{-5}$ m^3/sec) at which time, the degas bottle started putting air back into the main loop till the void fraction was about 0.07. This step took roughly 20 minutes. When the flow rate was cut back to half again, degas bottle stared de-aerating again. During the entire run, the fractional flow through the degas bottle was maintained between 3.5 - 4% of the total flow.

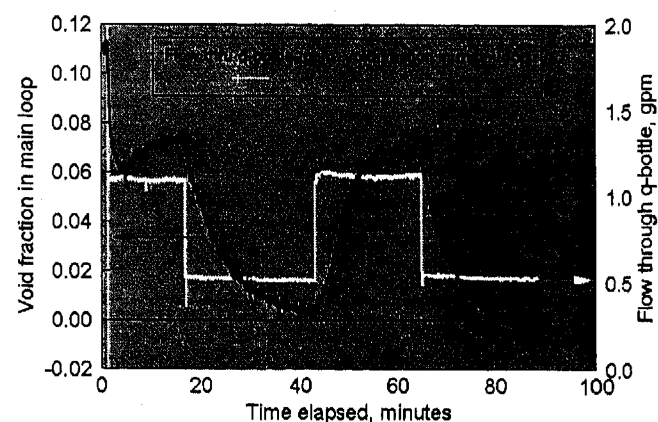


Fig 7: Degas Bottle Re-aeration with High Flow and Aged Coolant

It seems that the behavior of degas bottle is a complex function of maximum permissible flow as well as the aging of the coolant. When the coolant was changed, the results could not be reproduced. The long exposure of the coolant to the atmospheric oxygen may have been responsible for loosing the characteristics of the anti-foaming agents used as additives in coolant.

VEHICLE TESTS

Based on the success of the technique for monitoring void fraction in laboratory degas - bottle set up, additional tests were conducted on vehicles. In order to get the data based on wide operating conditions, tests were conducted on three different vehicles - 5.4 liter F250 pickup truck, 12 liter L9000 diesel truck and a 3 liter Taurus car. The data for the diesel truck were obtained in a chassis dynamometer set up whereas other two vehicle tests were conducted at fast idle conditions without putting any load. Data for both the trucks are presented here.

The de-aeration data are best obtained by monitoring the conductivity in a continuous flow loop and therefore, heater core loop was considered for all the

tests. Some data were also obtained at the degas bottle outlet and radiator inlet. A typical cooling system schematic is shown in Figure 8, showing the sensors at the inlet of heater core, where the coolant is hottest. The coolant flow through the conductivity sensor, where pressure, temperature and conductivity are monitored, and then the direction is reversed while installing the turbine flow meter. The outlet of flow meter is connected back to the heater hose. The additional instruments and hoses increase the cooling system capacity by about one liter, which roughly equals the additional amount of void artificially created at the time of each fill up.

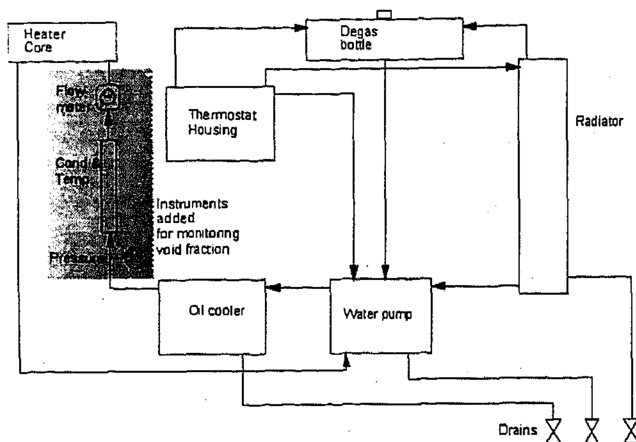


Fig 8: Cooling System Schematic

Following strategy for a cooling system de-aeration measurements was adopted for the vehicle testing:

(1) Install the instruments at the heater core

(2) Establish the exact capacity of the cooling system with incorporation of additional instruments and hoses. The cooling system is completely drained and measured amount of coolant added. Vehicle is started and a few minutes after thermostat opens up, shut down and then allowed to cool. The system is refilled with coolant and the cycle is repeated 2-3 times as necessary.

(3) Run the data acquisition system for completely filled cooling system for normalization.

(4) Run the de-aeration measurement tests at idle, typical operating and an intermediate rpm with a known volume of coolant which is deliberately under-filled.

(5) Repeat the tests at the above rpm's for different amounts of coolant so as to have different initial void fractions.

(6) As additional tests, run the vehicle at idle rpm and gradually raise the speed to speeds higher than typical operating speeds and observe the rate of air removal as a function of vehicle speed.

L9000 HEAVY DIESEL TRUCK - The de-aeration tests for L9000 diesel truck were made at APX International, 1120 John A Papalas Drive in Lincoln Park. The truck is equipped with a 12 liter engine from Detroit Diesel. All the measurements were made in the chassis - dynamometer set up wherein the vehicle was operated under closely monitored load to simulate the actual driving conditions. The set of instruments was installed at the heater core inlet. A sight glass was also provided at the radiator inlet for visual observation of the vehicle de-aeration.

To account for conductivity variation of water at different locations, samples of coolant and water from the site were carried over to calibration rig to get the actual conductivity - temperature relationship. Following conductivity - temperature equation was generated:

$$C = 835.92 + 31.714(T-20) + 0.20765(T-20)^2 - 0.00097753(T-20)^3 \quad (8)$$

The total capacity of the cooling system (including the degas bottle and additional pipelines / instruments) was found to be 36.1 liters and the runs on different days were conducted by under-filling the system by 2 to 5 liters, making sure that some liquid always existed in the degas bottle to avoid the possibility of re-entraining air back into the system. All the tests were conducted in the dynamometer room by running the truck under a known load and speed. A heavy duty fan installed in front of the truck was used for cooling the radiator and engine between two successive runs on the same day. The typical operating speed of this truck is at 1800 rpm, based on highway driving at 55 miles per hour, while the recommended maximum rpm is 2100.

The vehicle needed to be drained completely and refilled to desired capacity before every run. This was accomplished by draining coolant from the water pump, oil cooler and bottom of the radiator under gravity. Extended hoses and valves / pet cocks were provided for each line to facilitate easy drainage. When no more liquid flowed, compressed air was surged through a specially fabricated degas bottle cap which accepts ¼ inch compressed air line. Usually, compressed air at 10-20 psig (70000 - 140,000 N/m^2) was found to adequately drain the complete coolant in about 30 minutes. The drained coolant was always collected and measured to make sure that all coolant put in for the run was collected.

Figure 9 shows the temperature and de-aeration data at 1300 rpm. The truck was started by putting a load of 50 hp on each axle, simulating a speed of about 40 MPH which was maintained constant throughout the run. The temperature started rising within one minute of starting the truck and the thermostat started opening

within 7 minutes of the start up. The graph shows the temperature oscillations for about 2.5 minutes, going through 4 cycles. During this period, air removal rate is very high as observed on the computer screen as well as the sight glass at radiator inlet. Once the thermostat is fully open, the temperature continues to rise and is controlled by the radiator fan which kicks in at about 93 °C. The temperature oscillates between 83 and 93 °C, leaving the thermostat partially closed for a few seconds during each fan cycle, lasting about a minute each. Some air trapped in the cooling system shows up in the conductivity sensor and therefore, the graph shows continuous oscillations in the calculated void fractions. Rate of air removal tapers down during this period, showing the complete removal in over 30 minutes after the first opening of thermostat. Even if the vehicle is run for a long time, the characteristics of the void fraction curve do not change.

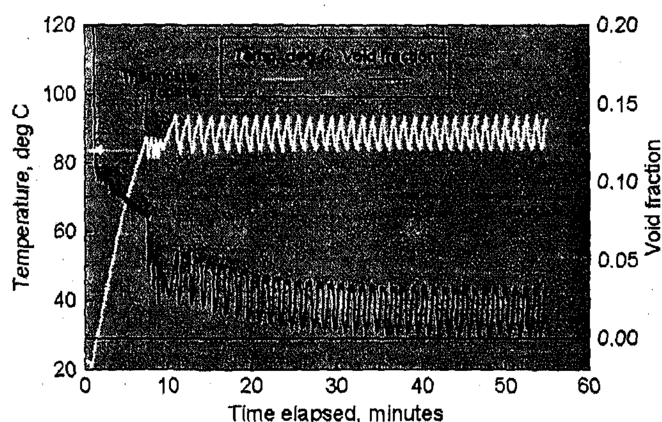


Fig 9: L9000 Heavy Diesel Truck De-aeration Data at 1300 rpm

The pressure and the flow rate at the heater core are shown in Figure 10. The pressure steeply rises till thermostat opens up, after which it cycles along with the thermostat opening and closing (decreases when thermostat closes after shutting off of radiator fan), with gradual overall drop in pressure. The flow rate remains fairly constant at 3 gallons per minute (0.00019 m^3/sec) with only minimal variations with the cycling of radiator fan.

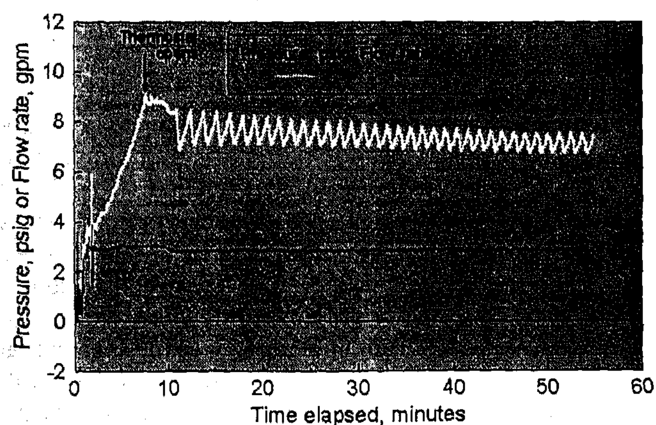


Fig 10: L9000 Heavy Diesel Truck Pressure and Flow Rate at 1300 rpm

It was noticed that the truck had to be operated with at least 1300 rpm (simulating a speed of 40 MPH) for reasonably good de-aeration. Figure 11 shows the temperature and void fraction data at 1000 rpm (based on 30 hp load on each axle, simulating a speed of 25 MPH). Although 10% void fraction had initially been created, it does not show in the graph because all air pockets are not removed by the flowing coolant. Thermostat takes longer to open (about 12 minutes) and the radiator fan cycling is much slower. Although the graph shows only small void fraction, all air had not been removed by the degas bottle, as evident from the fast approach of void fraction towards zero value.

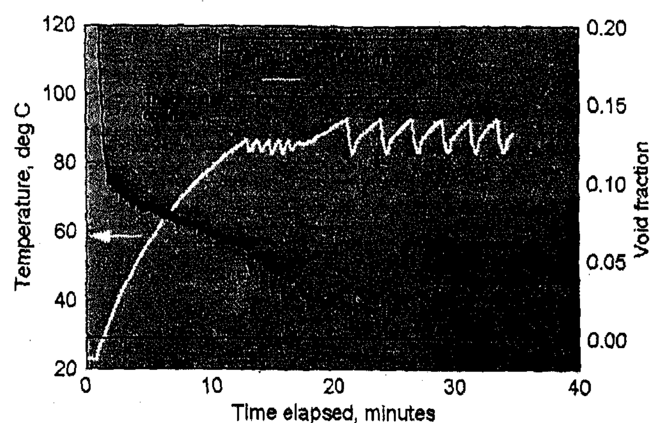


Fig 11: L9000 Heavy Diesel Truck De-aeration Data at 1000 rpm

In order to compare the de-aeration data at heater core loop and the radiator inlet, another conductivity sensor (2 inch diameter) was built and installed at the radiator inlet. Using the sight glass, it had been observed that the air is not uniformly distributed along the entire cross section. There is always more air at the top than the bottom part of the sight glass. Therefore, two sets of electrodes were used - one set was inserted from top and the other from bottom. The output of both was combined to get the data representative of the entire cross section. The comparison of void fraction observed at both the locations is shown in Figure 12 for one run at 1800 rpm. The data from the radiator sensor is meaningless before the opening of thermostat as well as immediately after opening of thermostat because of very large amount of air flow. The instruments could not detect a very rapid change in conductivity. However, the data is presented for the period after the thermostat had been open for a few minutes. Both the curves follow the void fraction cycling with the radiator fan, and are in phase. The de-aeration at the radiator inlet was slightly slower than the data at heater core. This confirms the fact that a small amount of air always resides in the cooling system and shows up only during the closing period of thermostat.

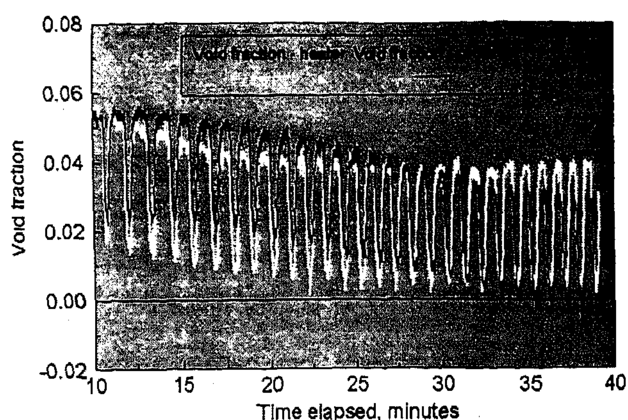


Fig 12: L9000 Heavy Diesel Truck De-aeration Data at 1800 rpm - Comparison of Void Fraction at Heater and Radiator

F250 PICK UP TRUCK - The tests on PN102 vehicle, an F250 pickup truck, .were conducted in the garage of Carron & Company at 26700 Princeton Avenue in Inkster. As explained above, instruments package was installed in the heater core loop. An additional set of instruments package was also tried at the degas bottle outlet for continuously monitoring the air removal efficiency of the degas bottle. However, the conductivity and flow sensors did not work because the flow was not continuous and suitable signals could not be generated with any reasonable confidence. The signal conditioning instruments, data acquisition system and the notebook computer rest on the passenger seat on a sheet metal tray. The conductivity of water and coolant mixture at this site was different than the values obtained at calibration rig, and therefore, a correction factor of 1.36 was applied for all on line calculations.

The total capacity of the cooling system was found to be 16.5 liters and the runs on different days were conducted by under-filling the system by 1 to 3 liters, making sure that some liquid always existed in the degas bottle to avoid the possibility of re-entraining air back into the system. All the tests were conducted on high idle speed in the garage without putting any load on the vehicle. A floor fan put directly in front of the radiator was used for controlling the temperature and for simulating the actual driving conditions.

The vehicle had to be drained completely and refilled to desired capacity before every run. This was accomplished by draining the engine drain block and bottom of the radiator under gravity. To facilitate easy drainage from engine, a pet cock was installed at the engine drain block. When no more liquid flowed, compressed air was surged through the degas bottle inlet (while clamping the line from top of the radiator which was connected to the degas bottle). The drained coolant was always collected and measured to make sure that all coolant put in for the run was collected and accounted.

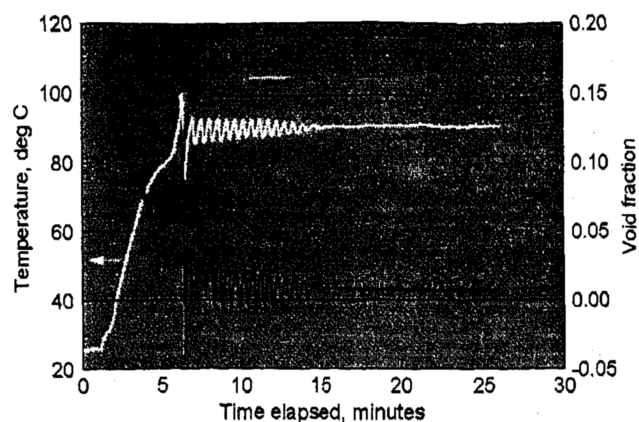


Fig 13: F250 Pickup Truck De-aeration Data at 2500 rpm

The data from a typical run are shown in Figure 13 wherein the temperature and calculated void fraction are plotted against time from the beginning of the data acquisition. For first minute, only the base conditions were recorded, at which time, the engine was started and quickly raised to desired rpm. The temperature in the heater core gradually rises and the thermostat opened up after about 5.5 minutes. Right at this moment, the degas bottle enters the coolant circulation loop and the void fraction starts dropping. For several minutes, the thermostat goes through the cycle of opening and closing (about 8.5 minutes for this graph) with an oscillation period of roughly 35 seconds between two consecutive cycles. The temperature subsequently becomes constant with thermostat in fully open position.

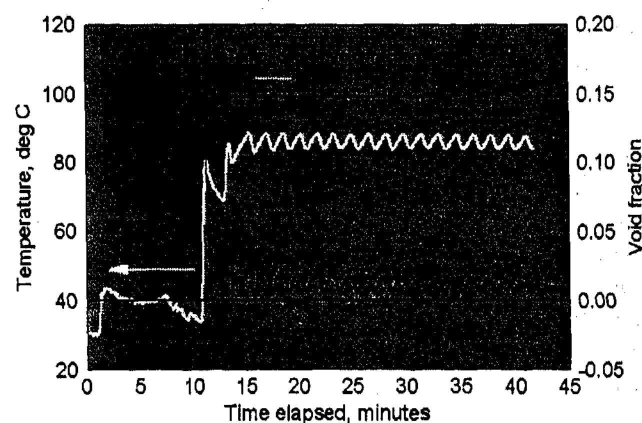


Fig 14: F250 Pickup Truck De-aeration Data at 700 rpm

In many instances, the temperature rise in heater core was not steady, rising marginally and jumping to about 85 °C as soon as thermostat opened. This could be attributed directly to the level of coolant in the degas bottle. When the cooling system was highly under-filled and / or the vehicle was operated at low rpm, the heater core temperature rise was never steady. Figure 14 shows the data at idle rpm, indicating that the thermostat opened up after 10 minutes of vehicle operation with temperature suddenly jumping to the operating temperature when the thermostat opened. A large amount of air was entrained in the system but could not

be picked up by the flowing coolant. Although the graph shows near zero void fraction, air still remained in the system. This is apparent based on high void fraction shown when the thermostat opened up. Further, even after 30 minutes of vehicle operation, the thermostat oscillations continued without any dampening effects in temperature. This shows that air entrained in the cooling system can not be removed at low rpm.

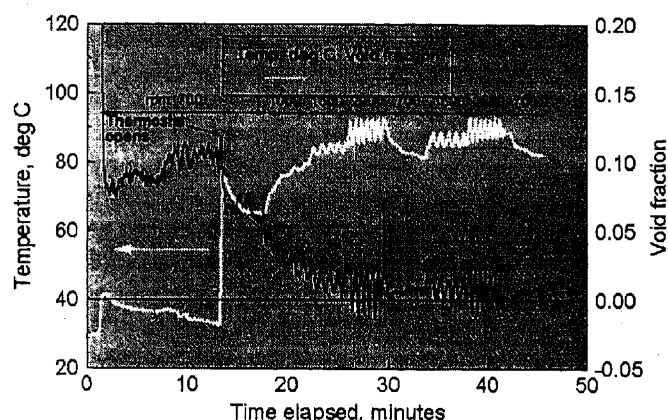


Fig 15: F250 Pickup Truck De-aeration Data with Varying Flow Rates Between 700 and 2900 rpm

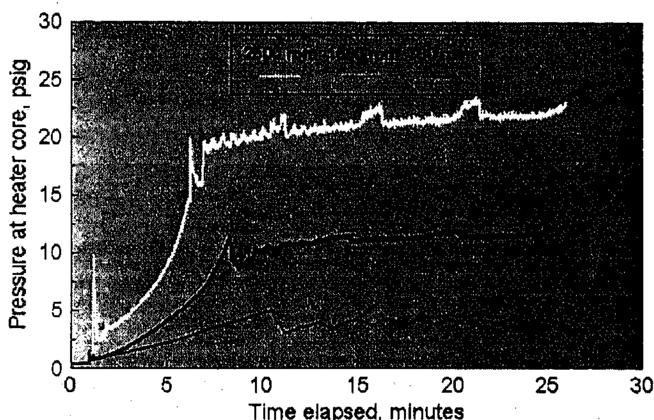


Fig 16: F250 Pickup Truck Pressure at Different rpm

The effect of engine rpm on the rate of de-aeration is illustrated in Figure 15. The vehicle was started at idle rpm and a few minutes after the thermostat opened, the speed was gradually increased to 2900 rpm. No de-aeration was observed at idle conditions while only a partial de-aeration took place at 1000 and 1600 rpm. Complete de-aeration was observed when the speed was further increased. Subsequent drop and increases in rpm had no effect on the level of de-aeration, though the oscillations of the thermostat followed the predicted behavior. The pressure in the heater core transducer was found to exceed 16 psig (110,000 N/m^2), the relief pressure set for the degas bottle cap. It is expected that the additional pressure in heater core should be about 3 psig (20,000 N/m^2) as compared to the pressure in the air space in degas bottle. Figure 16 shows the pressure in the heater core obtained on three different runs at speeds of 700, 1600 and 2500, with the same level of coolant. At lower rpm, pressure is below 12 psig (84,000 N/m^2), but at 2500 rpm, degas bottle cap is popping on periodically due to high pressure after opening of the thermostat.

CONCLUSIONS

- Electrical conductivity measurements provide a functional method for establishing the fractional amount of air in the flowing coolant, with an accuracy of ±0.5% absolute.
- The temperature - conductivity relationship of coolant - water mixture is quadratic in nature over the range of 20-120 °C, the typical operating range.
- The conductivity of different coolants and waters are not only different at ambient temperature, but may also have different temperature slopes. Therefore, the equation should be generated for each quality of coolant - water mixture, prior to actual vehicle tests.
- Laboratory tests have shown that degas bottle can remove up to 15% air within 10-20 minutes for typical operating conditions.
- The performance of degas bottle significantly deteriorates with high flow and aged coolant. If both these factors combine, the degas bottle starts aerating instead of removing the air.
- Vehicle tests have shown that at idle or near idle conditions, flowing coolant is not able to pick up the air; conductivity sensors may show de-aerated coolant while air pockets remain in the system. Tests showed that a good level of de-aeration could be achieved when the operating rpm was at least twice the idle rpm (1500 - 2000 rpm).
- The time required for complete de-aeration at intermediate to high rpm's was at least 20 minutes for diesel truck whereas the de-aeration in pick up truck was faster under comparable operating conditions. The slower rate of de-aeration in heavy truck may be due to much larger coolant volume but the circulation flow rate is not proportionately high.

970940

Comparison of Aqueous Mixtures of Propylene-Glycol and Ethylene-Glycol Under Simulated Engine Cooling Conditions

C. Branchi, S. Bhowmick, and E. V. McAssey, Jr.
Villanova Univ.

M. Gollin and G. Cozzone
ARCO Chemical Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Test results have been obtained on the thermal performance of aqueous mixtures of propylene-glycol and ethylene-glycol for conditions simulating automotive engine operation. For the present study, thermal performance is defined as the test section surface temperature for a given set of operating conditions. In addition to testing over a wide range of surface heat fluxes up to 2.3 MW/m^2, data were also obtained at different mixture concentrations. For all cases considered, both coolant mixtures provided similar results.

INTRODUCTION

An extensive test program has been accomplished to evaluate the thermal performance of propylene-glycol and ethylene-glycol aqueous mixtures under conditions similar to those existing in modern internal combustion engines.

Because of environmental concerns, there has been an increased interest in replacing ethylene-glycol as the primary engine cooling fluid. Although ethylene-glycol/water mixtures have performed satisfactorily in internal combustion engines for a number of years, leaks from cooling systems annually cause a large number of human and domestic pet poisonings. Propylene-glycol appears to be an ideal replacement fluid. It is less toxic and possesses very similar heat transfer characteristics when compared to ethylene-glycol. However, differences do exist. Ethylene-glycol/water mixtures have a lower viscosity and higher thermal conductivity than propylene-glycol/water mixtures at the same concentration. At the normal cooling system operating pressure of 200 kPa, the saturation temperature of a 50/50 by volume mixture of propylene-glycol/water is approximately 3 °C lower than the equivalent mixture of ethylene-glycol/water.

In a modern automobile engine, the cooling system operates over the entire heat transfer spectrum from single phase convection to saturated flow boiling. The objective of the present study was to obtain experimental data on the heat transfer characteristics of both coolant mixtures over a wide range of flow-rates and surface heat fluxes. In addition, the effects of variations in concentrations from the normal 50/50 by volume value were also examined.

BACKGROUND

Although ethylene-glycol/water mixtures have been used as engine coolant for over fifty years, there is very little data available in the open literature.

In an automobile cooling system, the working fluid is generally a mixture of water and either ethylene-glycol or propylene-glycol. The normal mixture concentration is 50/50 by volume. Finlay et al. (1987) presented experimental results for an ethylene-glycol/water mixture covering an operating range appropriate to automotive engine conditions. Most of the data were obtained under constant pressure conditions using a copper test section. Some data were also obtained for other test section materials including cast iron and aluminum, and for constant flow-rate operation. These results showed reasonable agreement with analytical predictions based upon correlations at low surface heat fluxes. However, at higher fluxes under sub-cooled boiling conditions, the analytical model tended to under-predict the surface temperature. Gehres (1963) examined the performance of ethylene-glycol/water mixtures at different locations in the engine under various operating conditions Results showed that,

for severe driving conditions, nucleate boiling takes place on approximately 60 percent of the cylinder heat transfer surface area. High heat flux regions can experience surface heat fluxes exceeding 0.3 MW/m^2. In general, experimental data using glycol mixtures at typical automotive conditions of high inlet temperature (85 °C) and system pressures of two atmospheres are limited.

McAssey, Stinson, and Gollin (1995) presented test results comparing propylene-glycol/water and ethylene-glycol/water mixtures for a range of conditions similar to those existing in normal engine operation. Test data were obtained for test section inlet velocities up to 0.5 m/s and surface heat fluxes up to approximately 0.8 MW/m^2. All the test data were obtained for a 50/50 by volume mixture at a system pressure of 205 kPa and a test section inlet temperature of 85 °C. For this range of test conditions, the resulting data spanned the spectrum from single phase convection to saturated boiling. These investigators concluded that the overall performance for both coolant mixtures was very similar. At high velocities (approximately 0.5 m/s) under predominantly single phase conditions, the ethylene-glycol/water mixture yielded slightly lower surface temperatures. Conversely, at low velocity (approximately 0.1 m/s) where sub-cooled boiling occurred, the propylene-glycol/water mixture provided lower operating temperatures on the heated surface. This paper also presented comparisons between analytical predictions and experimental results. In general, the analytical results under-predicted the surface temperature when the difference between the surface temperature and fluid temperature exceeded 60 °C.

Bhowmick, Branchi, McAssey, and Gollin (1996) presented additional data on both fluid mixtures for a wider range of conditions. In this paper, the inlet velocity varied from approximately 0.4 m/s to 2.5 m/s and the surface heat fluxes reached a maximum of 1.8 MW/m^2. The concentrations of the mixtures were limited to 50/50 by volume, and the test conditions were the same as the previous paper. In addition, comparisons were presented between experimental results and analytical predictions. The experimental results showed that both mixtures had similar thermal performances. In general, the analytical results under-predicted the experimental results by a significant margin.

DESCRIPTION OF TEST FACILITY

FLOW LOOP - The flow loop consisted of a test section, pump, accumulator tank, rejection heat exchanger, and required piping. Figure 1 presents a schematic layout of the test loop. For this program, the loop was operated in the controlled flow mode. In this mode of operation, the test section is provided a constant volumetric flow-rate

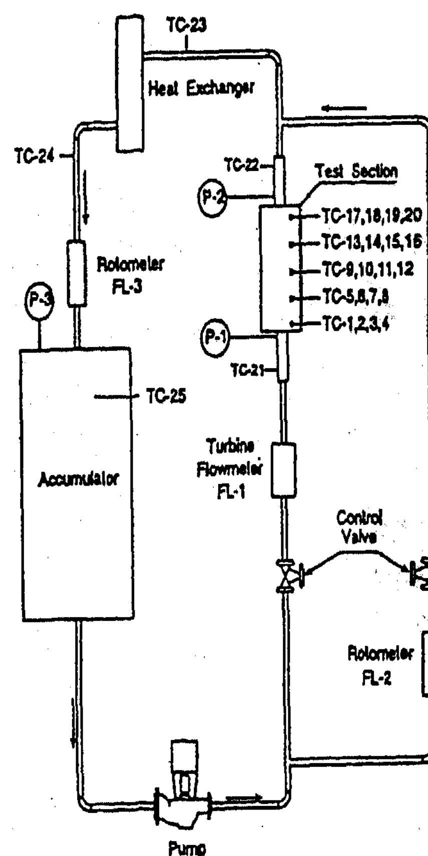


Figure 1 Schematic of Test Loop

under all operating conditions. Since most of the pressure drop occurs across the control valve, changes in the test section pressure drop due to heating have a small effect on the flow-rate to the test section. The bypass line around the test section was installed to allow operation on the constant head portion of the pump performance curve and to reduce the vapor content of the fluid entering the heat exchanger. The primary flow-rate instrument was a turbine flowmeter (FL-1) located upstream of the test section. Test section inlet pressure and temperature were measured by a pressure gauge (P-1) and a thermocouple (RTD-1) respectively. A similar arrangement (P-2 and RTD-2) located downstream of the test section was used to measure exit conditions.

The accumulator tank was used to maintain the system pressure. Pressure control was accomplished by means of adjusting the liquid level or the argon pressure. An argon blanket was maintained in the top of the accumulator tank and the pressure in this region was adjusted using an external argon supply. The tank also contained a heater which controlled the test section inlet temperature. The accumulator tank also produced a constant head source for the loop pump, which was a centrifugal pump capable of producing 0.3 m^3/min. at 400 kPa. A control valve located upstream of the test

section was used to provide the required flow to the test section. After leaving the test section, the bypass and test section flows combined before entering the heat exchanger.

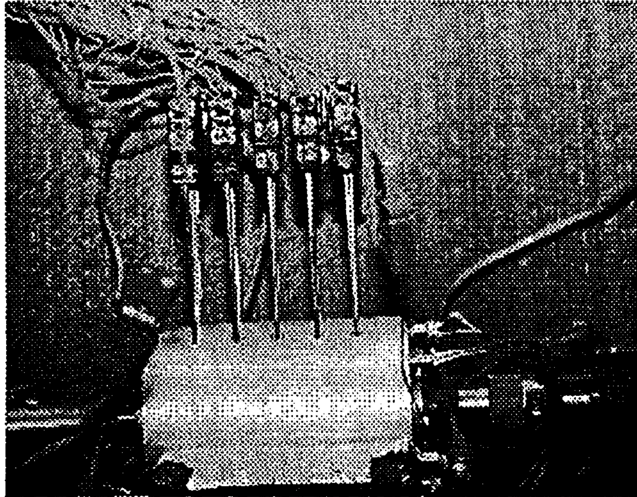


Figure 2 Test Section

TEST SECTION - The test section was constructed from a 114 mm diameter aluminum cylinder with a length of 165 mm. The horizontal flow channel consisted of a 9.53 mm diameter hole drilled along the center-line of the cylinder (see Fig. 2). The test section was heated by ten cartridge heaters located on a 97.9 mm diameter circle. Each heater was capable of producing 1,000 watts at 120 volts. By operating at slightly higher voltages, it was possible to dissipate approximately 11 kW in the test section.

The test section was connected to the flow loop by means of inlet and exit calming sections, which have the same inside diameter as the test section. The upstream calming length and the test section each had an L/D ratio of 17:1. This ratio was considered to be sufficient to provide fully developed flow to the heated length.

In order to minimize heat loss from the system, the entire flow loop and test section were insulated. The insulation on the test section consisted of over 200 mm of fiberglass insulation. Calculations of the adiabatic fluid exit temperature agree with the measured fluid exit temperature to within 2-3 °C.

The test section temperatures were measured by 20 E-type thermocouples mounted at five axial stations along the cylinder (see Fig. 2). Using two thermocouples on the same radial line, the surface temperature can be calculated as shown in Bhowmick et al. (1996). All thermocouple measurements were made in the same horizontal plane. This arrangement provided two essentially equal measurements at each axial location.

INSTRUMENTATION - The instrumentation for this program was chosen to determine various test parameters and test section temperatures. Test section flow-rates were measured with turbine flowmeters and rotometers. The rotometers were calibrated using a weigh tank with the mixtures at the required operating temperature of 85 °C. The turbine flowmeters were then calibrated against the rotometers. Initially the thermocouple calibration provided by the vendor was checked by calibrating each thermocouple in an environmental oven up to approximately 125 °C. Later in the test program, a high temperature thermocouple calibrator was used to calibrate each sensor over the complete operating range. Table 1 presents an estimate of the overall accuracy of the instruments used in these tests.

Table 1 Instrument Accuracy

Measurement	Type of Instrument	Accuracy
Temperature	E-type thermocouple	+/- 1.2 °C
Pressure	Pressure gauge	+/- 1% F.S.
Pressure Drop	Differential pressure transducer	+/- 0.25% F.S.
Test Section Flow-rate	Turbine flowmeter	+/- 0.50% of reading
Loop Flow-rate	Rotometer	+/- 0.75% F.S.
Test Section Fluid Temperature	E-type thermocouples	+/- 1.2 °C

RESULTS AND DISCUSSION

Test results were obtained over a range of test section inlet velocities from 0.44 m/s to 2.6 m/s. In addition, the surface heat flux varied from zero to 2.3 MW/m^2. For each test, the test section exit pressure was maintained at 205 kPa while the inlet temperature was set at 85 °C. Exit pressure was selected as the controlled parameter because it provided a more stable operating point. These conditions correspond to normal design operation of an engine.

The following procedure was used for each test:

1. The test section exit pressure, inlet temperature, and flow-rate were set.
2. The power to the test section heaters was increased in steps from zero to the maximum conditions. Generally, eight steps were used for each flow-rate.
3. For each power setting, the test conditions were maintained until steady state conditions had been achieved. Steady state was usually reached after 15-20 minutes of operation.

This procedure was followed for three coolant mixture concentrations, 50/50, 30/70, and 70/30 by volume for each type of fluid. Approximately, 400 tests were conducted in this program.

For this range of conditions, the full spectrum of heat transfer from single phase forced convection to saturated boiling was produced. Figure 3 presents the measured fluid exit temperatures at three inlet velocities for a 50/50 concentration of propylene-glycol/water and ethylene-glycol/water. At the lowest inlet velocity (0.445 m/s), both coolant mixtures' fluid temperatures reached saturation at approximately 1.2 MW/m^2 surface heat flux. At 50/50 concentration and these test conditions, propylene-glycol/water and ethylene-glycol/water have saturation temperatures of 127 °C and 130 °C respectively. At the maximum flux condition, adiabatic energy balance calculations indicated that approximately one half the channel experienced saturated boiling. Conversely, at the highest velocity, the exit fluid remained subcooled at approximately 100 °C. Local subcooled boiling will occur when the local surface temperature exceeds the onset of nucleate boiling (ONB) temperature. For the present test conditions, T_{ONB} is approximately 145 °C.

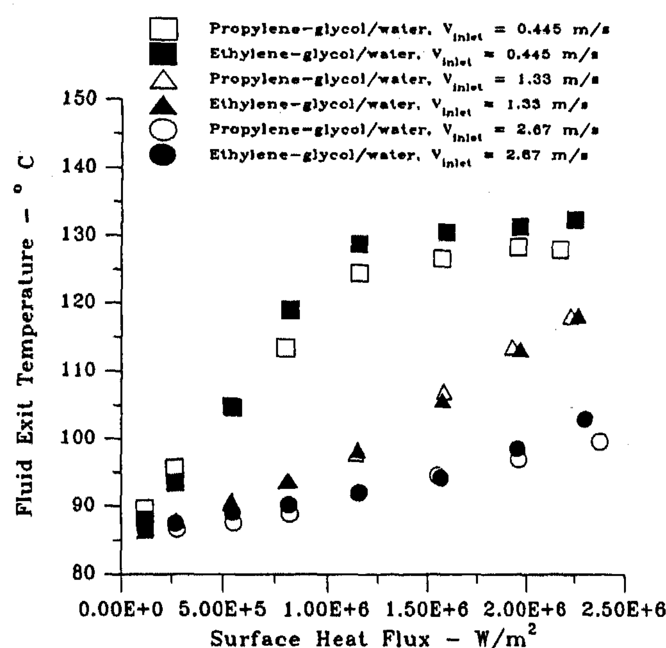


Figure 3 Comparison of fluid exit temperatures for 50/50 by volume propylene-glycol/water and ethylene-glycol/water mixtures. P_{system} = 205 kPa, T_{inlet} = 85° C

Figures 4, 5, and 6 present surface temperature results at the test section midsection (z/L = 0.5) as a function of surface heat flux for inlet velocities of 0.445 m/s, 1.33 m/s, and 2.67 m/s respectively. In these figures, data are presented for both propylene-glycol/water and ethylene-glycol/water. The data are drawn from two test sets performed approximately six months apart. The agreement between the data sets is quite good. The later data shows better agreement between the temperatures measured on each side

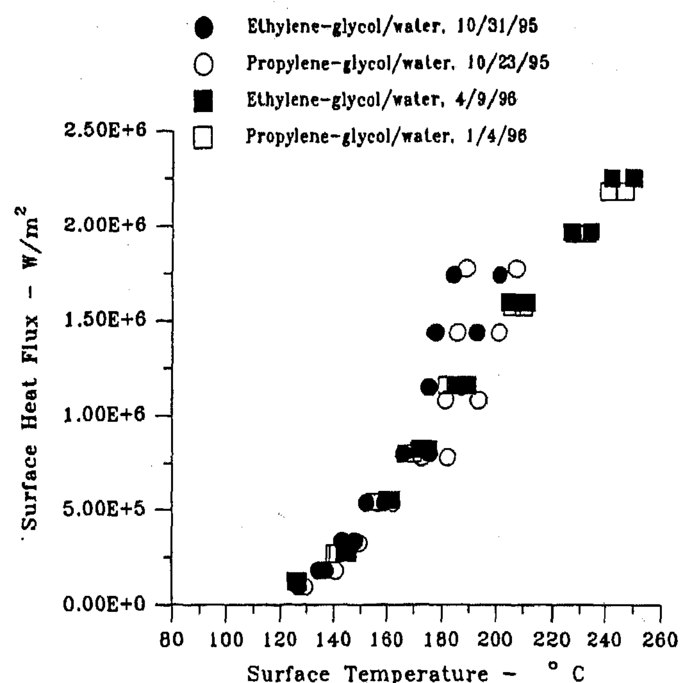


Figure 4 Comparison of propylene-glycol/water and ethylene-glycol/water at V_{inlet} = 0.445 m/s
z/L = 0.5, P_{system} = 205 kPa, T_{inlet} = 85° C

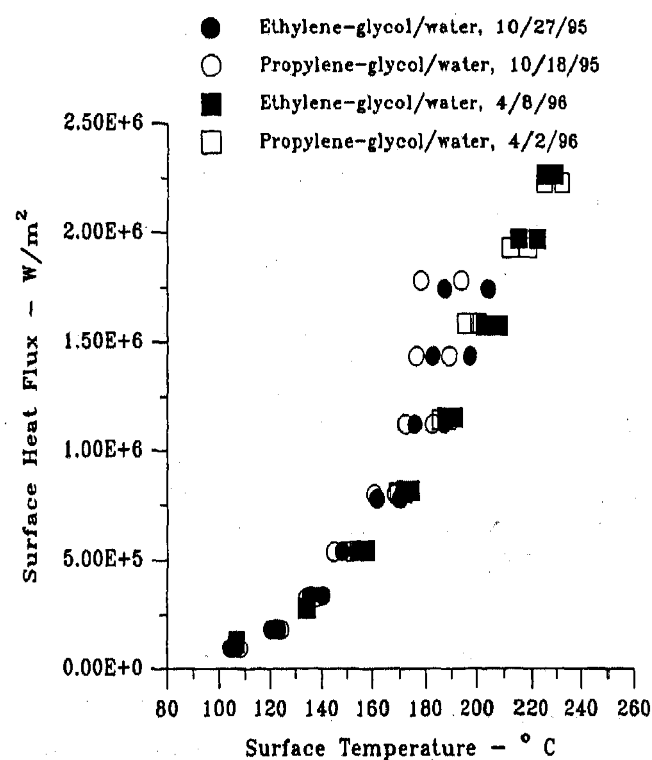


Figure 5 Comparison of propylene-glycol/water and ethylene-glycol/water at V_{inlet} = 1.33 m/s,
z/L = 0.5, P_{system} = 205 kPa, T_{inlet} = 85° C

of the flow channel. For the thermocouples used in the earlier data, the vendor-supplied calibration curves were

used. However, for the later data, each thermocouple was calibrated and matched. Both coolant fluids yielded very similar results over the range of heat fluxes and inlet velocities.

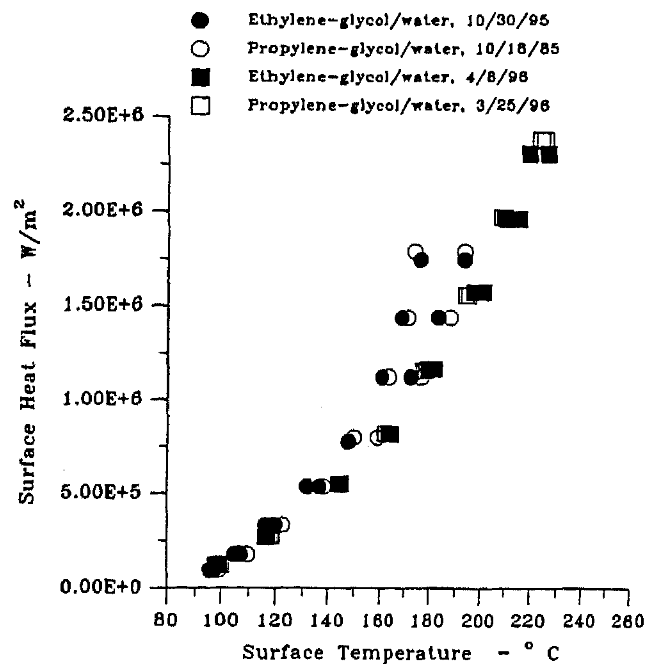


Figure 6 Comparison of propylene-glycol/water and ethylene-glycol/water at V_{inlet} = 2.67 m/s

z/L = 0.5, P_{system} = 205 kPa, T_{inlet} = 85° C

It is significant to note that over the entire range of test conditions, there was no evidence of heat transfer crisis. The data did not show the presence of critical heat flux, film boiling, or dryout. When local saturated boiling conditions were established, the local surface temperature increased more rapidly with heat flux. In Fig. 4, there is a change in slope above a surface heat flux of 1.5 MW/m². At this point in both fluids, saturated boiling is taking place and the heat transfer coefficient is slightly reduced. Conversely, in Fig. 6 for the highest inlet velocity, the coolant mixture local condition is well below saturation; however, the wall temperature is above the onset of nucleate boiling,and local subcooled boiling is present.

Table 2 Mixture Saturation Temperature

Concentration	Saturation temperature for Propylene-glycol/water	Saturation temperature for Ethylene-glycol/water
30/70 by volume	123.8 °C	124.4 °C
50/50 by volume	127.5 °C	130.3 °C
70/30 by volume	134.9 °C	138.9 °C

Figures 7, 8, and 9 present results showing the effect of coolant concentrations. In addition to the normal 50/50 by volume, tests were conducted with

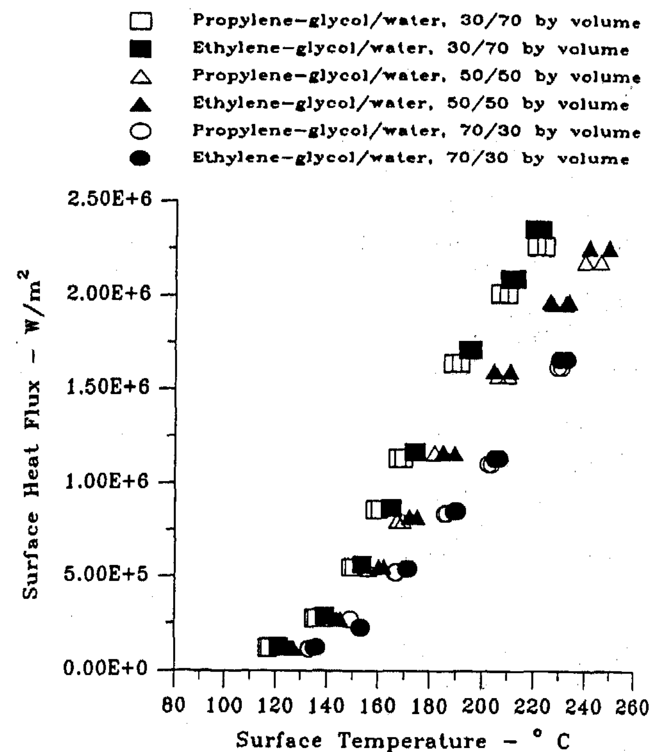


Figure 7 Concentration comparisons for ethylene-glycol/water and propylene-glycol/water at an inlet velocity = 0.445 m/s
T_{inlet} = 85° C, P_{system} = 205 kPa

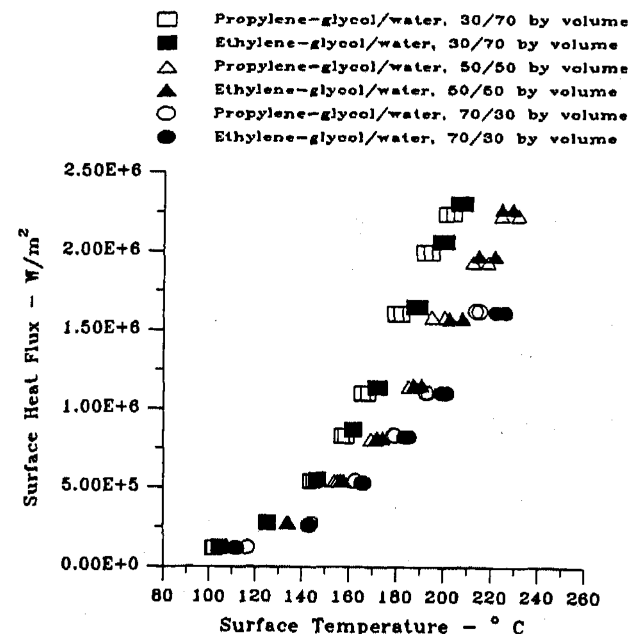


Figure 8 Concentration comparisons for ethylene-glycol/water and propylene-glycol/water at an inlet velocity = 1.33 m/s
T_{inlet} = 85° C, P_{system} = 205 kPa

70 % and 30 % coolant concentrations. Increasing the concentration of coolant raises the saturation temperature and reduces the effective heat transfer. Table 2 presents the mixture saturation temperature for each concentration tested.

For reference, the saturation temperature for 100 % water at the test pressure is 120 °C. From these

figures, it can be seen that the variation of wall temperature with concentration at the higher flux levels is approximately 40 °C for both coolant mixtures. Comparing the performance of propylene-glycol/water to ethylene-glycol/water, there is little or no difference as the concentration changes from 30 % glycol to 70 % glycol. At the higher velocities, the propylene-glycol mixture has a slightly lower surface temperature at the 30 % concentration; however, this difference (approximately 10 °C) is not considered significant.

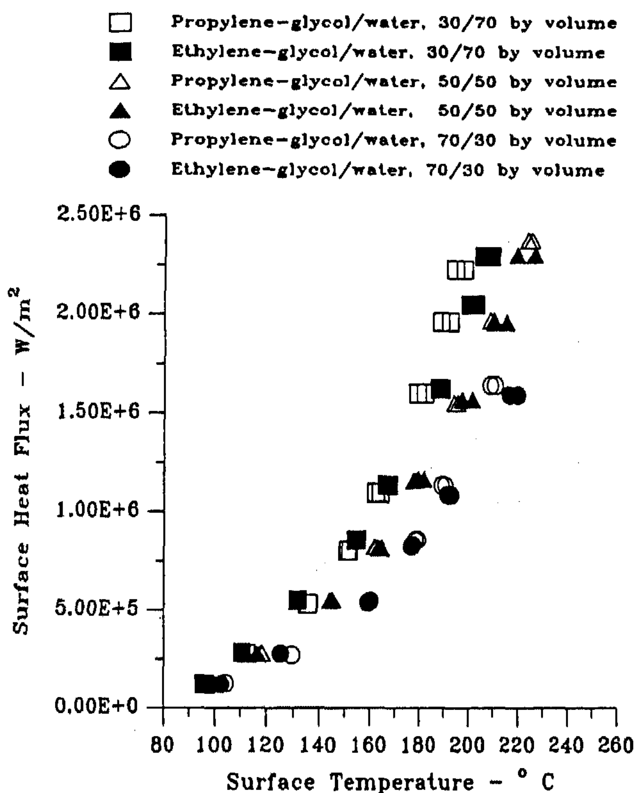


Figure 9 Concentration comparisons for ethylene-glycol/water and propylene-glycol/water at an inlet velocity = 2.67 m/s T_{inlet} = 85° C, P_{system} = 205 kPa

CONCLUSION

An experimental program has been conducted to examine the thermal performance of propylene-glycol/water and ethylene-glycol/water engine coolant mixtures over a wide range of heat fluxes and mixture concentrations.

The results of these experiments indicate that both coolant mixtures have similar performance with respect to surface temperatures.

For normal engine operating conditions of two atmospheres and 85 °C inlet temperature, two phase heat transfer exists over most of the heat flux range. At low inlet velocities, this two phase heat transfer involves both subcooled boiling and saturated boiling. Increasing the inlet velocity eliminates saturated boiling; however, even at the highest inlet velocity (2.67 m/s), subcooled boiling exists when the surface heat flux exceeds 0.6 MW/m². This result is particularly important because subcooled boiling can affect the hydrodynamics when alternate flow paths exist. This condition was not simulated in these tests.

Since boiling is an important aspect of heat transfer in engines with aqueous coolant mixtures, the effects of off design conditions such as lower system pressures and higher inlet temperatures are being considered in a separate test program.

The effect of surface finish is also under examination in tests planned using a cast iron test section.

ACKNOWLEDGMENTS

This research was sponsored by a grant from the ARCO Chemical Co. The authors also wish to recognize the support of the shop staff in the Department of Mechanical Engineering at Villanova University.

REFERENCES

Bhowmick, S., Branchi, C., McAssey, E.V., and Gollin, M., 1996, "Heat Transfer Performance of Engine Coolants Under Subcooled Boiling Conditions," ASME ICE-Vol. 26-2, 1996 Spring Technical Conference.

Finlay, I.C, Boyle, R.J., Pirault, J.P., and Biddulph T., 1987, "Nucleate and Film Boiling of Engine Coolants Flowing in a Uniformly Heated Duct of Small Cross Section," SAE Technical Paper Series 870032.

Gehres, E., 1963, "An Analysis of Engine Cooling in Modern Passenger Cars," SAE National Automobile Meeting, Detroit, MI, paper No. 660C.

McAssey, E. V., Stinson, C., and Gollin, M., 1995, "Evaluation of Engine Coolants Under Flow Boiling Conditions," Proceedings of the ASME Heat Transfer Division, HTD-Vol. 317-1, p. 193-200.

970943

Modelling of Car Dynamic Frontal Crush

Denis P. Wood and Stephen Mooney
Wood & Associates

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper examines the dynamic displacement-crushing force and dynamic displacement-absorbed energy behaviour of eight cars in full width barrier, 45% overlap rigid barrier and 30^0 angled barrier tests at 56, 50 and 56 kph respectively. This study shows the frontal crush behaviour of these cars can be divided into three regions or zones of constant force, these zones being associated with crushing of the front structure as far as the engine, the engine and rear front structure and the occupant compartment. The highest average crushing force is associated with crushing of the engine and rear front structure with lower average crushing forces required for the extreme front of the car and for the occupant compartment.

It is hypothesised that the energy absorbed-dynamic displacement behaviour in the full width barrier test represents the energy absorbed-mean displacement for all other crush configurations. This hypothesis is combined with the geometric crushing model for the car front (Wood 1993, 1995) to calculate the responses in the 45% overlap and the 30^0 angled barrier tests for one of the cars. A high degree of correlation is obtained. Further evaluation is needed to verify the hypothesis.

INTRODUCTION

This paper examines the frontal crushing behaviour of eight cars in full width and 30° angled barrier tests at 56 kph and in 45% overlap rigid barrier tests at 50 kph. The purpose is to obtain an insight into the overall crushing behaviour of car front structures and the manner in which this behaviour varies with depth of crush. The evaluation is carried out by examining the force-displacement and the absorbed energy-displacement responses of the cars. Also a model (Wood 1993, 1995) for frontal crush of cars is applied to obtain the energy absorbed responses in the 45% overlap and 30° angled barrier tests from the full width barrier response and these computed responses are compared with the actual responses obtained in the tests.

CARS EXAMINED

The details of the eight cars examined are detailed in Table 1. All are front wheel drive and with the exception of car 3 which has a longitudinal engine all are fitted with transverse engines.

Table 1 - Vehicle Data

Car	Curb Mass Kg	Length m	ds/L	de/L	db/L
1	860	3.96	0.033	0.081	0.220
2	809	3.96	0.045	0.091	0.250
3	1010	4.41	0.030	0.059	0.249
4	810	3.63	0.030	0.094	0.218
5	1090	4.56	0.026	0.103	0.208
6	990	4.41	0.035	0.084	0.236
7	745	3.70	0.052	0.086	0.238
8	1260	4.71	0.074	0.128	0.244

The ratios ds/L, de/L and db/L are the normalised distances from the fronts of the cars to the start of the two longitudinal members in each car, to the front of the engine and to the front of the bulkhead/firewall which divides the occupant compartment from the car front structure.

TEST RESULTS

DYNAMIC DISPLACEMENT RESPONSES

Figures 1 to 3 inclusive show the normalised crushing force-dynamic displacement responses of Car 4 in the full width, 45% overlap and 30° angled barrier tests. Figures 5 to 7 show the corresponding averaged responses for all eight cars. The full width barrier responses show that there is an initial peak force, the position of which corresponds with the location of the fronts of the longitudinal members and the initiation of their collapse.

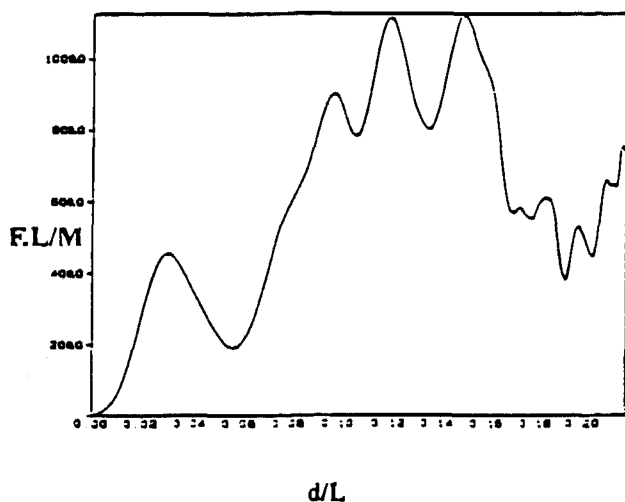


Figure 1 Full Width Rigid Barrier Response - Car 4

This is followed by a displacement over which there is a low force after which it climbs to a high force regime where the force oscillates about an average level before dropping to a lower force level.

The 45% overlap and 30° angled responses are similar but the initial low force region occurs over larger displacements before rising to a high force. This behaviour pattern was found for all cars with the exception of car 6 which did not show the reduction in force as maximum displacement was reached, the reason for which will be discussed later.

Figure 4 shows the absorbed energy-dynamic displacement response in the full width barrier test for car 4. Three distinct regions are immediately apparent. Each region can be represented by a straight line the slope of which is the average force in the respective region. The 45% overlap and 30° angled barrier responses for car 4 showed similar characteristics as did all the other cars in the full width, 45% overlap and 30° angled barrier tests (with the exception of car 6 which showed distinct regions 1 and 2 only).

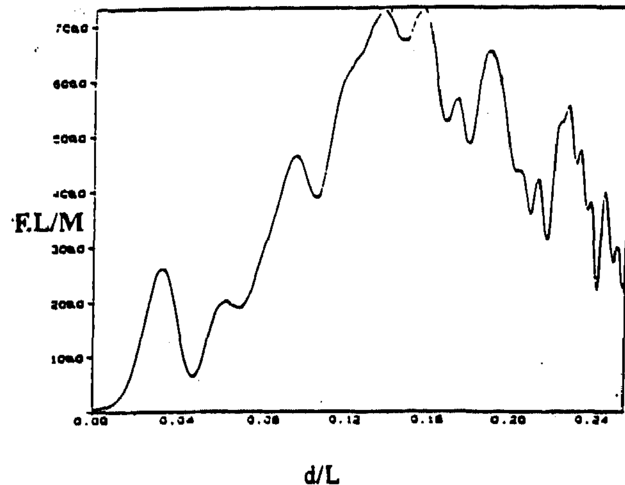


Figure 2 45% Overlap Response - Car 4

For car 4 in the full width barrier test the transition between regions 1 and 2 at a d/l of 0.087 is 93% of the

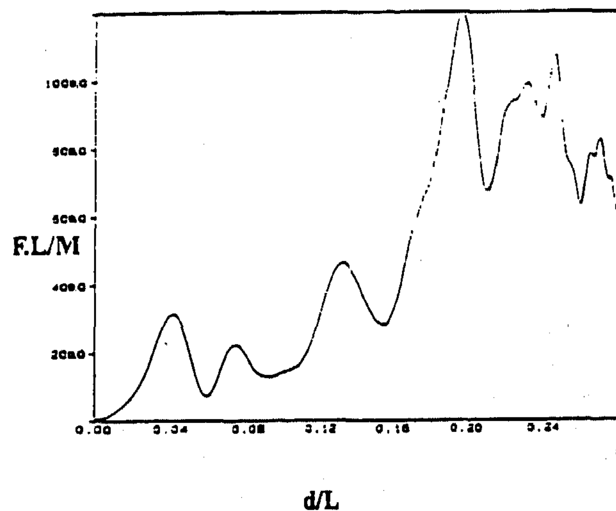


Figure 3 30° Angled Barrier Response - Car 4

distance to the front of the engine. The transition between regions 2 and 3 occurs at d/L of 0.164 which is 75% of the distance to the firewall/front bulkhead. Statistical analysis

for all eight cars shows that the position of the initial force peak corresponds with the location of the front of the longitudinal members and that the transition point between regions 1 and 2 in the full width barrier tests are, on average, 75% of the distance to the front of the engine and that the transition between regions 2 and 3 is at 75% of the distance to the firewall/front bulkhead.

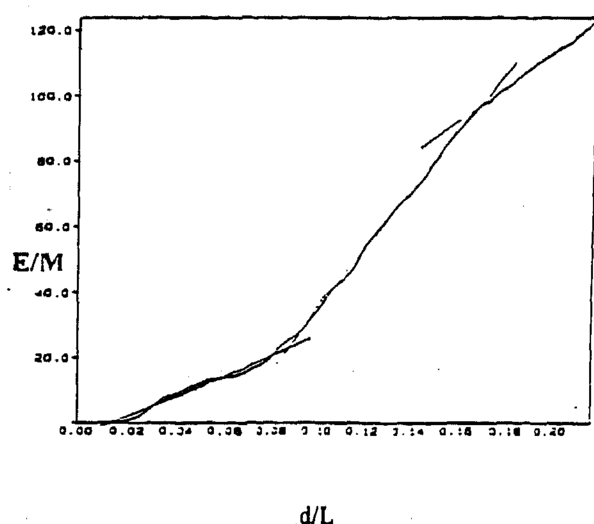


Figure 4 Full Width Absorbed Energy Response - Car 4

Jones et al (1990) shows that the depth of crushed material behind the crush face of tubes is such that the distance to the crush face is, at most, 75% of the total distance to the rear of the crushed material. When the depth of the crushed material immediately behind the crush face is taken into account and using the proportionality factor from Jones et al (1990) then for uniform full width crushing of the car front the transitions between regions 1 and 2 and between regions 2 and 3 correspond to the front of the engine and the firewall/front bulkhead respectively.
Therefore we can conclude that region 1 and the initial average force corresponds with crushing of the two longitudinals on their own with little contribution from the rest of the structure, that region 2 and the highest average crushing force is associated with crushing of the engine and rear front structure of the car front while region 3 and the lower average crushing force in this region is associated with crushing of the occupant compartment.

FULL WIDTH BARRIER AVERAGE FORCES

Table 2 shows the average forces in region 2 - the engine and rear front structures for each of the cars and the ratio of average forces between this region and regions 1 and 3 respectively. With the exception of car 6 which has a high region 1 force at 85.3% of region 2 the other seven cars have low region 1 forces which average 34% of their

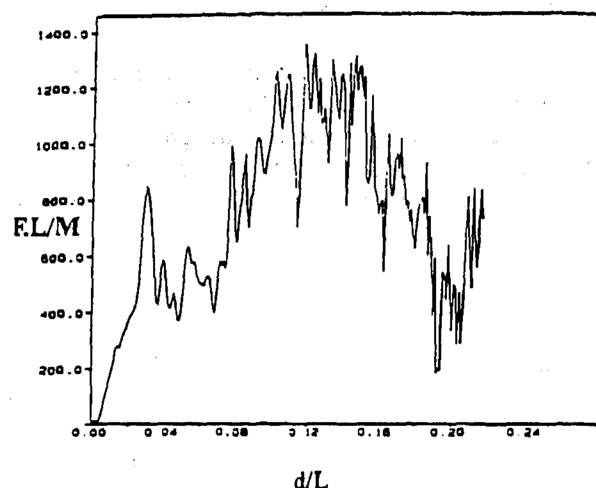


Figure 5 Average Full Width Response - All Cars

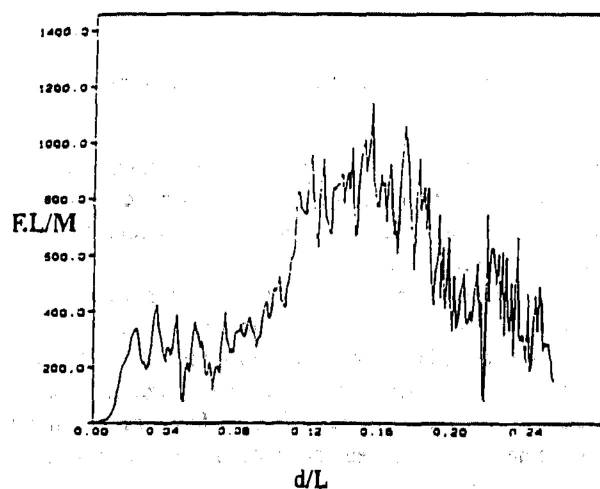


Figure 6 Average 45% Overlap Response - All Cars

region 2 values. All cars with the exception of car 6 showed evidence of a lower region 3 force. However the displacement in this region was sufficiently large for only four of the cars to compute the average force in region 3.

One explanation for the absence of evidence of a lower force region 3 for car 6 is that the crushing was not

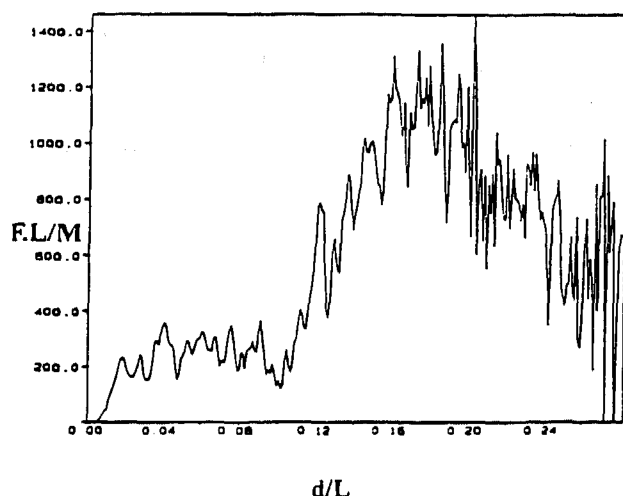


Figure 7 Average 30° Response - All Cars

sufficient to reach this region and involve the occupant compartment in the crushing process. The maximum crush in the full width test was 58% of the distance to the firewall/front bulkhead while it was 69% in the 45% overlap test i.e. both were less than the 75% crush face/crush depth ration from Jones et al (1990) necessary for the crushed material to reach the firewall/front bulkhead.

Table 2 Full Width Barrier Average Forces,kN.

Car	Region 2 Force,kN	Force1 / Force2	Force3 / Force2
1	227.3	43.6%	63.1%
2	187.1	48.7%	-
3	190.1	34.5%	-
4	204.2	33.0%	57.7%
5	519.2	15.5%	55.2%
6	294.1	85.3%	-
7	258.7	30.0%	43.6%
8	496.5	32.8%	-

Table 2 also shows that there is a high degree of variability in the average force required to crush the engine and rear front structure of the cars from a low of 187.1 kN to a high of 519.2 kN, a 2.8 to 1 force range. The mean force is 297.2 kN.

Turning to the 45% overlap and the 30° angled barrier tests Table 3 shows the ratios of the average forces in each of the three regions in these two tests by comparison with the full width barrier test.

Table 3 - All Cars, 45% Overlap & 30° Force Ratios.

Region/Test	45%	30°
Region 1/Region 1 Full Width	60.9%	53.6%
Region 2/Region 2 Full Width	78.0%	100%
Region 3/Region 3 Full Width	55.3%	95.2%

This data shows that the average forces in each region for the 45% overlap test are less than those in the full width barrier test. However in the 30° angled barrier test only the region 1 force is significantly less than for the full width test. In regions 2 and 3 the average forces in the 30° angled barrier tests are similar to those obtained in the full width barrier tests. An extended analysis is given in Wood (1996).

DYNAMIC DISPLACEMENT RESPONSES

DYNAMIC CRUSH MODEL

It is hypothesised that the absorbed energy-displacement characteristic of any individual car type in the full width barrier test represents the absorbed energy-average displacement characteristic for that car type in all frontal crush configurations. Wood (1993, 1995) has previously proposed that in frontal crushing the portion of the car not in direct contact with the barrier or opposing car i.e. the induced crush portion can also be regarded as crumpling rearward, see Figure 8, and that the average crush depth should be calculated over the deformed front profile in the manner shown in Figure 9.

APPLICATION OF CRUSH MODEL TO CAR 4

Figure 10 shows the absorbed energy per unit mass for car 4 as a function of dynamic displacement in the full width barrier test i.e. as a function of average dynamic displacement as per hypothesis. The relation between the dynamic displacement of the C.G. in the 45% overlap and 30° angled barrier tests and the corresponding average dynamic displacement is calculated on the basis of the car front deforming in a simple geometric manner, refer Wood (1993). For the 30° angled barrier tests the car front was considered to progressively increase in angle until 30° was reached after which the car maintained a 30° angle while

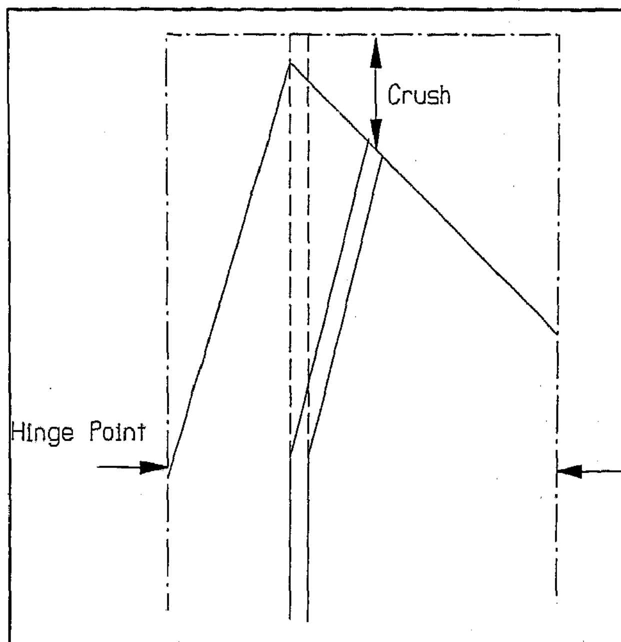


Figure 8 Crush Model

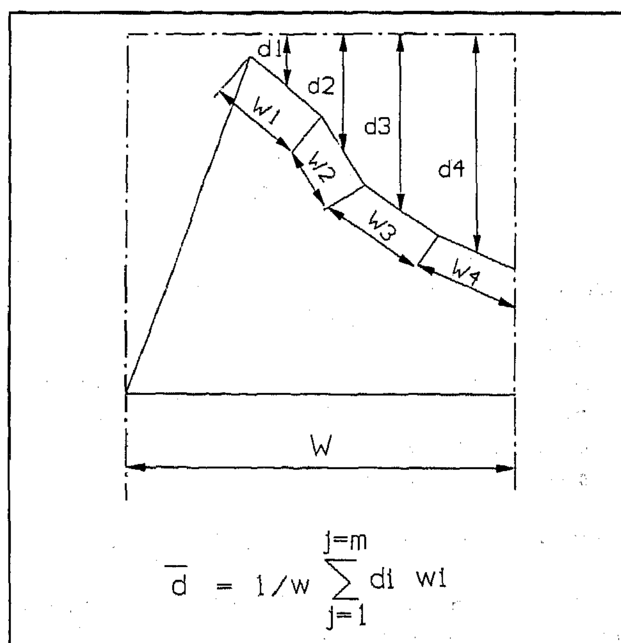


Figure 9 dbar Calculation Method

continuing to crush to maximum dynamic displacement.

The comparisons between the actual and calculated absorbed energy-dynamic displacement are shown in Figures 11 and 12 for the 45% overlap and 30° angled barrier tests respectively. There is a high degree of agreement between the actual and calculated responses in both tests. Towards maximum dynamic displacement there is some deviation between the computed and actual responses. However in the 30° angled barrier comparison the calculated response is less than the actual response whereas the reverse occurs in the 45% overlap comparison, i.e. the calculated response is greater than the actual response.

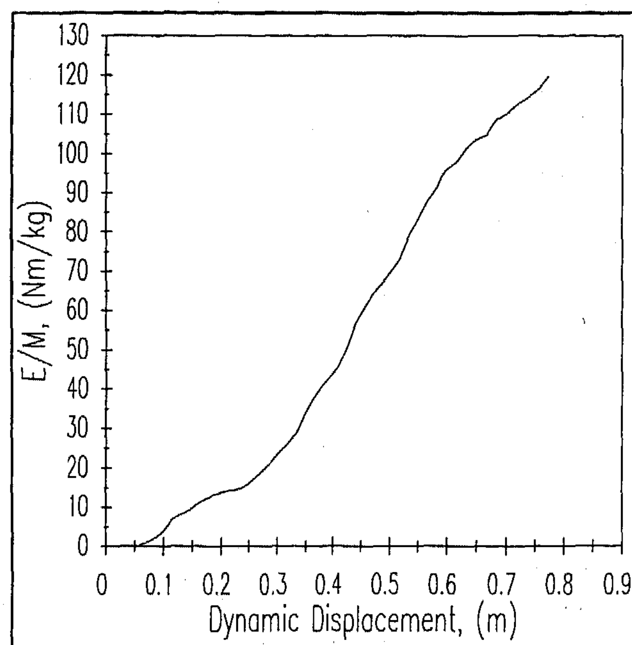


Figure 10 Car 4 Energy Absorbed -v- Displacement Full Width Barrier

DISCUSSION

Examination of the crush behaviour of eight cars in full width, 45% overlap and 30° angled barrier tests in terms of absorbed energy-dynamic displacement shows that the crushing can be divided into 3 regions or zones, each with a constant force. Comparison of the transition points between the regions or zones shows that the zones are associated with the immediate front of the car and the two longitudinal members, the engine and rear front of the car and the occupant compartment respectively. The data for the eight cars shows that the lowest force zone is the first i.e. the immediate car front and the plastic collapse of the longitudinal members, the highest force zone is the engine and the rear front structure while the occupant compartment has a lower crushing force level than the engine and rear front structure. The reason why collapse is progressive despite the occupant compartment having a lower crushing force than the engine cum rear front structure is due to dynamic inertia effects under impact conditions, Johnson (1970).

It is hypothesised that the absorbed energy-dynamic displacement behaviour in the full width barrier crushing

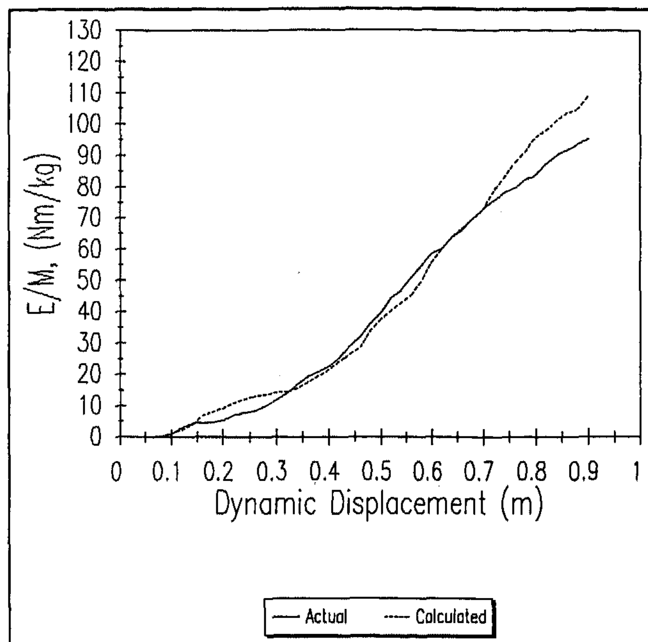


Figure 11 Car 4 Energy Absorbed -v- Displacement 45% Rigid Offset Barrier

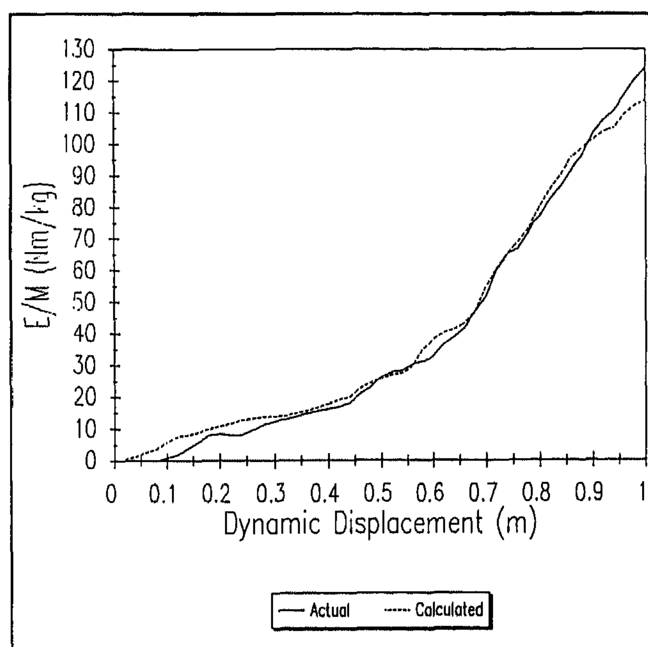


Figure 12 Car 4 Energy Absorbed -v- Displacement 30° Angled Barrier

represents the absorbed energy-average dynamic displacement in all other frontal crush configurations. Evaluation of the hypothesis combined with the geometric model, Wood (1993, 1995) for one car, car 4, shows a high degree of correlation between the actual and calculated dynamic responses in the 45% overlap and 30° angled barrier tests. Comparison of the hypothesis with the actual responses for other cars is needed to verify the model in the context of dynamic crush behaviour. The basic hypothesis of regarding the car front as crushing in a geometric manner and considering the full width barrier characteristic as being the characteristic for average crush have already been successfully applied to residual crush data for pole, offset and car to car collisions, refer Wood (1993, 1995).

APPLICATION TO ACCIDENT RECONSTRUCTION

Before this approach can be applied to the residual crush data obtained from crashed cars and to accident reconstruction a reliable method of correlating dynamic and residual crush for individual car types is required. Hight et al (1985) details residual and dynamic crush values for a number of cars. Wood (1992, 1995) derived a general correlation between normalised or non-dimensional residual and dynamic crush for the car population which can be used to estimate dynamic crush values from residual crush measurements when the data for a particular car type is not available. Previous analysis (Wood, unreported) showed that there was considerable variation in the residual/dynamic characteristics as between different car types.

The test data also indicates that care is needed in extrapolating the crush force response data to higher impact speeds. The full width barrier data indicates that the average full width barrier force during crushing of the occupant compartment varies between a low of 112.8 kN and a high of 286.6 kN. Also the full width barrier response data for numbers of car types is such that crushing of the occupant compartment is not sufficiently developed to allow average force levels for this region to be derived. Methods of using the 30° angled barrier and the 45% overlap test data to derive this force level is being examined. Consideration of high speed, 80/100 kph full width barrier tests of various car types in order to have sufficient data to reliably calculate speed from residual crush in high speed collisions is desirable in the light of the test data in this paper.

ACKNOWLEDGEMENTS:

We wish to thank Mr. P. Hight for his helpful advice and comments.

REFERENCES:

1. Hight P.,Lent-Koop B. and Hight R. "Barrier Equivalent Velocity, Delta V and Crash 3 Stiffness in Automobile Collisions". SAE Technical Paper 850437, 1985.

2. Johnson W. and Mamalis A.G. " Crashworthiness of Vehicles" Mech. Eng. Publ., London, 1978.

3. Jones N. and Birch R.S. "Dynamic and Static Axial Crushing of Axially Stiffened Tubes" J. of Mech. Eng. Sc., Vol 204, No. C5, pp 204-210, 1990.

4. Wood D.P., "A model for the Frontal Impact Characteristics of the Car Population". Proc. of the 4th International Conference on Structural Failure, Product Liability and Technical Insurance, pp. 396-406, Vienna, Austria, July, 1992.

5. Wood D.P., Mooney S.J. and Doody M. "Application of a Generalised Frontal Crush Model of the Car Population to Pole and Narrow Object Impacts" SAE Technical Paper 930834, 1993.

6. Wood D.P. "Determination of Speed from Crush" Chapter 7, Forensic Accident Investigation: Motor Vehicles, T.L. Bohan and A.C. Damask eds, Publ. Michie Butterworth, 1995.

7. Wood D.P., O'Riordain S. and Vallet G. "Car Frontal Crush - A New Perspective" 1996 International IRCOBI Conference on The Biomechanics of Impact, pp 201-210, 1996.

SAE #970943
Reviewer's Discussion
by Philip V. Hight, P.E., Accident Research, Santa Barbara
Modelling of Car Dynamic Frontal Crush
Denis P. Wood, Stephen J. Mooney, Authors

This paper proposed a dynamic model of automobile crush using data from 90 degree barrier impacts at 50 and 56 km/h (30 and 35 mph). At this time a comparison for the other cars has not been accomplished. This study had divided the frontal crush into three zones of constant force. This would enable researchers to calculate responses in 45% overlap and 30 degree angled barrier. It is hoped by this reviewer that by the presentation that further analysis will be available for comparison concerning the other seven front wheel drive automobiles. A preliminary review indicates a high degree of variability in the average force required to crush the engine and rear front structure of the cars, nearly a 3 to 1 force range. The concept of the paper is interesting but it seems premature to evaluate the success of the hypotheses until after further cars have been analyzed. The basic hypotheses that the front of the car crushes in a geometric manner and can be applied to crush profiles from pole impacts may be disputed by other researchers. It appears probable that this concept can be applied to residual crush of offset and car to car collisions.

970946

Passenger Vehicle Braking Performance with a Disabled Vacuum Power Booster

Wade Bartlett
Mechanical Forensics

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

When the vacuum-powered brake booster in a passenger vehicle becomes disabled, the brake force gain of the system is reduced significantly, and the brake pedal force required to lock the tires increases beyond the ability of some adults. In such cases, the maximum braking deceleration acheived by those individuals will be something less than the upper boundary as defined by available traction.

This paper's goal is to review the design of vacuum boosters, the literature concerning human ability to depress a brake pedal, and FMVSS 105 requirements which must be met by vehicle manufacturers, and to present performance data with and without the booster operational for four passenger vehicles. Furthermore, the application of this information to accident investigations involving disabled boosters will be discussed.

INTRODUCTION

The vacuum power booster (booster) is the most common form of power assistance unit now found in passenger vehicle brake systems. It derives motive force from an air pressure differential across a large surface area, and significantly reduces the force an operator must apply to the brake pedal to achieve satisfactory deceleration.

Vacuum booster systems are fairly reliable, but can fail. One study completed in 1977 concluded that approximately one failed booster per year per 11,700 vehicles could be anticipated.[1] Though passenger vehicle component reliability has improved since the 1970's, a finite number of failures can still be expected. If the booster becomes disabled, the operator must apply a much greater force to the brake pedal to achieve a given deceleration. If such a failure occurs unexpectedly, an accident could result.

It is the goal of this paper to outline the design and operation of vacuum booster systems and their possible failure modes, review the portion of the Federal Motor Vehicle Safety Standard 105 (FMVSS 105) which lists braking performance criteria new vehicles must meet in the event of a booster failure, and review of some of the literature concerning the brake pedal application forces of which drivers have been found capable. Results of brake performance testing for four passenger vehicles, with and without their boosters operational, will be presented, and the application of this information to accident investigations involving disabled boosters will be discussed.

Hydraulicly activated power boosters (hydro-boosters) are occasionally utilized in passenger vehicles, but often share a hydraulic pump and pressure accumulator with the power steering system, and incorporate a variety of plumbing arrangements which vary by manufacturer and vehicle model. These complications prevent the sort of general treatment possible for vacuum-powered systems which are simpler and less expensive to manufacture, and have enjoyed more widespread use. There is overlap in the performance and regulation of the two system types, but this paper will be devoted exclusively to consideration of the vacuum-type.

The inclusion of anti-lock braking systems (ABS) in passenger vehicles in recent years has had little effect on the design and operation of brake boosters. The pedal force required to engage the ABS is significantly increased with a disabled booster, but the ABS will operate normally if called upon.

This paper will not address issues related to surfaces having a low coefficient of friction, such as snow or ice. All testing was conducted in conditions where available traction was greater than the decelerations attempted. Deceleration resulting from tire forces can never exceed the traction available at the tire / road interface, regardless of brake system design, booster condition, or ABS operation.

BOOSTER DESIGN AND OPERATION

The purpose of a vacuum power booster is to reduce the brake pedal application force required to generate sufficient pressure in the hydraulic system to stop the vehicle at a particular rate of deceleration. Figure 1 shows how deceleration varies pedal with force for a typical passenger vehicle. Tire rolling resistance and other minor sources of drag provide a slight deceleration even without brake application, shown as the area adjacent to 1 in Figure 1. A modest pedal force is required to overcome internal resistances and various springs in the brake system. This system "kick-in point" is

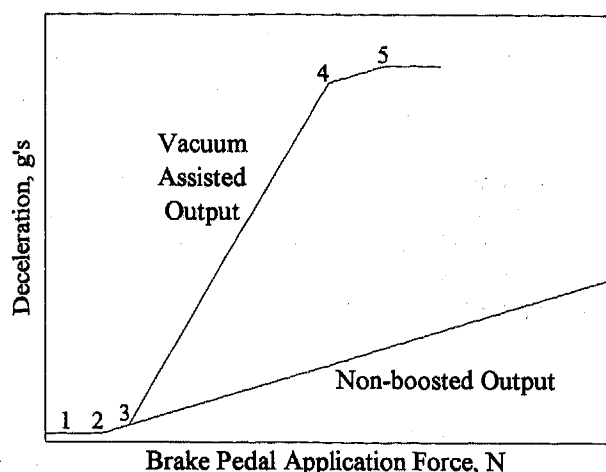


Figure 1: Vehicle deceleration as a function of brake pedal application force; 1 deceleration from minor drag sources, 2 brake system kick-in point, 3 booster crack point, 4 booster run-out point, 5 maximum deceleration.

shown as point 2 in Figure 1. Once the pedal force exceeds the kick-in force, deceleration increases in an essentially linear fashion in a manual (unboosted) brake system or in a booster-equipped system when the booster is not functional. It should be noted that the slope of the performance curve for a manual system will likely be steeper than that for a system with a disabled booster.

The incorporation of a booster adds two features to the curve. Some pedal motion is required to operate the valves controlling the admittance of air into the rear chamber, allowing the booster to begin to augment the pedal application force. This point, point 3 in Figure 1, is commonly referred to as the "crack point," so named because at this point the valves first "crack open." The crack point is quite low and may not be noticeable in full-vehicle test data. The second feature added by a booster is the "run-out point" or "knee point," shown as point 4 in Figure 1. This is the point at which the rear chamber has reached atmospheric pressure, and no additional force augmentation is available. The names are derived from the booster's having "run out" of additional boost, and from the shape of the performance curve. Beyond the run-out point, additional hydraulic pressure is generated at the same rate as for the "unassisted" case.

Point 5 in Figure 1 shows the plateau defined by the available traction. This plateau will drop off once wheel slip increases beyond the point at which peak traction occurs. A discussion of this phenomenon is beyond the scope of this paper. Further information about how traction varies with slip can be found in most handbooks which discuss tire performance, including those by Gillespie and Bosch.[2][3]

Physically, a vacuum booster is a canister, typically measuring 200 to 270 mm in diameter, and 40 to 70 mm in depth.[4] It contains at least two isolated chambers, one forward and one rear. When not in use, both chambers are exposed to the partial vacuum of the intake manifold.* When the operator presses on the brake pedal, the booster input shaft moves forward, isolating the rear chamber from the partial vacuum in the forward chamber, and allowing a controlled amount of air at atmospheric pressure to enter the rear chamber. This creates a pressure differential of up to approximately 85 kPa.[3][5] Typical actuation force gain in the booster is 4:1 to 5:1,[6] causing the assisted output to rise much more quickly than the unassisted output. Thus, the force applied to the master cylinder piston is 4 to 5 times greater than the force applied to the booster input shaft. A drawing of the principle components of a vacuum booster unit is shown in Figure 2. Some designs incorporate a second "tandem" diaphragm for additional boost where the unit diameter must be minimized.

The booster input rod operates directly on the master cylinder, regardless of the booster's operation; therefore, in the event of booster failure, the operator can still apply force to the master cylinder, though the pedal application force to lock the wheels may more than quadruple from about 160 N to over 800 N.

A booster fails to augment brake pedal application force when there ceases to be a pressure differential across the diaphragm. Such a loss of pressure differential can occur at a variety of onset rates, depending on the nature of the breach in vacuum system integrity. If the failure allows a significant amount of air at atmospheric pressure into the forward chamber (for example, if the check valve body breaks off or falls out of the booster housing) there may not be any residual capacity, resulting in a sudden loss of all power assist.

Other disablement modes may not be as sudden, as when the engine ceases to run or when the vacuum system between the engine and the check valve is breached. In these cases, the check valve should preserve a partial pressure differential for one or more power-assisted brake applications, depending on the severity of each application. Additional failure modes may only result in a slow degradation of power assist, as in a small tear in the diaphragm material.

Though FMVSS 105 does not explicitly discuss the reserve capacity of boosters, the Society of Automotive Engineers (SAE) has published Standard J1808, Vacuum Power Assist Brake Booster Test Procedure,[7] which offers a protocol to test the reserve capacity of a booster once disconnected from the vacuum source. Performance requirements are still under development as SAE standard J1902, at the time of this writing. The protocol, interestingly enough, tests the number of times that a booster supplies some

* The intake manifold vacuum is generally used; however, some types of engines, including diesels, do not develop enough partial vacuum to power a booster. In these cases, an auxiliary vacuum pump is required.

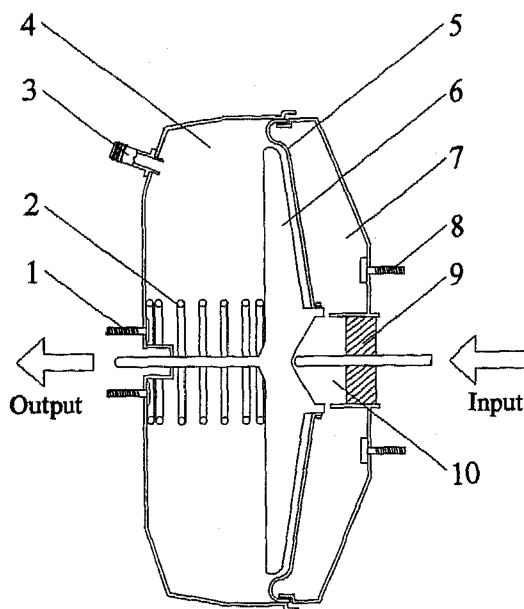


Figure 2: Simplified schematic of a typical vacuum power booster; 1 master cylinder mounting bolts, 2 return spring, 3 vacuum connection and check valve, 4 partial vacuum chamber, 5 diaphragm, 6 diaphragm plate, 7 rear chamber, 8 booster mounting bolts, 9 inlet air filter, 10 air control valve hardware location.

assistance during applications which only reach up to 25% to 30% of the knee point. This test does not imitate a panic stop.

When a booster becomes disabled the brake pedal remains high, and feels "hard" to the operator. Repeated pumping of the brake pedal, as many drivers have learned to do in case of brake troubles, exacerbates the problem by depleting whatever reserve vacuum remains in the front chamber.

In addition to component failure, a loss of pressure differential can be caused by incorrect assembly of the system, two recent examples of which caused vehicle manufacturers to issue vehicle recalls.[8][9]

Though not a factor in analyses involving diminished braking performance, additional booster failure modes exist which can cause the brakes to be applied, usually lightly, despite no application of the brake pedal.

FMVSS 105 REQUIREMENTS

Part 571 of the Code of Federal Regulations includes FMVSS 105, which outlines the requirements for hydraulic brake systems.[10] Though there are several testing options from which manufacturers can choose to meet the requirements for a disabled booster, a typical test is outlined in Section 7.10.2: With the booster depleted of all reserve capacity, a passenger vehicle must be capable of stopping from 96.6 kilometers per hour (kph) (60 miles per hour) at a deceleration rate of 2.13 meters per second per second (m/s^2) (7 feet per second per second), using a pedal application force of not more than 667 N (150 lb). The equivalent stopping distance is 168.9 m (554 feet), or a deceleration of 0.218 times that of gravity (0.218 g's).

An array of additional requirements address different situations. Most requirements involving fully functional brake systems (with or without the inclusion of some form of power boost system) allow pedal application forces up to 667 N (150 lb) to meet prescribed stopping distances.

OPERATOR CAPABILITY

The Human Factors Design Handbook [11] recommends an upper bound for brake pedal force of 267 N (60 lb.), though it states that "When power assist can be provided, this force can and should be considerably reduced." It has been this author's experience during testing for this paper, that control forces below 130 N were noticeably easier to modulate accurately than higher forces. Control forces above 550 N (125 pounds) were found very difficult to apply and modulate smoothly within 10% of a given desired value.

Radlinski and Price [12] measured the maximum average sustained pedal force which 105 women, ages 18 to 62, could apply over five seconds. The maximum force which could be applied by their female test subjects was measured first in a vehicle with a normal brake pedal height, then in the same vehicle with a reduced pedal height, then in a second vehicle with a normal pedal height. In the first vehicle, with the pedal at a normal height, the results ranged from 378 to 1,290 N (85 to 290 pounds), with a mean of 778 N (175 pounds). Twenty percent of the volunteers could not exceed an average sustained pedal force of 605 N (135 pounds). In a second vehicle, also with the brake pedal at its normal height, the same group of women generated average sustained brake pedal forces of between 400 N and 1,379 N (90 lb and 310 lb), with a mean of 863 N. In the first vehicle, when the brake pedal was lowered to simulate a partial system failure, the mean maximum pedal force measured was 587 N, significantly lower than the "normal pedal height" value of 778 N.

Test results reported in the Human Factors Design Handbook indicate that the distance between the seat and pedal plays an important role in operator capability, as does the relative height of the pedal to the seat.

Mortimer et al. [13] used a test buck with a rigid seat to measure the ability of 276 females to apply a brake pedal. When coached to press as hard as possible, the test subjects were able to apply forces from 270 N to more than 1,300 N (60 to 290 lbs), with a mean of 860 N (192 lbs) Five percent of those tested could not exceed 448 N (100 pounds), while twenty percent could not exceed 623 N (140 pounds). These data correlate well with Radlinski's findings.

Mortimer et al. additionally reported a range of 506 to well over 1,340 N (113 to well over 300 lbs) for males when

tested in the same fashion. The fiftieth percentile male was capable of pedal forces in excess of 1300 N (290 lbs). The fifth percentile value for males was 823 N (185 pounds).

Eaton and Dittmeyer [14] tested the pedal-force application ability of female operators when faced with an unexpected increase in control force required to stop the vehicle. They determined that a fifth percentile female was capable of applying up to 448 N, which is the same value reported by Mortimer et al.

The above data indicate that males are more likely to be *capable* of exerting enough brake pedal force to achieve maximum braking; however, the actual performance of every driver, whether male or female, is dependent on their response to an unfamiliar situation.

This author has investigated three cases where the booster became disabled in a male-operated vehicle. In two of those cases there was no warning and the failure occurred as they entered intersections. Both drivers described their vehicle's problem as "brake failure," reporting that they had "no brakes." Both of these operators reported that they responded to the "hard" pedal by pumping it, which in one case effectively exhausted any partial vacuum remaining in the booster, and in the other case had no effect, as the failure was in the diaphragm itself. In the third case, the operator knew of the failure prior to attempting to stop, and was able to apply enough pedal force to control the vehicle, but reported that he feared the forces required would damage the seatback or the steering wheel.

VEHICLE TESTING

Two sets of tests were performed on a 1995 Saturn SL2, a four-door sedan. For all tests, a Valentine Research G-Analyst [15] was used to measure and record the accelerations generated during each stop. The G-analyst's internal parameters were set to those described in the instructions as being appropriate for a slightly sporty sedan. The highest acceleration value maintained for at least 0.5 seconds was recorded. A 4 kN load cell was affixed to the brake pedal, to measure pedal application force. The digital readout was placed such that it could be viewed easily by the operator during testing. The vehicle was accelerated to about 40 kph and shifted into neutral before brake application.

During the first set of tests, the brake pedal application force was applied manually by the operator, with the steadiest possible pedal force, while for the second set of tests an air piston mounted between the steering wheel and load cell was used. The piston was activated using a compressed air tank carried in the vehicle.

Figure 3 shows the braking performance of the Saturn with and without the vacuum booster functional. All data are included, illustrating that the results of the two pedal application methods were not significantly different.

Some of the features described in Figure 1 can be seen

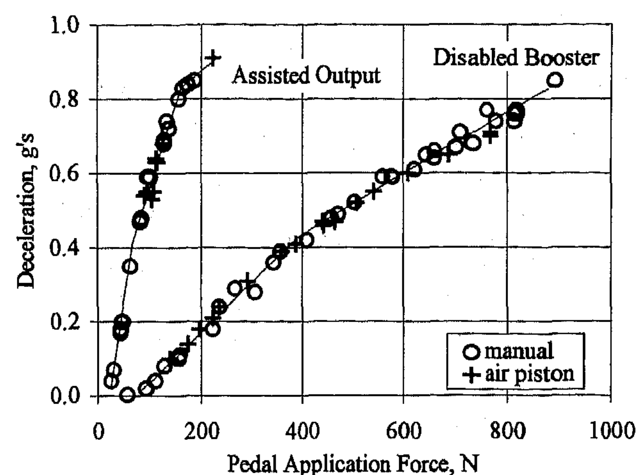


Figure 3: Deceleration of a 1995 Saturn SL2 with and without the vacuum booster functional.

in Figure 3, including the run-out point, at approximately 0.84 g's on the "Assisted Output" curve. At the low end of the curves, the two main features, the kick-in point and the crack point, become blurred, and cannot be discerned accurately.

The curves in Figure 3 show an additional knee near 0.4 g's, which is a result of the inclusion of a proportioning valve in the vehicle's braking system. Among other things, proportioning valves modify rear wheel brake pressure after a predetermined input pressure to optimize the front-to-rear brake balance. This balance is a compromise which attempts to account for dynamic variations which result from braking on surfaces offering different traction and for variations in vehicle load. The slope-change is a function of valve design, and can be seen in some whole-vehicle test results more clearly than others, as will be shown below. A vehicle's proportioning valve may be combined with other valves in a single body, commonly known as a combination valve. Such valves are included in the brake system of virtually every modern passenger vehicle. Most vehicle handbooks offer a more detailed explanation of the purpose and operation of these valves.

Hyung Dae Kim and Younchul Rhim [16] offered a model to predict deceleration as a function of pedal force, which incorporated all the major elements of modern passenger vehicle braking systems. Although they showed an excellent correlation between their predictions and test results, the model required a significant amount of information which is not typically available to an accident reconstructionist, including exact performance characteristics for the proportioning valve and vacuum booster, caliper piston diameter, drum diameter, friction material parameters, pedal mechanical advantage, master cylinder diameter, etc. Though the model is not helpful to this analysis, testing which they conducted to support their work is. They tested the braking performance of two vehicles,

one of which they tested a second time with a variety of components altered or replaced. Figures 4, 5, and 6 present their findings, comparing braking performance with and without the booster operational as a function of pedal force.

Note that the effects of the proportioning valve are not as visible in Figures 4 and 5 as in the earlier example.

A summary of the data presented in Figures 3, 4, 5, and 6 is shown in Figure 7, along with markers to indicate the pedal force application ability reported by Mortimer, et al for both male and female operators. Despite the variations between vehicles, the four "Assisted Output" performance curves are similar, as are the "Disabled Booster" curves.

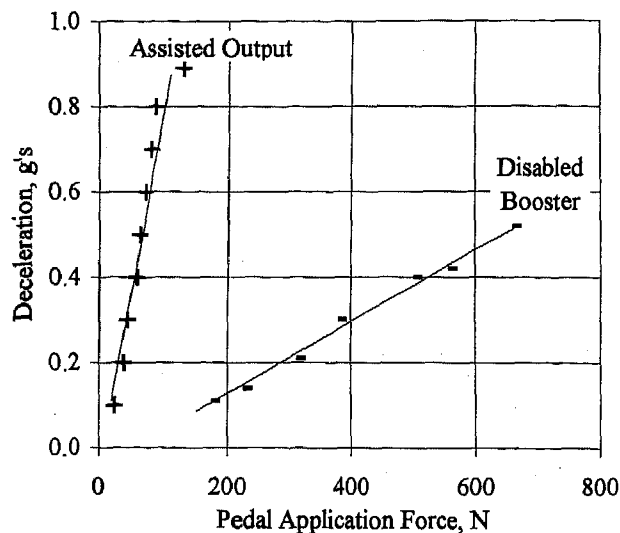


Figure 4: Deceleration with and without the vacuum booster functional, as reported in [16], "Test 1."

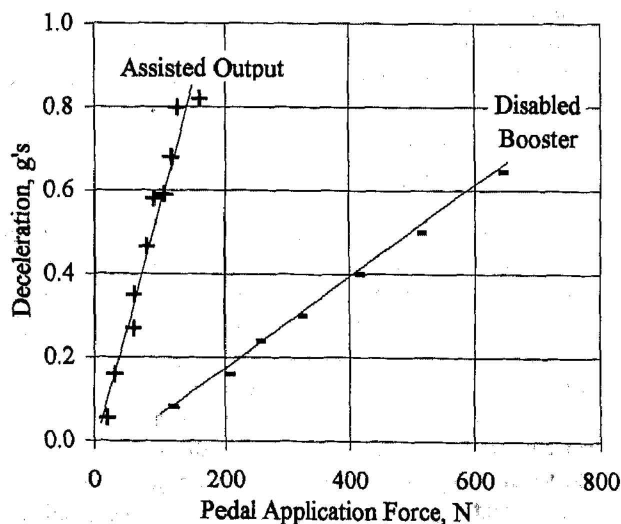


Figure 5: Deceleration with and without the vacuum booster functional, as reported in [16], "Test 2A."

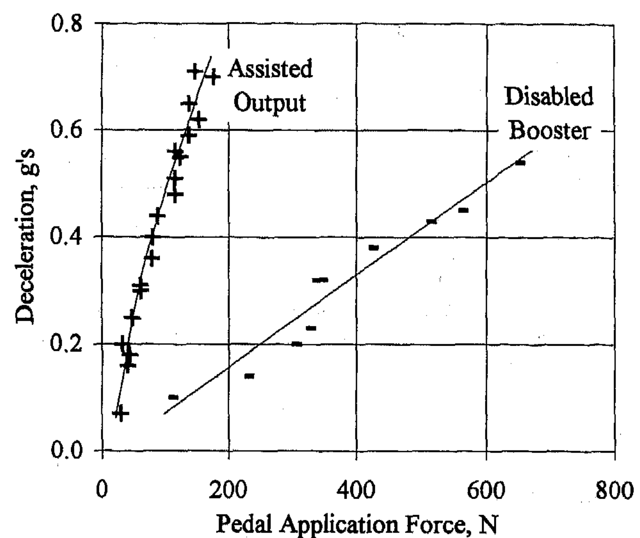


Figure 6: Deceleration with and without the vacuum booster functional, as reported in [16], "Test 2B."

DISCUSSION OF TEST PROCEDURE

The benefits of using the pneumatic piston, instead of manually applying the brakes, included a narrower data spread, and more accurate control over advance selection of the pedal application force. The piston was found to offer very little stiction, and the analog air tank pressure gage was found to be fairly accurate. Using the air pressure and the piston working area, it was possible to calculate the pedal application force within a few percent of that recorded by the loadcell. This method is sufficiently accurate for braking evaluation as it relates to accident reconstruction, and does not require an in-line loadcell.

Use of the pneumatic piston did, however, limit steerability during testing, and applied a significant force (up to 800 N) to the lower portion of the steering wheel ring. This did not result in any permanent damage to the test vehicle's steering wheel or steering column, but it did deflect the lower edge of the steering wheel ring by more than 4 cm. An alternative means of securing the piston should be sought when the possibility of interior damage must be avoided.

Despite the difficulty in modulating high pedal application forces, as noted earlier, the manual results compare favorably with the pneumatic piston results. The loadcell's digital display was not as informative as an analog device would have been, being essentially incapable of offering information concerning the rate at which the force was changing. A compact "pedal effort gage," incorporationg an analog dial and intended for repair facility brake system diagnostic testing is marketed by Kent-Moore Automotive Group.[17] It is anticipated that such a tool would provide

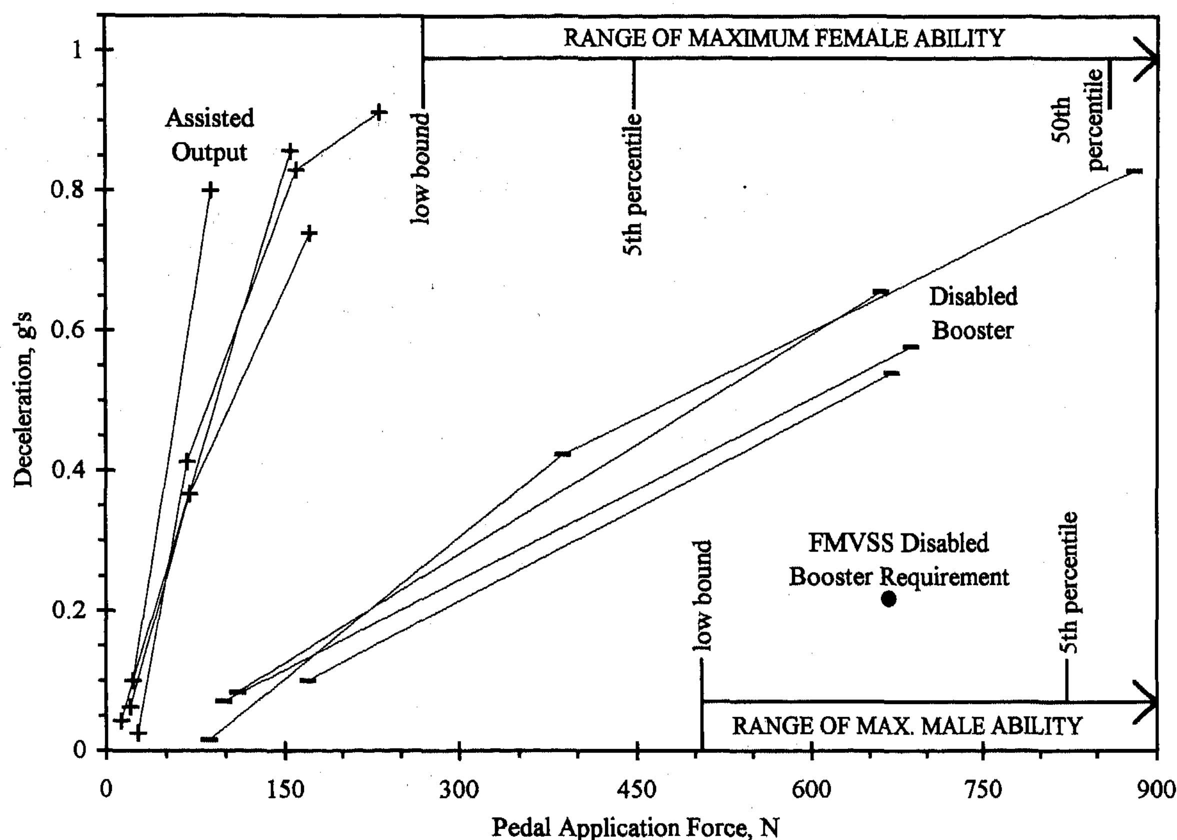


Figure 7: Composite plot of four vehicles' braking performance with and without the vacuum power booster disabled (as presented in Figures 3, 4, 5, and 6), showing ranges for male and female pedal application abilities, as reported by Mortimer, et al.[16]

sufficient accuracy for testing related to accident reconstruction, given other variables in such analyses. The added information supplied by the analog gage would probably improve the accuracy of manual testing, as compared to the use of a digital readout.

DISCUSSION OF RESULTS

The variability between operator capabilities and the variety of vehicle designs precludes determination of deceleration values a particular operator may have effected preceding an accident while faced with a failed vacuum power booster, without conducting specific testing. A female operator may be capable of generating maximum braking (ie, locking the wheels), but may also be only capable of generating sufficient brake pedal force to acheive 0.2 g's, depending on her own capabilities and on the particular design of the vehicle, as shown in Figure 7. A fiftieth percentile male operator is almost certain to be capable of generating sufficient brake pedal force to acheive maximum braking, while a male at the low end of the scale may only be capable of generating 0.4 g's.

Using statistical methods, Mortimer, et al offered a means of evaluating the likelihood that a person of a particular percentile ability would be capable of generating the pedal force required to slow their vehicle at the rate they desired. A statistical approach, however, was not the goal of this study.

In order to evaluate the possible brake pedal forces effected by a particular operator immediately prior to an accident, their ability to depress a brake pedal must be directly measured. Research noted earlier showed that the ability of operators to depress the brake pedal varies noticeably between vehicles and also with pedal height and seat location, thus the following recommendations are made: An operator's ability should be tested in an exemplar vehicle of the same interior geometry as the one they were operating during the accident. The seat should be in the same location. The booster should be depleted, so that the pedal height will be the same as at the time of the accident. If a force gage is mounted on top of the brake pedal, its height should also be considered.

The most accurate means of assessing the effect of the loss of power assistance on a vehicle's braking characteristics is to conduct direct testing.

With the two portions of the driver-vehicle system tested independently, the decelerations which could have been effected by a driver can be evaluated with some confidence. This process does not, however, take into account the possible effects the stress an impending accident may have on an operator, which must be considered, nor does it account for operator perception and reaction time. Further complicating an analysis of such accidents, some operators respond to a booster failure by pumping the brake pedal, which does not increase the brake line pressure significantly, but does expend precious seconds. The amount of time spent in inappropriate responses must be evaluated on a case-by-case basis.

Gillespie noted that only 20% of brake applications during normal driving resulted in decelerations greater than 0.2 g's, which would indicate that the control forces employed are generally below 50 N (referring to Figure 7). Based on the data presented in Figures 3 through 7, it can be deduced that if the booster became disabled, 80% of pedal applications would initally generate decelerations only barely more than those resulting from minor losses and rolling resistance. Many drivers would call an event of this nature "total brake failure."

In many driving situations (stopping in traffic for a red light, for instance), the availability of additional braking capacity, given unusual brake pedal force application, is of no benefit to an operator. Anticipating "normal" braking capacity to be available, operators may not have enough distance to stop prior to impact to perceive the problem and react to it, even if they are capable of applying sufficient pedal force under controlled test conditions to generate "normal" decelerations.

As noted above, in addition to the time required for an operator to perceive the problem and react to it, there is the very real possibility that an operator will respond with a course of action which is not the most desireable. Two such courses of action are pumping the brake pedal and simply "freezing," and failing to apply a significantly higher pedal force. Such failure to respond to an unfamiliar hazard with any action at all is not uncommon.

Since many drivers have the ability to generate braking decelerations well in excess of the 0.2 g's often effected during normal driving, even with a disabled booster, the time required to perceive the problem and press more firmly on the pedal will be an important factor. Stops from low speeds will tend to be more problematic, as the operator may only anticipate requiring two seconds to stop, thus a typical perception/reaction time of 1.2 seconds will place them very near the location at which they intended to stop. The longer time anticipated by drivers to stop from higher speeds allows more opportunity to take advantage of the additional braking capacity which is available, given higher pedal application forces.

CONCLUSIONS

Depending on the vehicle and driver combination, it is possible that an operator, particularly a female, will be incapable of generating sufficient pedal force to lock the wheels, regardless of how much warning and distance they are allowed.

Is is not possible to state, without specific operator and vehicle testing, what deceleration a particular person would be capable of in a particular vehicle, when faced with a failed power assistance unit. If an operator applies a "normal" amount of brake pedal force, the braking deceleration will be about 15 to 25 % of that acheived with the booster functional.

Direct testing of the driver and the vehicle is recommended in order to make accurate estimates of maximum possible braking performance in a particular situation.

Even with direct testing, the investigator must consider the operator's actual response to the perceived problem when evaluating stopping distances and times.

ABOUT THE AUTHOR

Correspondence should be addressed to: Wade Bartlett, P.O. Box 1958, Dover, NH 03821-1958.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Ian Jones and Frederick Hochgraf for their encouragement; Richard Radlinski for his generous offers of assistance; FTI Corporation for the use of their G-Analyst and loadcell; and C.R. Eddie, J. Rolly Kinney, and Lisa DeLucia for their constructive comments and suggestions.

REFERENCES

1 Fisher, Franklin G., H. Cook, F. Oldham, R. Gloss, R. Eidemiller, "Optimized Brake Inspection Technical Volume," DOT Report No. HS-803 375, U. S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C., 1977.

2 Gillespie, Thomas D., Fundamentals of Vehicle Dynamics, Society of Automotive Engineers, ISBN 1-56091-199-9, Warrendale, Pennsylvania, 1992.

3 Bosch Automotive Handbook, 3rd Edition, Robert Bosch GmbH, ISBN 01-56091-372-X, Stuttgart, Germany, 1993.

4 Rossigno, L.P., J.J. Horsting, J.J. Carré, "Design of Lightweight Vacuum Power Brake Boosters," SAE Paper No. 820024, Society of Automotive Engineers, Warrendale, Pennsylvania, 1982.

5 Chilton's Guide to Engine Repair and Rebuilding, edited by Kerry Freeman and Richard Rivele, Chilton Book Company, ISBN 0-8019-7643-X, Radnor, Pennsylvania, 1985.

6 Baker, A.K., Vehicle Braking, Pentech Press Limited, ISBN 0-7273-2202-8, London, England, 1986.

7 1995 SAE Handbook, Society of Automotive Engineers, ISBN 1-56091-585-4, Warrendale, Pennsylvania, 1995.

8 *Highway & Vehicle Safety Report*, Volume 22, Number 8, Stamler Publishing Company, ISSN 0161-0325, Branford, Connecticut, 1995; 1996 Dodge Dakotas may have an improperly clamped booster vacuum hose.

9 *Monthly Defect Investigation Report*, U.S. Department of Transportation, National Highway Traffic Safety Administration, June, 1996; 1996 Honda Civic booster check valves were assembled with a soapy lubricant which could cause the check valve to stick.

10 U.S. Department of Transportation, National Highway Traffic Safety Administration: 49 CFR 571.105, Federal Motor Vehicle Safety Standard Number 105, Hydraulic Brake Systems.

11 Woodson, Wesley E., The Human Factors Design Handbook, McGraw-Hill Book Company, ISBN 0-07-071765-6, New York, New York, 1981.

12 Radlinski, Richard W., James I. Price, "The Brake Pedal Force Capability of Adult Females," NBS Technical Note 557, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1970.

13 Mortimer, R.G., L. Segel, H. Dugoff, J.D. Campbell, C.M. Jorgeson, R.W. Murphy, "Brake Force Requirement Study: Driver-Vehicle Braking Performance as a Function of Brake System Design Variables," Highway Safety Research Institute Report No. HuF-6a, Highway Safety Research Institute, Ann Arbor, Michigan, 1970.

14 Eaton, Dennis A., Henry J. Dittmeier II, "Braking and Steering Effort Capabilities of Drivers," SAE Paper No. 700363, Society of Automotive Engineers, Warrendale, Pennsylvania, 1970.

15 Valentine Research Inc., 10280 Alliance Road, Cincinnati, Ohio, 45242.

16 Hyung Dae Kim and Younchul Rhim, "A Study on the Characteristics of Automotive Brake Pedal Force," SAE Paper No. 923879, Society of Automotive Engineers, Warrendale, Pennsylvania, 1992.

17 Brake Pedal Effort Gage, p/n J-28662, Kent Moore Automotive Group, 29784 Little Mack, Roseville, Michigan, 48066-2298, (800)-345-2233. As of this writing, the gage costs less than $190.00.

ADDITIONAL READING

Kroemer, K.H. Eberhard, "Pedal Operation by the Seated Operator," SAE Paper No. 720004, Society of Automotive Engineers, Warrendale, Pennsylvania, 1972.

Limpert, Rudolf, Brake Design and Safety, Society of Automotive Engineers, ISBN 1-56091-261-8, Warrendale, Pennsylvania, 1992.

Automotive Brake Systems, 1st Edition, Robert Bosch GmbH, ISBN 01-56091-708-3, Stuttgart, Germany, 1995.

SAE #970946
Reviewer's Discussion
by C.R. Eddie, C.R. Eddie Engineering Inc.
Passenger Vehicle Braking Performance With a Failed Power Booster
Wade Bartlett, Author

Bartlett assesses the physical force requirements for various decelerations for a specific but typical late model vehicle with and without the vacuum power booster operating. He compares this with the range of ability for male and female drivers, based on publications by others. He notes that for a specific incident it is not possible to know the braking effort a particular person might be able a generate if the power assist unit failed. Depending on the vehicle and driver combination he finds that an operator might be incapable of generating sufficient pedal force to lock the wheels. This paper provides a straight forward and succinct presentation which will be a valuable starting point for anyone undertaking tests and analyses to reconstruct the effects of a brake booster failure.

SAE # 970946
Reviewer's Discussion
by J. Rolly Kinney
Passenger Vehicle Braking Performance with a Failed Vacuum Power Booster
Wade Bartlett, Author

The paper suggests a significant portion of the driver population probably cannot activate the maximum vehicle braking capability when driving passenger cars with un-boosted or disabled booster systems. Those drivers probably achieve a maximum performance level similar to heavy trucks and inexperienced motorcyclists, so they are not exceptional in the traffic mix but are mismatched to the likely performance of passengers cars. When a booster is abruptly disabled a more serious mismatch probably occurs between the driver's expectations and the car response.

The author suggests some accidents may occur because drivers believe a disabled booster means the brake system failed and do not attempt to apply a larger foot force. A similar driver reaction sometimes occurs when the brake pedal stroke increases because one part of the dual brake system fails. Driver failure to use the remaining capability in both type incidents probably is best moderated by improved driver training practices that include actual operation with the booster disabled and with one part of the dual system disabled.

970947

SMAC-97 - Refinement of the Collision Algorithm

Brian G. McHenry and Raymond R. McHenry
McHenry Consultants, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The Simulation Model of Automobile Collisions (SMAC) computer program, developed in the early 1970's, includes a complex collision algorithm for monitoring, detecting and modeling the collision interactions of motor vehicles. A detailed review of some aspects of the logic, rationale and, in particular, limitations of the original SMAC collision algorithm is presented.

This paper presents refinements in the definition of the collision interface, the definition of collision type, the vehicle proximity and collision detection logic, and the form of supplementary impulsive constraints on relative motions.

The effects of the modifications of the SMAC algorithm on reconstruction results are presented in the form of direct comparisons of results obtained with the original and modified algorithms.

INTRODUCTION

The SMAC program was initially developed in the 1970's when all development of computer code was performed on time-share mainframe computer systems. The capabilities of computers at that time were limited by the maximum amount of available memory (e.g., limit on program size) and users were charged for computer use based on memory and CPU utilization. The costs associated with the development and execution of the SMAC program were relatively high (e.g., [1][1] ,circa 1971,p 48, "The range of costs,...,has been approximately $25.00 per application run" for the SMAC program).

These limitations during the original development of the SMAC program guided the selection of many of the simplifying assumptions of the mathematical model.

Since the early 80's and particularly by the mid 1990's, the prevalence of powerful mini-computers and more recently extremely powerful and inexpensive Pentium PC's, creates an availability of virtually unlimited and inexpensive computer resources. This has inspired a detailed re-evaluation and refinement of computer codes, particularly those developed in the 1970's. The general approach to the reported refinements of the SMAC computer program has been to reconsider the initial simplifying assumptions based both on the availability of additional full-scale test results and the virtually unlimited computer resources.

The reported research identifies and discusses artifacts and/or shortcomings of the original SMAC and the EDSMAC computer programs that have been encountered by the authors in relation to applications (SMAC) and evaluation of applications (EDSMAC) to actual accident cases.

It should be noted that any references to the original SMAC [2] computer program are also generally applicable to the EDSMAC [3] computer program. The original SMAC program and the widely distributed EDSMAC clone are essentially identical. No analytical refinements have been made by the distributors of EDSMAC which produce any significant changes in the results.

BACKGROUND

In the early 1970's, NHTSA sponsored a research project to develop a computer program that would achieve improved uniformity, as well as improvements in accuracy and detail, in the interpretation of physical evidence in highway accidents. The resulting prototype computer program was the Simulation Model of Automobile Collisions (SMAC) [1,2,4,5]. At the completion of the NHTSA sponsored research at Calspan in 1974, a preliminary version of the SMAC program was delivered to the NHTSA and it has subsequently been distributed as the NHTSA SMAC computer program.

Subsequent follow-up contracts for research and development of the SMAC program sponsored by NHTSA went to other organizations [6,7,8]. Further research and development on the SMAC program was also continued independently at Calspan [9] and additional corporate-sponsored research to support criticism of the SMAC program [10,11] was also performed.

[1] Numbers in brackets [] indicate references at end of paper

There were no significant changes by NHTSA into the 1974 NHTSA SMAC at the completion of the NHTSA follow-up contracts.

In 1986, Day and Hargens created EDSMAC[12], a PC version of the 1974 NHTSA SMAC program converted to the BASIC programming language. Subsequent reports related to the EDSMAC program [3,13,14] reveal that except for very minor modifications, the EDSMAC program is essentially the same as the original 1974 NHTSA SMAC program. Related development efforts by the distributors of the EDSMAC program have been directed towards a mini-computer based high-end graphics environment [14,15,16,17,18].

In 1988, a number of suggestions for further refinement and extensions of the SMAC program were presented [19]. In 1989, some suggestions for avoiding misapplication of computer programs, including the EDSMAC program [20] were presented.

The widespread distribution of the EDSMAC program has dramatically increased its utilization for the reconstruction of individual accidents (e.g., accidents that are involved in litigation). This creates a situation where in many instances, either through misuse, misapplication or due to shortcomings in the original NHTSA SMAC (and therefore EDSMAC), there have been applications of the program which include significant effects of artifacts of the original programming logic.

The current reported research defines some important refinements of the SMAC program, particularly with respect to the collision modeling algorithm. This paper also extends the suggestions in [19] and [20] to assist users in avoiding possible misapplication of the SMAC program. Note that any references to the "original" SMAC program refer to the 1974 NHTSA SMAC program and therefore to the EDSMAC program.

This paper constitutes a continuation of the research presented in [19] and more recently in [21]. Specific refinements to the original SMAC program collision modeling routines discussed herein are as follows:

1. Definition of the collision interface.
2. Collision type specification.
3. Supplementary impulsive constraints on relative motion.
4. Vehicle proximity and collision detection logic.

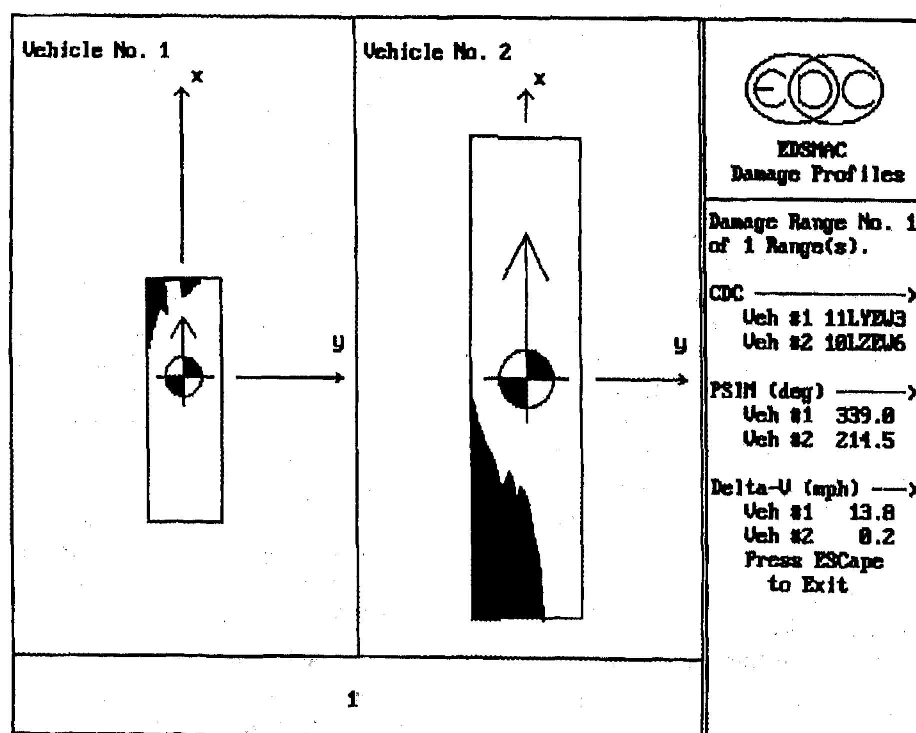


Figure 1 EDSMAC damage display for bus/auto impact simulation with 90 degree vector artifact demonstrated [22]

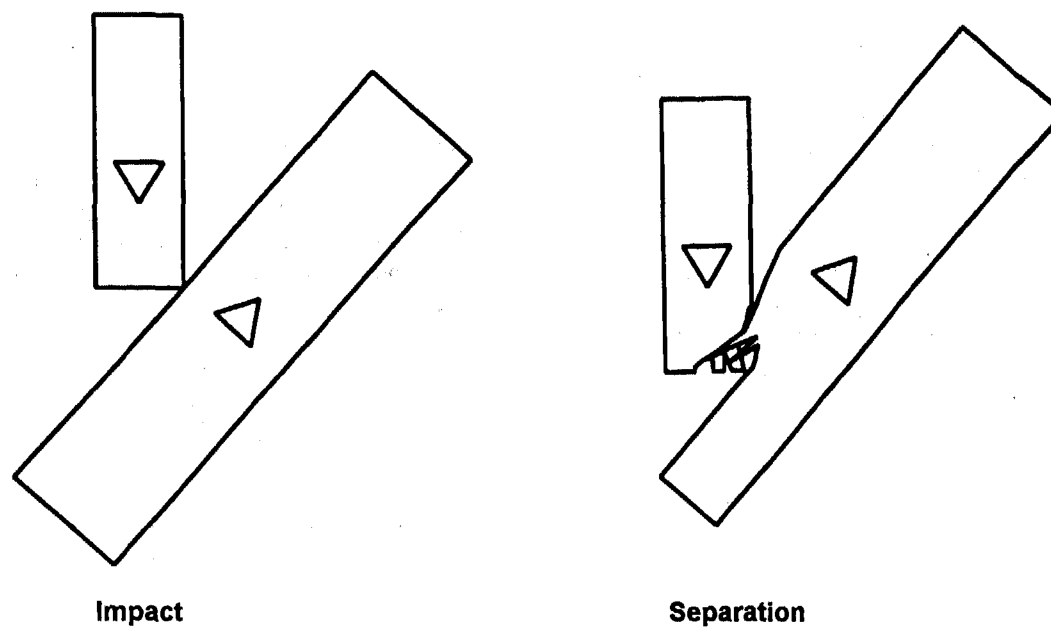


Figure 2 Original SMAC(EDSMAC) bus/auto impact simulation at 0.140 sec after initial impact with 90 degree vector artifact demonstrated.

DEFINITION OF THE COLLISION INTERFACE

Problems encountered in applications of the original version of the SMAC (and EDSMAC) programs in oblique impacts particularly when a long vehicle is struck on the side at a position away from the center of gravity (e.g., **Figure 1** and **Figure 2**) originate in limitations imposed by the selected original form of analytical definition of the vehicle peripheries.

In particular, radial vectors with origin at the center of gravity are used to define the peripheries of the two vehicles (e.g., **Figure 3**). During a collision, the lengths of the radial vectors in contacted regions are adjusted by an iterative procedure which seeks an equilibrium dynamic pressure on the collision partners. The dynamic pressures at individual peripheral points are assumed to be proportional to their displaced distances from the undeformed periphery.

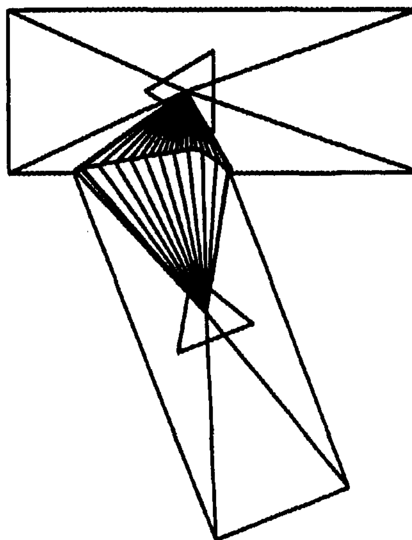


Figure 3 Mathematical representation of the SMAC periphery by radial vectors.

In the cited problem configuration, some of the radial vectors of the two interacting vehicles approach the condition of being perpendicular to each other and, as a consequence, the iterative adjustment procedure for the lengths of radial vectors cannot find a proper solution. The inter-vehicle pressure associated with these vector points cannot find an equilibrium. As a result, an unrealistic equilibrium interface is produced that includes a series of jagged notches (e.g., **Figure 1** and **Figure 2**). A close-up illustration of the phenomenon is included in **Figure 4** and a close up detail including display of the radial vectors is included in **Figure 5**. The error flag normally associated with the problem in the SMAC program is "PRESI (or PRESIJ) tends negative". In the original program the error flag sets the associated pressure to 0.0 and attempts to continue program execution. The setting of the zero pressure creates the jagged interface wherein only some of the vectors are "active". Other vectors have the error and, therefore, they are undeformed or partially deformed from the periphery. The original SMAC included a program stop if a sufficient number of PRESI/PRESJ type errors occurred in an individual simulation run.

The perpendicular vector problem is most obvious in the original SMAC with longer vehicles as a result of the large distance of the contact region from the origin. Individual vectors on the long vehicle are required to define a larger area of the vehicle periphery than that in a more typical impact configuration. This tends to amplify the near perpendicular vector problem where one radial vector on the long vehicle may affect several on the other vehicle.

An initial approach that was investigated for a generalized fix to the phenomenon was an increase in the number of vectors and, thereby, the detail included in the collision routine. In the original SMAC program the capabilities of the collision routine were limited due to computer memory storage considerations. The specific limitations were the maximum number of interacting vectors (100) and the maximum number of iterations the program could utilize to attempt to find equilibrium in inter-vehicle pressure (100). In consideration of the capabilities and capacities of modern day (1990's) computers, the capabilities have been increased to permit the use of 1 degree vectors and smaller values for the acceptable error in inter-vehicle pressure. Users of SMAC should never make use of an angle greater than 3 degrees between radial vectors. The use of 1 degree increments reduces the problems associated with near 90 degree vectors by decreasing the area of coverage of individual vectors for longer vehicles. It also provides a generalized increase in the detailed equilibrium interface of the collision routine for normal simulations. However, the described changes did not solve the problem with perpendicular vectors.

The indicated problem is resolved in the revised SMAC program by means of locating the origin of the radial vectors of the collision interface at a position on the vehicle other than the center of gravity. In this manner, the origin of the vectors defining the vehicle periphery can be moved longitudinally to be near the location of the collision contact (e.g., see **Figure 6** and **Figure 7**) and/or moved laterally away from the initial contact location (e.g., see **Figure 8** and **Figure 9**).

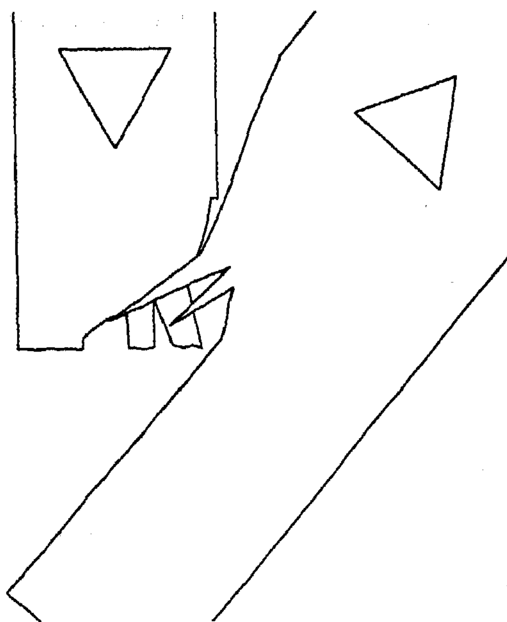


Figure 4 Close-up detail of NHTSA SMAC (EDSMAC) collision interface problem of perpendicular radial vectors (0.14 sec after initial impact)

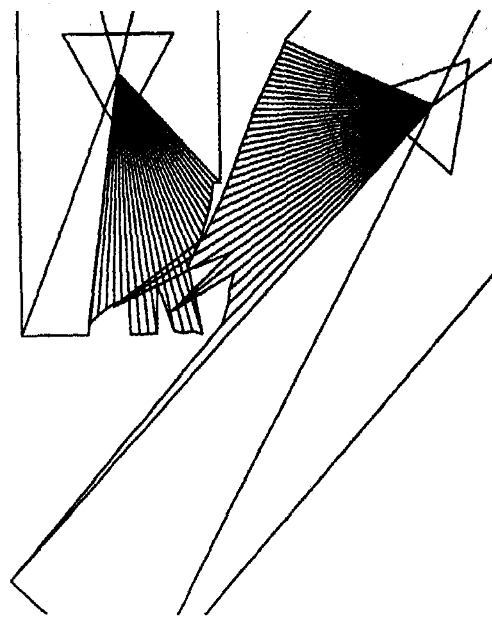
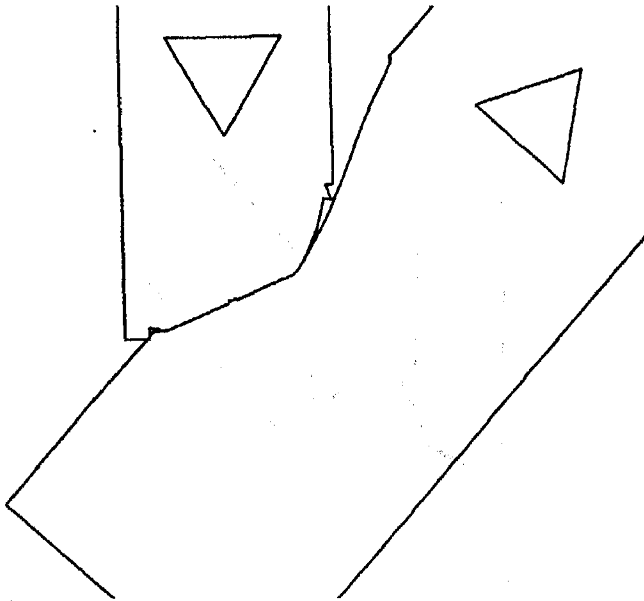


Figure 5 Close-up detail including radial vectors



'igure 6 Close-up detail of variation of the X location of the center of collision interface origin to avoid perpendicular vectors (0.14 sec after initial impact)

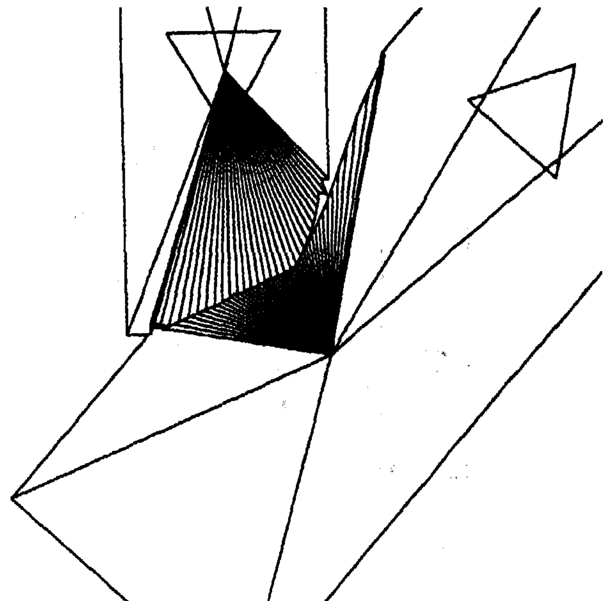


Figure 7 Close-up detail, including radial vectors

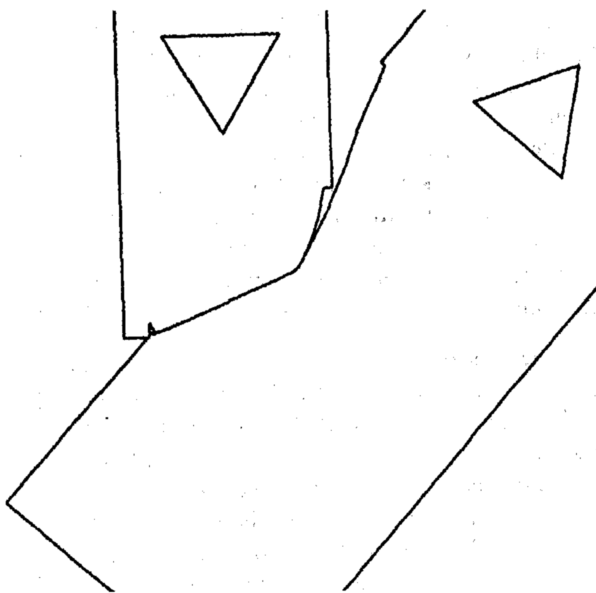


Figure 8 Close-up detail of variation of the X and Y :ation of the center of collision interface origin to avoid perpendicular vectors (0.140 sec after initial impact)

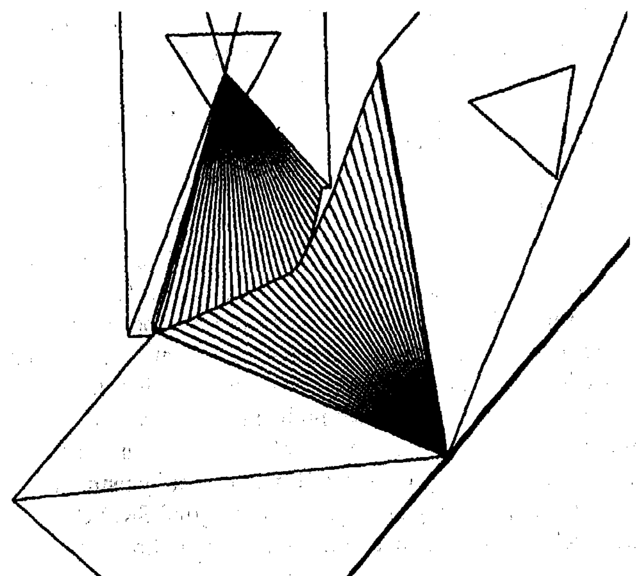


Figure 9 Close-up detail, including radial vectors

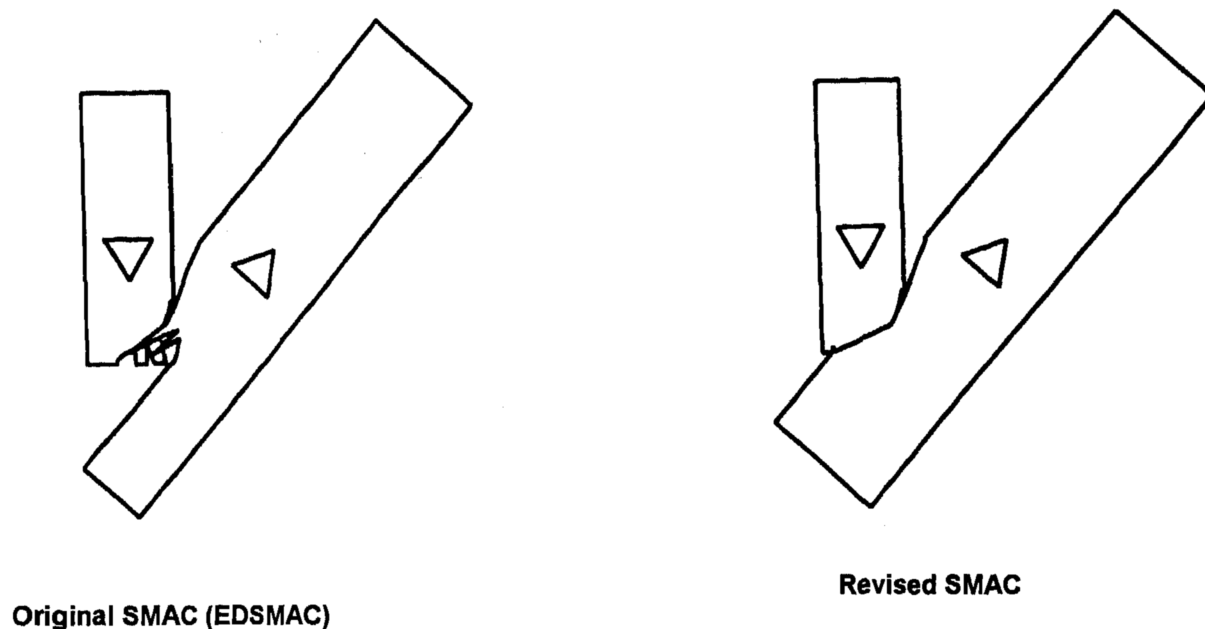


Figure 10 Comparison of predicted damage of original SMAC (EDSMAC) and revised SMAC (0.14 sec after initial impact)

A comparison of the effects of moving the origin of the collision interface from the CG is displayed in **Figure 10**. The results of the car/bus impact presented in **Figure 1** also indicate an incorrect value for the impact speed-change of the automobile. The EDSMAC program indicated an impact speed-change of 13.8 MPH for the automobile, whereas the correct value is 31.5 MPH (The impact velocities were 43 mph for the automobile and 15 mph for the bus).

The long vehicle example dramatically demonstrates problems which can occur with perpendicular vectors. Users of the original SMAC (EDSMAC) should also be aware that to a lesser degree, dependent on the impact configuration and severity level, normal length vehicles may occasionally encounter perpendicular vector problems. The problem is dramatically demonstrated on a smaller scale when users attempt to simulate a pole or small tree impact by using a small rectangular or square object. The original SMAC (EDSMAC) program is not valid for the simulation of pole and/or small tree impacts. The fundamental problem of pole impact simulations with the original SMAC(EDSMAC) (i.e., non-homogeneous crush properties for small contact areas) is increased by the perpendicular vector problem when the pole and/or small tree impact location is offset from the vehicle center of gravity. A separate research report will include discussion of revisions to SMAC to accommodate pole and/or small tree impacts[23]

The original form of SMAC can also produce application problems in reconstructions in which the equilibrium interface approaches the center of gravity of one of the vehicles. Therefore, lateral movement of the origin of the collision interface vectors may be required for simulations of severe side impacts where the collision interface can approach and/or exceed the ½ width of the vehicle. In the original SMAC (and EDSMAC) program these types of collisions normally produce either a program error or an effective softening of the crush characteristics (e.g., [11]).

The present discussion should not be interpreted as an indication that valid results can be obtained at very high severity levels. Rather, the related modifications of SMAC are aimed at minimizing the effects of artifacts, in the computer implementation of the analytical relationships, on the predicted results. Evaluations of the ranges of validity of a mathematical technique should not, of course, be obscured by the presence of significant effects of artifacts.

Research is underway to establish a generalized fix to automatically locate the origin of the collision interface vectors on the basis of impact configuration and severity. An automatic adjustment of the origin is considered to be highly desirable from the viewpoint of eliminating any possibility of related deviations from uniform interpretations of evidence. In the existing revised version of the SMAC program an option is presently included to permit the user to move the origin, where appropriate.

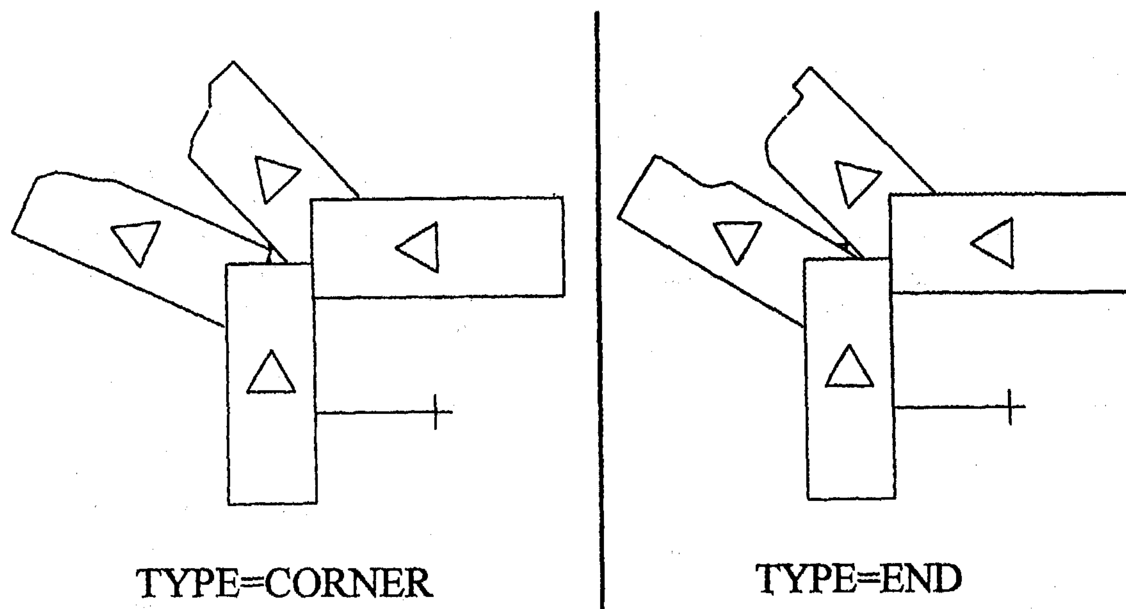


Figure 11 Demonstration of effects of 1" change in Y of vehicle produces different definition for type of impact configuration and therefore different damage profiles.

COLLISION TYPE SPECIFICATION

In the original SMAC (EDSMAC) program, the collision interface handles three types of impact configuration: End, Side, and Corner. Logic in the program calculates the extent of collision contact overlap at the instant of contact and determines which of the three types of configuration logic to use for calculation of the collision forces. Occasionally the program may exhibit a disproportionate sensitivity to minor changes in a particular impact configuration which is near a "logic" transition point. Minor changes in the impact configuration can cause the program to switch between the different types of collision force logic (e.g., Figure 11). Research is underway to update the logic associated with determination of the impact type as part of a general fix of the collision logic to avoid this phenomenon. To avoid sensitivity problems a manual override of the automatic logic is suggested and has been provided in the revised SMAC program. The manual override of the impact configuration logic eliminates the sensitivity and provides for a manual check of the effects of changing the designated impact configuration type.

SUPPLEMENTARY IMPULSIVE CONSTRAINTS ON RELATIVE MOTION

There are some impact configurations for which the simple combination of compressive forces and coulomb friction of the original SMAC program are inadequate. The actual structural interactions between colliding vehicles can include significant tensile forces and/or moment constraints on relative rotation in addition to the primary compressive interaction. Also, significant alternative load paths can occur which do not produce sheet metal crush. In such cases, supplementary impulsive constraints on relative motions can be applied without detracting from the principles of conservation of linear and angular momentum in relation to trajectory analyses. In other words, such impulsive constraints on relative motions of the two vehicles continue to conserve the linear and angular momentum of the two-body system with regard to interactions between the two vehicles. Any benefits derived in the form of improvements in the match of the position and heading at rest of vehicle#1 will have equal and opposite effects on the predicted position and heading at rest of vehicle#2. If both simulated vehicles can be brought into acceptable matches of their positions and headings at rest by means of such supplemental impulsive constraints on relative motions, the required pre-impact linear and angular momentum of the two vehicle system will be unaffected and the approximation of initial speeds and impact speed-changes can proceed as usual.

It should be noted that supplementary impulsive constraints on relative motions can have the effect both in reconstructed accidents and in actual collision evidence of disturbing the relationship between sheet metal crush and the corresponding ΔV. For example, in a small-overlap offset frontal collision, interlocking front wheels can support large interaction forces with only a limited extent of sheet metal crush. Also, a pocketing sideswipe can include a major load path (e.g., wheel/axle/suspension) that does not produce corresponding crushing of sheet metal.

It is therefore necessary for the reconstructionist to evaluate the role of alternative load paths when interpreting sheet metal crush. In those cases where alternative load paths play a major part in the vehicle interactions it is necessary to place greater emphasis on scene evidence as opposed to sheet metal crush and, in the case of SMAC applications, to make use of supplementary impulsive constraints.

To work around the indicated form of shortcoming of the original SMAC program some of the RICSAC reconstructions performed by Jones [24] and some users of EDSMAC made use of a value of the inter-vehicle friction coefficient substantially greater than 1.0 attempting to compensate for the lack of tensile forces and/or moment constraints on relative rotation. The problem with the high friction approach is two-fold.

First, the inter-vehicle friction coefficient of the SMAC collision model is based on coulomb friction which is independent of the sliding velocity and has a coefficient value less than or equal to 1.0. During the collision, as relative sliding of the surfaces occurs, the magnitude of the inter-vehicle friction force depends on the existence and magnitude of the collision normal force. Supplemental tensile forces and/or moment constraints occur in vehicles as the vehicles begin the separation phase when the normal force may be very small or zero. After the primary collision, as a result of the impact configuration, the vehicle contact area and/or the vehicle component contacted, intermeshed components of the two vehicles can offer additional resistance to separation in the absence of the normal forces required for coulomb friction. Using a value for the inter-vehicle friction coefficient greater than the recommended range (normal value 0.55, recommended range 0.3-1.0) may unrealistically and/or adversely affect the primary impact phase, since the additional forces and moments produced by the elevated friction can act to redirect the forces and moments during the primary impact.

A secondary problem which may occur with original SMAC (EDSMAC) use of a high value of inter-vehicle friction can occur in the post-processing program which may not be able to "match" the accelerations to the damage region. As a result invalid ΔV values and clock direction may be reported. An EDSMAC application summary page (used by an expert witness as an exhibit to deposition testimony) which contains invalid ΔV values for the collision partners was produced by the use of a value for the inter-vehicle friction coefficient of 2.0 and is shown in **Figure 12.** (Note that $m_1\Delta V_1 \neq m_2\Delta V_2$. In the original SMAC a supplementary page included diagnostics which report all of the speed changes calculated in a given run to permit a check of the values reported by the post-processing routine. This auxiliary page appears to have been omitted in the EDSMAC program.).

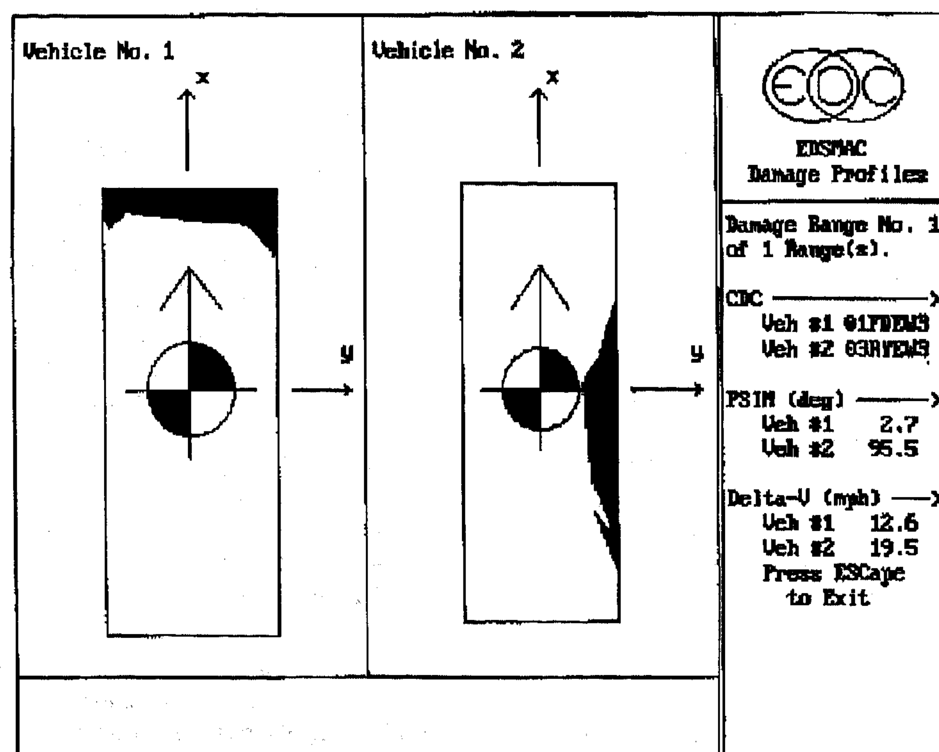


Figure 12 Example usage of invalid inter-vehicle friction coefficient of 2.0 which produces invalid results (Weight1=3226 lb., Weight2=3953 lb.) Therefore $m_1\Delta V_1 \neq m_2\Delta V_2$.)

Revisions to the SMAC program to model supplementary impulsive constraints include the impulsive-constraint "SNAG" option as previously presented in [19]. In **Figure 13** the original SMAC program does not adequately simulate an impact on the LR wheel of vehicle #1. The impact on the LR wheel in the investigated accident produced a 180 CCW rotation of vehicle #2 due to a momentary snagging of the LR wheel of vehicle #1 on the front of vehicle #2. In **Figure 14** two different impulsive constraints ("SNAG") are applied to the vehicles in the area of the left rear wheel of vehicle #1. The results indicate that applying a 1000 lb.-sec impulse is adequate to spin vehicle #2 180 degrees and, thereby, it constitutes a reasonable approximation of the magnitude of the actual snagging of the vehicle structures.

Efforts are being directed toward the development of automatic calls for impulsive constraints and for standardized related input values on the basis of the impact configuration and the closing speed. Prior to completion of that development the timing and the magnitudes of any impulsive constraints should be reported along with corresponding reconstruction results.

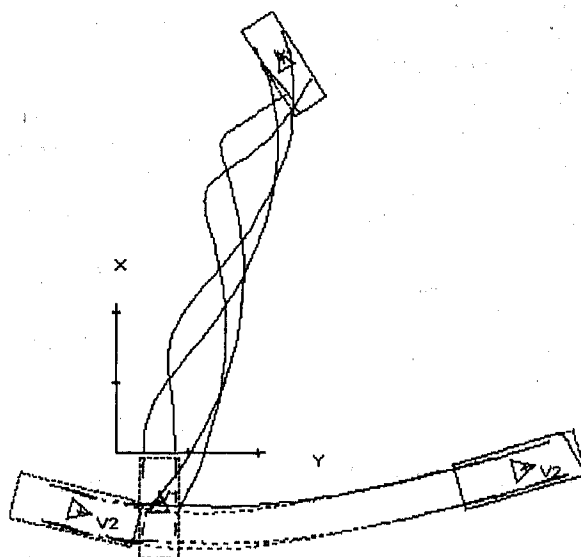


Figure 13 Original SMAC Simulated Impact on LR wheel of Vehicle No. 2

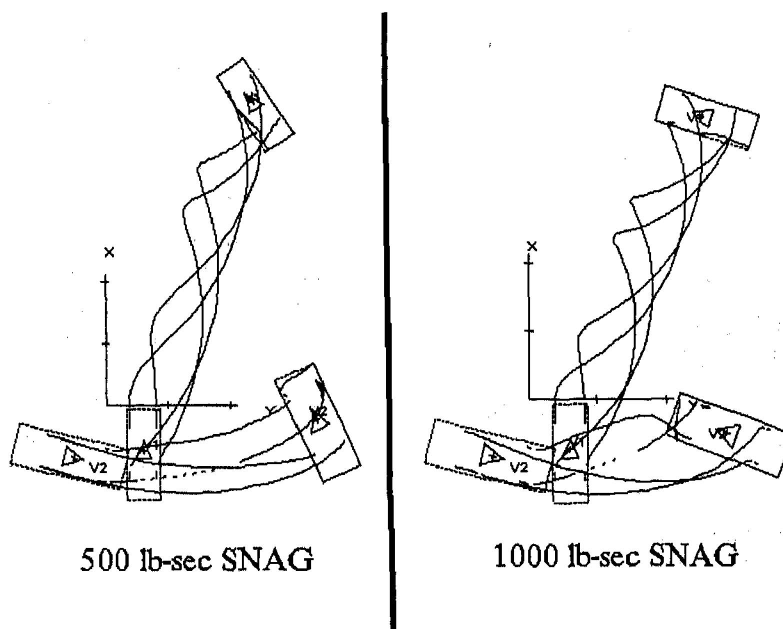


Figure 14 Revised SMAC simulated impact with 500 lb-sec and 1000 lb-sec impulsive constraint applied

VEHICLE PROXIMITY AND COLLISION DETECTION LOGIC

In the SMAC program, each vehicle is represented by a rectangular box with the length and width dimensions of the simulated vehicle. Collision detection is accomplished by continually checking the corners of each vehicle "box" to determine if it is within the periphery of the other vehicle "box". Once a corner point is found to be in contact, the program begins calling the collision routine to use the collision model radial vectors to scan for interference and contacts and to calculate the associated collision forces. The program also changes the integration time increment to the user specified collision integration interval (DTCOLL, normally 0.001 seconds). The end of the collision event is assumed when a fixed number of time increments have passed wherein the accelerations for each vehicle is below 1 g-unit. The program flags the system into a separation mode and utilizes the separation time increment (DTCOLT, normally 0.005 seconds). In the original form of the SMAC program, the separation time increment was utilized for a fixed number of increments (originally 100). It was the intent of the separation time increment to catch side-slaps, etc. while the program was still in a relatively small time increment. If the fixed number of DTCOLT time increments are passed without any accelerations greater than 1 g-unit then the program shifts to the trajectory time increment (DTTRAJ, normally 0.01 seconds) for the remainder of the simulation run.

Problems have been found with applications of the EDSMAC program where the program missed the accelerations associated with "side-slap" impacts. The reasons for the problems with the EDSMAC program have been found to be mainly due to the choice of time increment size.

The recommended time increment sizes for use of the SMAC (or EDSMAC) program are as follows:

DTTRAJ, Trajectory time increment, not to exceed 0.01 seconds.

DTCOLT, Separation time increment, not to exceed 0.005 seconds

DTCOLL, Collision time increment, not to exceed 0.001 seconds.

These time increments should be assumed to be absolute maximums. With the speed of modern day Pentium computers, a complete SMAC ten second simulation run of an impact and spinout can be performed in less than 10 seconds of real time with all time increments set to 0.001 seconds. EDSMAC documents recommend the use of 0.01 for the DTCOLT separation time increment, 0.05 for the DTTRAJ trajectory time increment, 0.001 for the DTCOLL collision time increment **[25]**. These suggested times must have been based on suggested times for mainframe applications from 1974 contained in **[26]**.

A problem with larger time-increments is the fact that they may reduce the accuracy of the predicted results. Also, when simulating a collision which may include a side-slap secondary collision, the program may miss the forces and moments associated with the side-slap event. After a primary collision event, once the SMAC (EDSMAC) program has changed to the EDSMAC recommended large DTCOLT or DTTRAJ time increments, the occurrence of a sideslap may not be detected or detected so late that the sideslap event is over and no associated accelerations are developed in the simulation due to the sideslap. The logic in the original SMAC (EDSMAC) associated with sideslap detection does not return to the DTCOLL small time-increment until the acceleration on either of the vehicles exceed 1 g-unit.

The following illustrates what occurred in an application of the EDSMAC program (used by an expert witness as an exhibit to deposition testimony). The impact configuration is displayed in **Figure 15** and the EDSMAC output damage page is contained in **Figure 16.** The damage display contains sideslap "damage" for which the accelerations associated with the sideslap "damage" were not simulated. The time history of the acceleration which 'misses' the side-slap is contained in **Figure 17** while the time-history of the corrected simulation is contained **Figure 18.** The differences which may occur between the incorrect simulation (which missed the accelerations associated with the sideslap) and the corrected simulation (which included the sideslap) are significant differences in the total amount of rotation and direction of travel to rest.

Revisions to the SMAC program have been implemented to make the post-impact interval where the program stays in DTCOLT a fixed *duration* of 0.30 seconds rather than a fixed number of time steps. In this manner, when the user sets the DTCOLT increment to 0.001 the program will continue to scan for greater than the 100 time steps . It is recommended that in accidents which include a sideslap collision that users use a DTCOLT of 0.001 seconds.

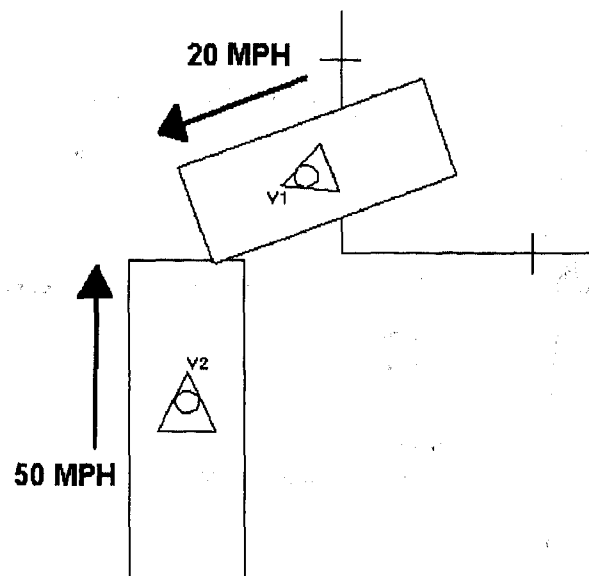


Figure 15 Impact Configuration which produces side-slap

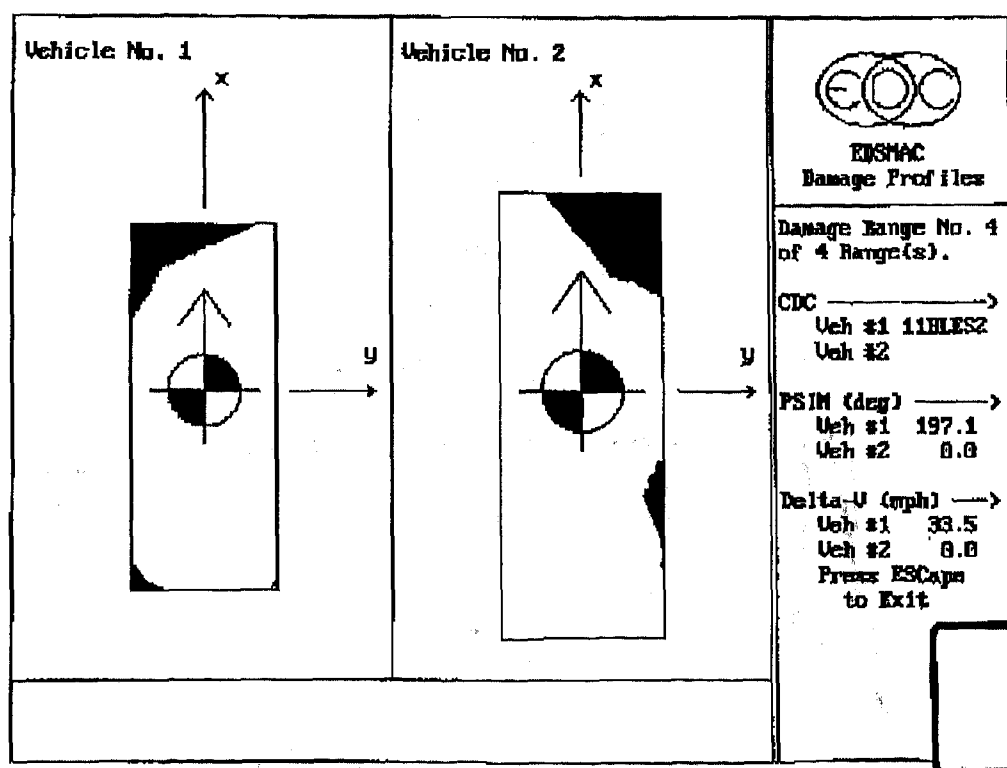


Figure 16 Example instance where EDSMAC program misses a side-slap event (no accelerations)

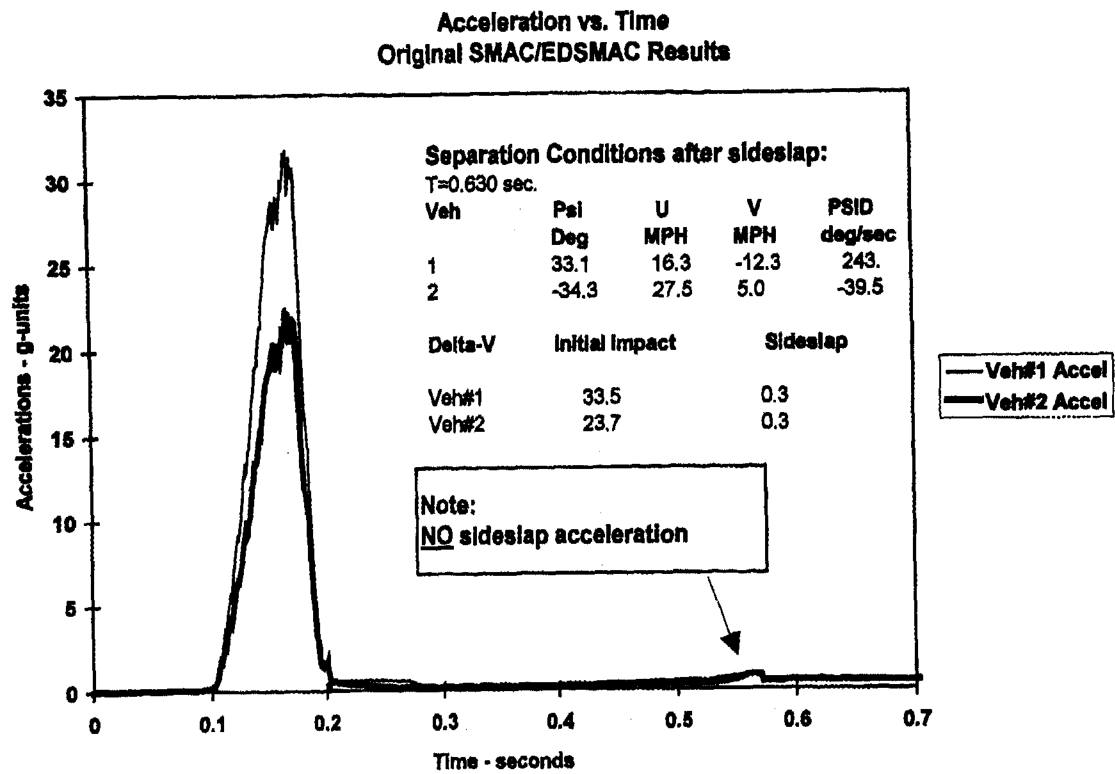


Figure 17 Acceleration time-history of example EDSMAC run which misses the side-slap

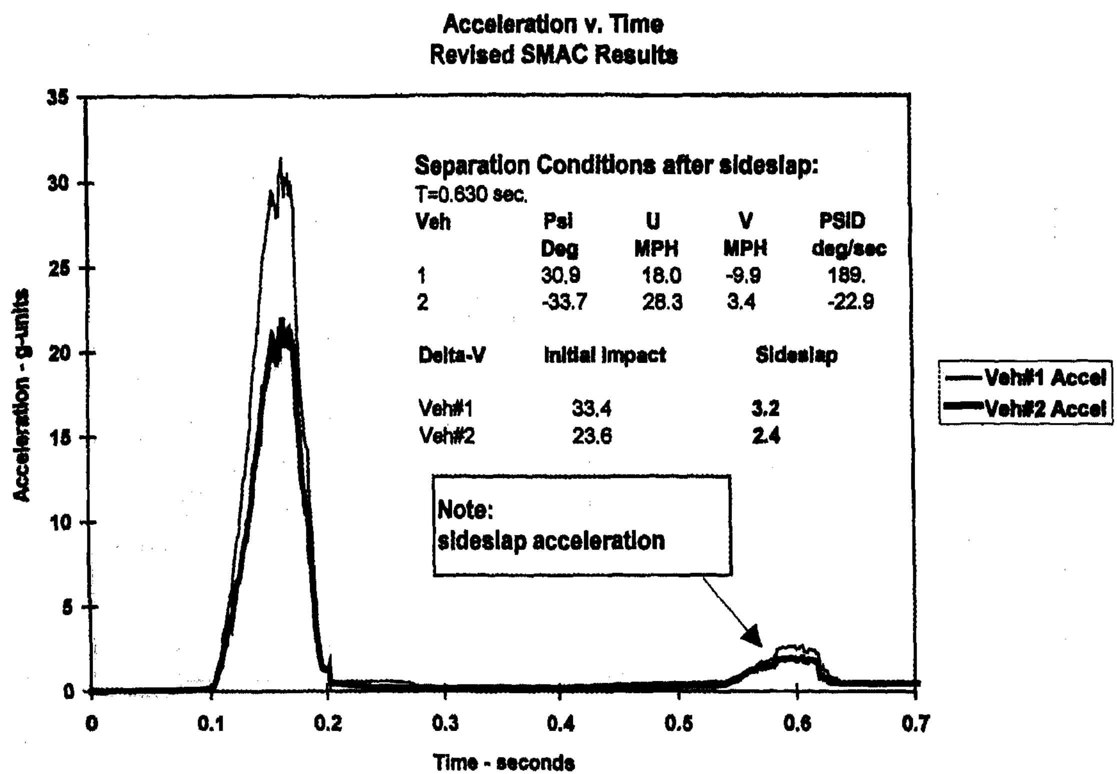


Figure 18 Acceleration time-history of example EDSMAC run on revised SMAC program which includes simulation of the side-slap impact

SUMMARY

With the rapidly changing capabilities and capacities of modern day computers, computer programs in general and accident reconstruction computer programs in particular require continuing efforts to check and refine results while critically evaluating the underlying simplifying assumptions. This paper has presented some areas of needed refinement in the original SMAC collision algorithm.

REFERENCES

1. McHenry, R. R., "Development of a Computer Program to Aid the Investigation of Highway Accidents", Contract FH-11-7526, NTIS PB 208537, Calspan Report VJ-2979-V-1, December 1971
2. McHenry, R.R.,"A Computer Program for Reconstruction of Highway Accidents", SAE Paper 73-0980, Proceedings of the 17th Stapp Car Conference, November 1973
3. Day, T. D., Hargens, R. L., "An Overview of the Way EDSMAC Computes Delta-V", SAE Paper 88-0069
4. McHenry, R. R., Segal, D. J., Lynch, J. P., Henderson, P.M., "Mathematical Reconstruction of Highway Accidents",Contract DOT-HS-053-1-146,NTIS PB 220150, Calspan Report ZM-5096-V-1, January 1973
5. McHenry, R. R., Jones, I. S., Lynch, J. P., "Mathematical Reconstruction of Highway Accidents - Scene Measurement and Data Processing System", Calspan Corporation, Contract DOT-HS-053-3-658, Calspan Report ZQ-5341-V-2, December 1974
6. Chi, M., Neal, E., Tucker, J. R., "Revision of Simulation Model of Automobile Collisions (SMAC) Computer Program: Investigation of New Integration Algorithm", Contract DOT-HS-7-01545, Report No. DOT-HS-803 294
7. James, M. E., Ross, H. E., Whittington, C., "Improvement of Accident Simulation Model and Improvement of Narrow Object Accident Reconstruction", April 1978, Contract DOT-HS-5-01262, DOT-HS-7-01656, Report DOT-HS-803 620
8. Moffatt, C.A., Byrd, J.,"A Computer Model to Operate the SMAC Program Automatically", 1980, Contract DOT-HS-8-01820, in Highway Collision Reconstruction, ASME winter meeting, 1980, Library of Congress 80-69198
9. McHenry, R. R.,"Computer Aids for Accident Investigation", SAE Paper 76-0776
10. Warner, C.Y., Perl, T. R.,"The Accuracy and Usefulness of SMAC", SAE Paper 78-0902
11. Perl, T.R., Anderson, D. O., Warner, C.Y., "Improvements to the SMAC Program", SAE Paper 83-0610
12. Engineering Dynamics Corporation, "Vehicle Analysis Package - EDSMAC Program Manual, Version 2", Lake Oswego, OR 1986
13. Day, T. D., Hargens, R.L., "Further Validation of EDSMAC using the RICSAC Staged Collisions", SAE Paper 90-0102
14. Day, T. D., Siddall, D. E., "Validation of Several Reconstruction and Simulation Models in the HVE Scientific Visualization Environment", SAE Paper 96-0891
15. Day, T. D., "The Scientific Visualization of Motor Vehicle Accidents", SAE 94-0922
16. Day, T. D., "An Overview of the HVE Developer's Toolkit", SAE Paper 940923
17. Day, T. D., "An Overview of the HVE Vehicle Model", SAE Paper 95-0308
18. Day, T. D., "An Overview of the HVE Human Model", SAE Paper 95-0659
19. McHenry, B.G., McHenry, R. R., "SMAC-87", SAE Paper 88-0227
20. Day, T. D., Hargens, R.L., "Application and Misapplication of Computer Programs for Accident Reconstruction", SAE Paper 89-0738
21. McHenry, R. R., McHenry, B.G., "Effects of Restitution in the Application of Crush Coefficients", SAE Paper No. 97-0960, To be published at the 1997 SAE International Congress and Exposition.
22. The inputs for the various examples demonstrated in this paper are available upon request from the authors. E-mail requests to: mchenry@interpath.com
23. Research completed on adding a pole impact to the SMAC program will be published in the near future.
24. Jones, I. A., Baum, A. S., "Further Validation of the Simulation Model of Automobile Collision Program", SAE Paper 79-1025, 23rd Stapp Conference Proceedings
25. Engineering Dynamics Corporation, "EDSMAC Users Manual, version 1.0", Lake Oswego, OR
26. Solomon, P. L., "The Simulation Model of Automobile Collisions (SMAC) Operators Manual", NHTSA, October, 1974

CONTACT INFORMATION

Questions or comments on the paper are welcomed and can be addressed to the authors by:

e-mail: **mchenry@interpath.com**

Postal Service Mail: **103 Brady Court, Suite 200**
Cary, NC 27511 USA

WWW: **http://www.mchenrysoftware.com**

SAE #970947
Reviewer's Discussion
by Gary W. Cooper
SMAC-97 Refinement of the collision Algorithm
Brian McHenry, Raymond R. McHenry, Authors

The authors describe some of the limitations of the original SMAC algorithm. The paper focuses on limitations in four collision modeling routines in the SMAC program. The authors discuss the reasons that limitations existed in the original program, how the limitations manifest themselves in simulations, and then describe programming "fixes" for these limitations. The authors point out there have been no significant changes in the original 1974 NHTSA SMAC algorithm and many of the limitations (simplifying assumptions) were directly related to the cost and capabilities of the computers used at the time the program was developed. However, with the powerful PC's that now exist the simplifying assumptions are no longer required and the program should be refined. The authors do an excellent job of describing the limitations, how the limitations impact a simulation, and the way to resolve the limitation through improving the algorithm. Everyone should agree that any refinements that better model the "real world" should be incorporated in programs to enhance mathematical models used to simulate collisions.

SAE #970947
Reviewer's Discussion
by Wesley Grimes
SMAC-97 Refinement of the Collision Algorithm
Brian G McHenry, Raymond R. McHenry, Authors

The authors have done a good job in summarizing the history of the SMAC algorithm and have pointed out many of the simplifying assumptions that are a part of the original SMAC program. It seems that the most significant refinement presented is in positioning the origin of the radial vectors used to model the impact and deformation of the vehicles. The refinement presented here is a logical step in improving the behavior of the SMAC algorithm, but the only clear advantage presented is in collisions where the damage on one of the vehicles is a long distance from the center of mass. It seems that the refinement will also aid in all the collisions studied, but a validation study must be completed before these advantages are quantified. The other three refinements are also of interest and should be studied carefully by analysis. Overall this paper is summary of refinements that are possible. Intuitively, the refinements make sense and should improve the overall behavior of the SMAC algorithm. A detailed validation study needs to be completed so the overall improvements can be presented.

970949

CRASH-97 - Refinement of the Trajectory Solution Procedure

Brian G. McHenry and Raymond R. McHenry
McHenry Consultants, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The *trajectory solution procedures* of the original CRASH program included both the SPIN routine and an exploratory *trajectory simulation option* to approximate and refine the linear and angular velocities at separation. The resulting separation speeds were then used to determine the impact speeds by means of application of the principle of conservation of linear momentum.

This paper presents a detailed review of the logic, rationale and limitations of the trajectory solution procedures of the original CRASH program and discusses a number of refinements including: incorporation of the principle of conservation of angular momentum, approximations of the effects of changes during collision in the positions and orientations of the two vehicles and of the effects of external forces and moments that act on the two-body system during the collision, and adaptations of optimization techniques for error reduction and convergence in iterative solutions.

The overall effects of the refinements to the CRASH trajectory algorithm on reconstruction results are illustrated by direct comparisons of results with SMAC reconstructions of full scale collision tests.

INTRODUCTION

The original form of the CRASH **[1, 2, 3, 4]**[1] computer program, which culminated in the CRASH3 version, was not intended to be a detailed, highly accurate reconstruction program. Rather, it was developed to serve as a simple preprocesser for the SMAC program. While the results of CRASH3 applications can be useful in providing approximate measures of accident severity for use in statistical studies, where the average error is most important, it has been demonstrated in validation tests to produce results which when compared to those of full-scale crash tests can include individual errors as great as 45%**[4]**. The possible error levels of the CRASH3 computer program are also generally applicable to the EDCRASH **[5, 6, 7]** computer program, since the CRASH3 program and the widely distributed EDCRASH clone are essentially identical. No significant analytical refinements have been made to the trajectory solution or trajectory simulation procedures of EDCRASH. The EDCRASH program, while claiming to be "within -6 to +7 percent of the combined impact speeds at a 95 percent level of confidence"**[6]** is subject to errors in individual speeds as great as 43.5% (Table 2, case 12, vehicle No. 1 **[6]**). Any "improvement" of the EDCRASH results over CRASH3 is mainly due to the "optimization" of the inputs to EDCRASH (to produce better correlation with known results) and modification of the error reporting techniques **[6]**.

This paper presents refinements to the trajectory procedure of the CRASH3 program, herein referred to as the '***refined CRASH3***' program. Refinements to the CRASH3 trajectory routine include the approximation of separation speeds through automatic iteration of the trajectory *simulation option*, inclusion of angular momentum equations, and inclusion of an approximation of the effects of the external forces (i.e., tire forces) which occur during the collision, in the CRASH3 trajectory *solution procedure*. Also to be presented is a discussion of the adaptation of optimization techniques for error reduction and convergence in the iterative solutions.

One of the problems associated with the development and refinement of any accident reconstruction technique and therefore with our research related to CRASH3 is that of demonstrating correlation with full-scale tests. The RICSAC tests **[8, 9, 10]** were specifically designed to serve as standards for such comparisons. Unfortunately, during review and utilization of the results (e.g., **[11, 12, 13]**), and particularly in prior studies which included evaluating the correlation of computer codes with RICSAC (e.g.,**[6, 14, 15, 16, 17]**) there have been various levels of interpretation and acceptance of the measured results. As a part of many of the cited projects, questions have been raised as to the validity of some of the reported RICSAC test results.

Since there has been no consensus on the interpretation of some of the results of the RICSAC tests, an intensive independent effort has been applied toward

[1] Numbers in brackets [] indicate references at end of paper

achieving proper and generally acceptable interpretations of the RICSAC test data. That research is being performed as part of a separate project reported in [18].

To permit a continuation of refinements of the CRASH3 computer code, which required comprehensive comparisons for validation, the approach reported herein has been to make use of the SMAC [19,20] computer program to generate "test results" data. The SMAC program has been demonstrated to correlate well with full-scale test results [10] [17]. It has been generally accepted that, in the absence of significant external forces, the SMAC program correctly conserves linear and angular momentum of the two-body system. For example, **Figure 1** and **Figure 2** demonstrate comparisons of the linear and angular momentum and kinetic energy for two different RICSAC test configurations simulated on test track friction (mu=0.87) and on a frictionless surface (mu=0.0).

In the SMAC program the simulated vehicles are set up to run into each other at impact locations, orientations and speeds which are identical with those of the corresponding RICSAC tests. The response data generated by SMAC includes ΔV values for the centers of gravity, including effects of vehicle rotation and external forces, separation velocities and positions of rest. Use of the SMAC results for preliminary comparisons of the effects of refinements of computer codes avoids the complications which can occur when utilizing "raw" full-scale test data due to sensor locations and test measurement variations. The SMAC program also provides a means of refining imprecise definitions in the reported test data for such items as the effective drag at the individual wheels (e.g., driveline drag for automatic or manual transmissions in gear, rubbing of damaged sheet metal on rotating wheels) and steer angles. The refinement of the items can be achieved with SMAC on the basis of matching the positions and headings at rest between SMAC and the full-scale test results.

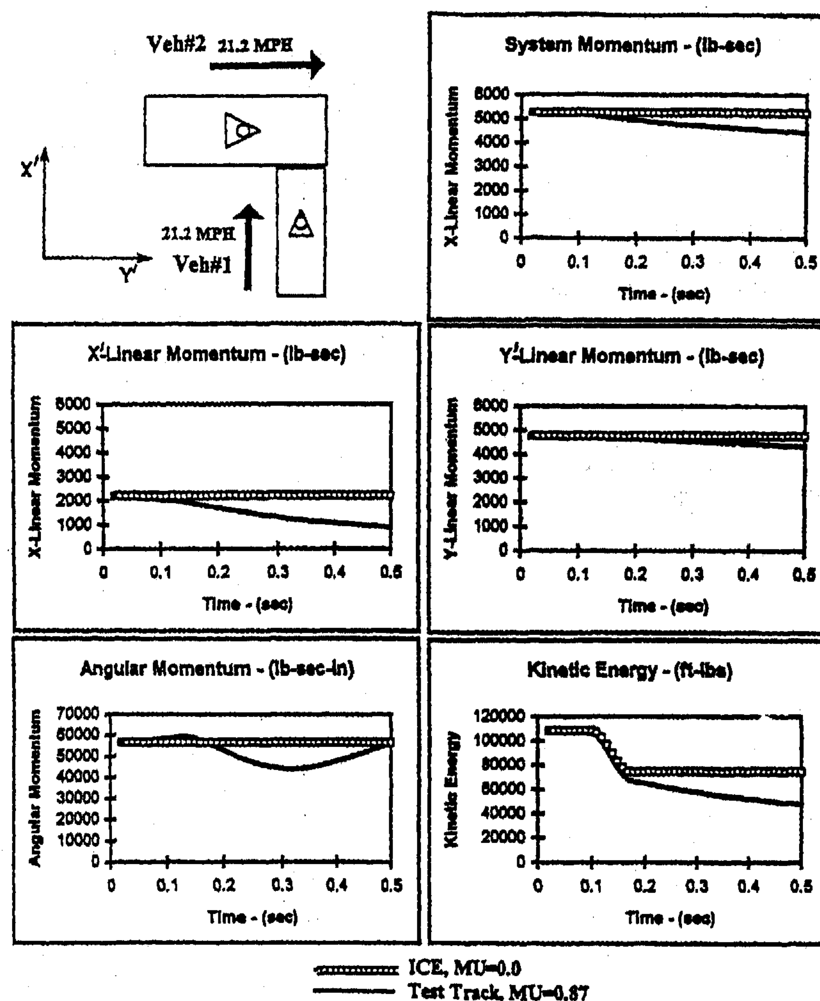


Figure 1: RICSAC Test 9: Comparison of SMAC generated System Momentum and Kinetic Energy time-histories for Test Track Friction and No Friction (note: impact at 0.10 sec, separation at 0.208 sec) (Veh#1 1975 Honda Civic,2270 lbs., Veh#2, 1974 Ford Torino, 4930 lbs.)

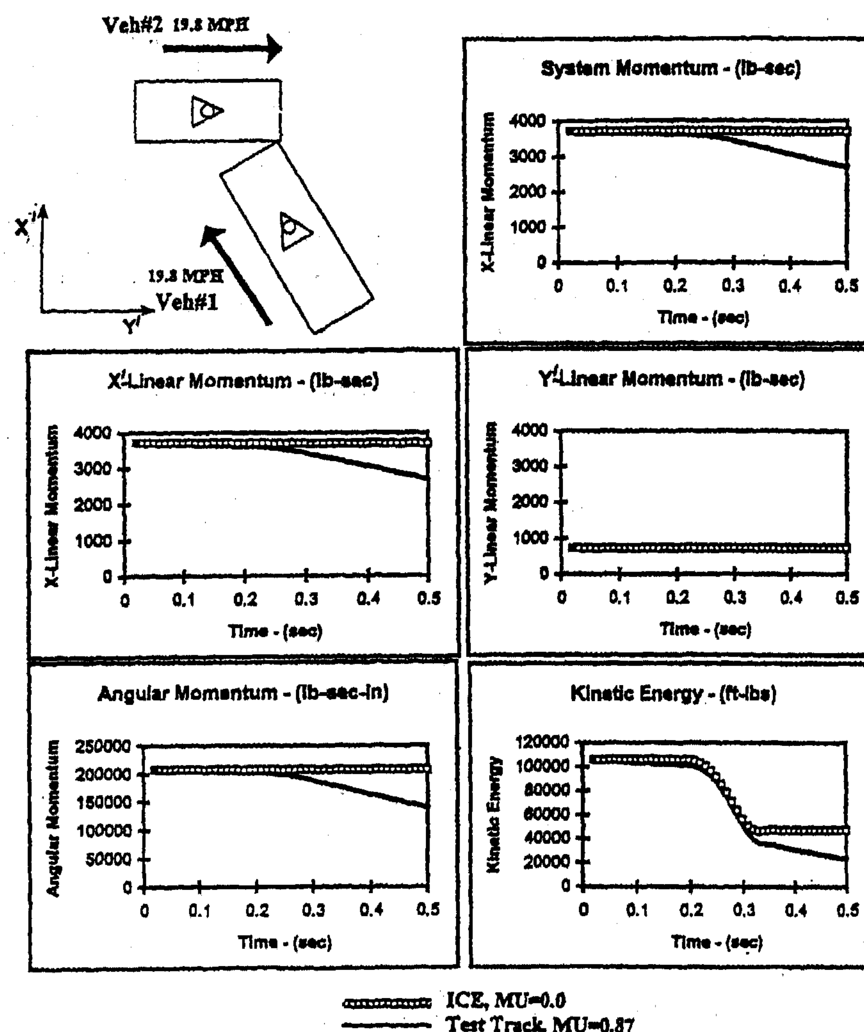


Figure 2: RICSAC Test 1- Comparison of SMAC generated System Momentum and Kinetic Energy time-histories for Test Track Friction and No Friction (note: impact at 0.18 sec, separation at 0.343 sec) (Veh#1 1974 Chevrolet Chevelle, 4650 lbs., Veh#2 1974 Ford Pinto, 3110 lbs.)

The inputs used in the present study for the SMAC runs were very similar to those that have been used in relation to SMAC validation [**10, 17**]. However, some of the cited prior SMAC inputs were found to be questionable (e.g., in some, the drag factor and steer angles at individual wheels were arbitrarily varied over time) or clearly erroneous (e.g., some had incorrect weights, tire cornering stiffness, intervehicle friction coefficient, etc.). The adjustment procedure for the SMAC inputs in the present research project was to initially check and correct any erroneous inputs and then to calculate a constant value for the wheel steer and drag factors based on the values reported in [**10, 17**]. The SMAC reconstructions were then re-run to determine the correlation of the SMAC "test results" with the full-scale test positions and orientations at rest. Minor refinements were then made, as required, to the constant values for the wheel steer angles and drag factors to achieve an approximate match of the SMAC reconstructed "test results" vehicle positions of rest with the RICSAC measured positions of rest [**21**].

Table 1 is a summary of the SMAC reconstructions of the RICSAC full-scale tests. This summary of the RICSAC tests represents a "best information" set of mathematically correct SMAC reconstructions for the individual RICSAC full-scale crash test conditions and results. The "best information" for the reconstruction results was derived from a review of the full-scale test reports [**8, 9, 22**], and a review of the reported results by Smith and Noga [**11**] to guide the SMAC reconstructions. Test numbers 1 through 12 were run at Calspan [**8, 9**]. Test number 14 is Test 3 run at Texas Transportation Institute [**22, also see 23**].

While the wheel drag factors and steer can be varied over time in the SMAC program, "validation" datasets should be representative of what can be achieved in real world applications. For normal reconstructions a constant value is generally required since detailed time-history information is not available. The intent of the refinements to the SMAC input datasets performed for this project was to create without undue delay a representative collection of mathematically

correct "test results" for comparison with CRASH predicted results. A more comprehensive review and refinement of the SMAC inputs will be performed as a part of the research reported in [**18**].

A set of input datasets for CRASH3 were then prepared which were derived from the SMAC "test results". The inputs for CRASH3 consisted of the SMAC impact and rest positions and headings, the SMAC steer angles and drag factors and the SMAC vehicle specifications (weights and dimensions) (see **Appendix 2** for the CRASH3 input datasets used). No further adjustments or refinements were made to the CRASH3 or SMAC input datasets as part of this project.

These inputs were then run with the CRASH3/EDCRASH programs and the results are depicted in **Figure 3** and **Figure 4.** These figures represent the starting point for the ***refined CRASH3*** research project and they demonstrate the general inability of the CRASH3/EDCRASH programs to consistently reconstruct the mathematically correct SMAC predicted "test results".

Table 1: Summary of RICSAC Tests Impact conditions and SMAC ΔV reconstructions

RICSAC	RICSAC TEST Impact Speed		SMAC Delta-V	
Test No.	Veh 1	Veh 2	Veh 1	Veh 2
1	19.8	19.8	14	20
2	31.5	31.5	21	30
3	21	0	10	15
4	38.7	0	16	25
5	39.1	0	15	27
6	21.5	21.5	9	14
7	29.1	29.1	13	20
8	20.8	20.8	14	12
9	21.2	21.2	20	7.7
10*	33.3	33.3	36	16
11	20.4	20.4	28	17.5
12	31.5	31.5	40	27
14	26	38.5	18.9	19.8

***Note: The ΔV for Test #10 is the sum of the primary and secondary (side-slap) ΔV.**

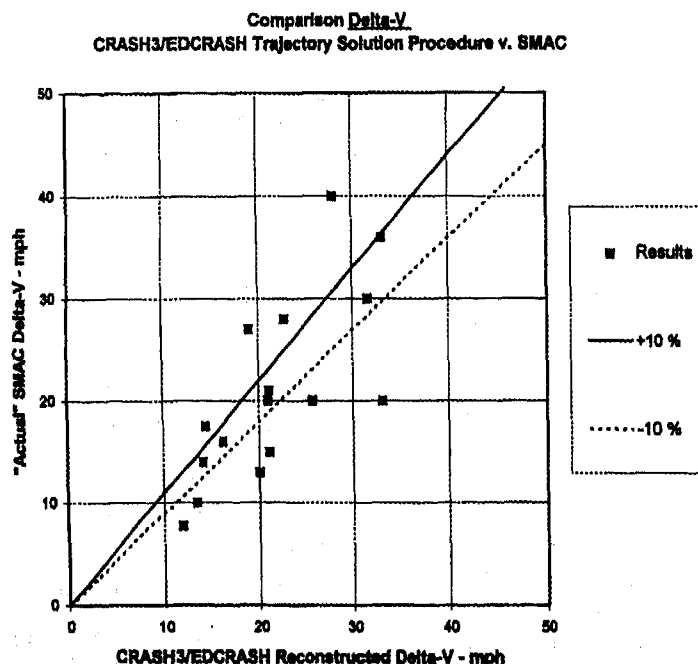


Figure 3 Initial comparison of reconstructed ΔV of CRASH3/EDCRASH Trajectory Solution vs. SMAC

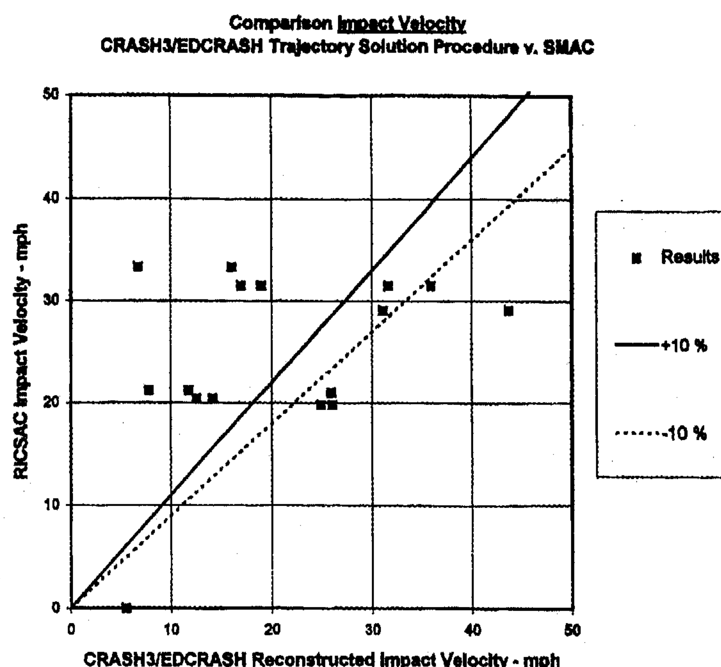


Figure 4 Initial comparison of reconstructed <u>Impact Velocity</u> of CRASH3/EDCRASH Trajectory Solution vs. RICSAC Test Results

BACKGROUND

When vehicles separate after a collision, they move to rest positions against resistance forces produced primarily by tire-ground friction. Analysis of the total energy dissipated as the vehicles travel from separation to their positions of rest and determination of their corresponding linear and angular velocities at separation constitute the essence of a *trajectory-based* reconstruction of a collision. The principles of conservation of linear and/or angular momentum are applied to the directions and magnitudes of the system momentum at separation to determine the velocities which must have existed prior to the collision.

In simple spinout motions, the actual paths traveled from separation to rest can be approximated, with reasonable accuracy, by straight lines between the separation and rest positions. The separation velocities can then be estimated on the basis of the total work done by each vehicle against tire-terrain friction forces between separation and rest. The rate of energy dissipation by tire forces is dependent on the heading direction of the vehicle in relation to its direction of motion, the rate of yawing rotation, and the extent of rotational resistance at the individual tires. For example, in a broadside slide, all tires produce full-friction resistance forces. In forward or backward motion, only those tires with applied brakes, damage effects, large steer angles or driveline braking produce significant drag forces. For the case of linear motion combined with angular rotation about a vertical axis (i.e., yawing rotation) a vehicle alternates between the two conditions of resistance of motion.

The CRASH3 program includes a simplified analytical procedure for approximating the linear and angular velocities of a vehicle subsequent to a collision. The developed procedure is referred to as SPIN2. A subcontract to refine SPIN2 undertaken in 1979 [24] (see **Appendix 1** for a discussion of SPIN2) led to the conclusion in **[24]**:

"To achieve a general improvement in the reliability and accuracy of approximations of the angular and linear velocities at separation, a step-by-step time history form of trajectory solution should be implemented."

The CRASH3 program also includes an exploratory *trajectory simulation* option solution procedure based on the SMAC trajectory model. The optional *trajectory simulation* procedure (USMAC) includes routines from the trajectory portion of the SMAC program to permit time-history simulations in CRASH of the spinouts of the individual vehicles from separation to rest.

The USMAC *trajectory simulation* model is a three degree of freedom (X, Y, PSI) mathematical representation of planar motion. The tire side force calculations are based upon a nondimensional side force function whereby the small-angle properties of the tires "saturate" at larger angles. The "friction circle" concept is used to approximate the interactions between side and circumferential (braking or tractive) tire forces. The "friction circle"

concept is based on the assumption that the maximum value of the resultant tire friction force is independent of its direction relative to the wheel plane.

The purpose of the USMAC routine in CRASH3 was to serve as a check of the SPIN2 approximations of separation speeds. An optional iterative procedure was also included in the CRASH3 *trajectory simulation option* to automatically adjust the SPIN2 separation velocities in an attempt to reduce errors in the predicted vs. actual final rest positions. The initial form of the trajectory iteration routine was implemented merely to demonstrate feasibility and was not thoroughly tested and evaluated. The costs of a CRASH run increased ten-fold by the use of the exploratory iterative trajectory solution procedure (USMAC) (e.g., **[25]** , circa. 1976, p1.,"The computer costs ... of the CRASH program ...range from approximately $1.00 to $10.00 per case. The upper end of the indicated cost range corresponds to a run in which the option for testing and refining the trajectory analysis portion of the calculation has been exercised"). There was no further NHTSA sponsored development of the original exploratory implementation of the USMAC routine.

By the mid 1990's, with the prevalence of extremely powerful and inexpensive Pentium PC's, and therefore the availability of virtually unlimited computer resources, consideration was given to internal research by McHenry Consultants, Inc. to further develop the *trajectory simulation* routine of the CRASH3 program. The objective in our refinements of the CRASH3 accident reconstruction procedures has been to simplify the input requirements of the program while providing a significantly improved correlation of the reconstruction results with known test results. A secondary consideration in the form of the refinement has been to limit the total computational time for convergence on a solution to a reasonable amount of time.

Effective refinements of the *trajectory simulation* procedure can substantially increase the usefulness of the simple "closed-form" CRASH3 accident reconstruction procedure by producing a general refinement of the *trajectory solution* procedure. The CRASH3 *trajectory solution* procedure requires a minimum amount of input information about the accident scene and vehicles. An effective improvement of the *trajectory solution* procedure can be expected to substantially improve the correlation or "validation" of the CRASH3 model when comparing the reconstruction results with full-scale test results. A 1989 study **[6]** concluded that the original form of the CRASH3/EDCRASH *trajectory simulation* option can actually degrade the *trajectory solution* results of the CRASH3 program.

A secondary task required in order to further refine and enhance the *trajectory solution* procedure of the CRASH3 program was a reactivation and refinement of the angular momentum solution procedure. The original CRASH program included conservation of linear momentum in the trajectory based solution to determine the impact speeds based on the separation velocities. A contract performed on CRASH2 to implement an angular momentum solution achieved mixed results **[26]**. A major hurdle for any procedure which includes an angular momentum solution is the need to approximate movement of the vehicles during the collision. In the CRASH2 formulation the impact and separation positions and headings were assumed to be identical. The research in **[26]** revealed that the accuracy of an angular momentum solution procedure for accident reconstruction which includes the assumption of no movement between impact and separation will produce unacceptable error levels (>>20%) in many cases.

Other analytical accident reconstruction techniques which include provision for an angular momentum solution procedure and/or which are based on conventional momentum analyses, include the somewhat subjective input requirement that either a vehicle-to-vehicle contact "point" **[27]**, or a "point of maximum engagement" **[28]** or an "impact center" **[29]** be specified. The additional input is required to compensate for the cited solution procedure's lack of an independent determination of separation positions and orientations.

The requirement that the user specify either an arbitrary impact contact "point" or an arbitrary "point of maximum engagement" detracts from the objectivity of the reconstruction techniques.

Figure 5 and **Figure 6** show representative changes in positions and orientations during the contact phase of collisions.

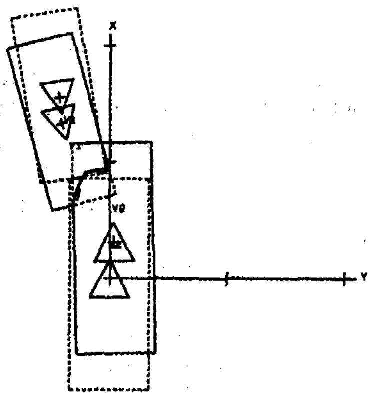


Figure 5 Impact and Separation Positions and Orientations for RICSAC Test #12

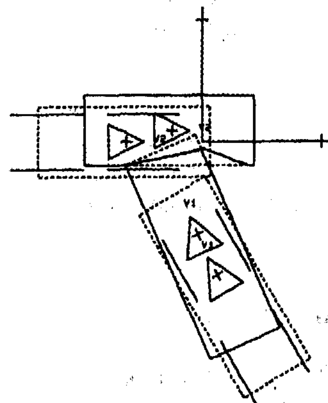


Figure 6 Impact and Separation Positions and Orientations for RICSAC Test #1

The subjective choice of a "point" can produce a large variation in the predicted results. During "validation," when the results are known, the user has some guidance in choice of the subjective "point." In real-world applications, where the answer is not known, the determination and arbitrary specification of a "point" can and will produce a wide range of predicted results. The normal input requirements of accident reconstruction programs of damage dimensions and approximate impact configurations should provide more than adequate information for any accident reconstruction program to independently achieve the function of any contact "point" or "point of maximum engagement" without user intervention. The movement of the vehicles between impact and separation can be initially approximated, for example, by moving the vehicles in their initial directions of motion to positions where the damage regions match. The procedure to determine a separation position should be automated to prevent subjective variations between users in the positions of match and therefore the results.

Other assumptions of the cited techniques **[27,28,29]** which may detract from the validity of their impact models for objective application to accident reconstruction are:

1. During the impact no consideration is given for ***tire-to-ground "external" forces***
2. The ***impact duration*** and time for exchange of momentum is assumed to be infinitesimally small.

TIRE-TO-GROUND "EXTERNAL" FORCES: The effects of tire-ground forces must be considered in a motor vehicle collision reconstruction. During the early development of the SMAC program **[19, 20]** tests were performed to determine the effects of external tire forces on the collision solution procedure. It was concluded that "The conventional assumptions that the effects of vehicle deformations and of tire forces can be neglected in analytical reconstructions of collisions can lead to significant errors. This is particularly true for intersection-type collisions at low to moderate vehicle speeds, in which prolonged or multiple contacts and significant movements of the involved vehicles occur" and that "therefore it is essential in a general procedure for reconstruction calculations that both the collision and tire forces be considered simultaneously."

IMPACT DURATION: The duration of a motor vehicle collision cannot be assumed to be infinitesimally small. Normally the exchange of momentum requires 50 to 125 milliseconds. Significant changes in positions and orientations can occur during the collision which can produce changes in the collision moments acting on the collision partners. Any accident reconstruction

solution procedure which contains the assumption of an instantaneous exchange of momentum should be carefully evaluated.

The importance of the inclusion of external forces in collision analysis and use of a finite time increment for the impact duration has also been reiterated more recently by Fonda [30].

RESEARCH APPROACH

The refinement of the trajectory solution procedure of CRASH3 required the following two basic tasks:

1. **Refinement of the trajectory *simulation option*** to improve the prediction of separation velocities.
2. **Refinement of the trajectory *solution procedure*** to include implementation of an angular momentum solution procedure and to include provisions for vehicle movement and external forces during the collision.

REFINEMENT OF THE TRAJECTORY *SIMULATION OPTION*

The refinement of the trajectory *simulation option* of CRASH3 began with determination of what constitutes the minimal information required for a trajectory based reconstruction. The minimum information required for a trajectory based reconstruction is contained in **Table 2**.

Table 2: Minimum Information Required for a trajectory based reconstruction:

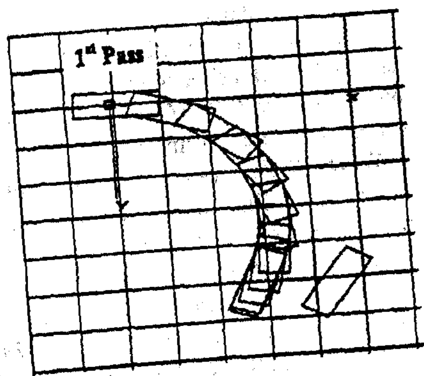
1. Impact and rest positions and headings.
2. Approximations of wheel steer and drag.
3. Vehicle specifications (weights, dimensions).

The CRASH3 program includes optional inputs for points on curve, end of rotation position and other inputs which were intended to permit use of such information when it is clearly indicated by physical evidence. However, the cited options have frequently been applied as arbitrary, subjective inputs. One intention in the reported improvement of the trajectory simulation routine has been to eliminate the need for any supplemental, sometimes subjective, inputs and to thereby create a reconstruction technique which will provide *uniform interpretations of physical evidence.* The use of sometimes arbitrary points-on-curves and end-of-rotation positions as inputs can detract from an objective analysis.

One major advantage of the present refinement of the trajectory simulation procedure for CRASH3 over the original development was the availability of essentially unlimited computer resources. The number of iterations of the trajectory algorithm for a given reconstruction is virtually unlimited on a PC. The only limitation is clock time and the feasibility of attaining a solution in a reasonable amount of time. There are no computer resource charges for execution of a simulation on a PC. Also, on a Pentium-60 or faster CPU, a 10 sec trajectory simulation of a vehicle spinout requires much less than 1 sec of real time. Therefore large numbers of iterations can be performed, if required. In the CRASH3 form of the trajectory simulation, computer resource concerns dictated an arbitrary limitation on iterations of 5 per run and in many instances, five iterations required more than 10 minutes real clock time to execute in a time-share computer environment [31].

The next phase in the refinement of the trajectory *simulation option* for CRASH3 was to determine how to create an iterative scheme which could converge on the separation speeds required to travel a given distance in a given direction and to match the extent of rotation. The only items of information to be provided to the routine were those listed in **Table 2**. The routine would then determine the required separation speed, course angle and angular velocity.

Figure 7 and **Figure 8** show representative results of iterative adjustments of the separation velocities by revised CRASH3 that produce convergence towards matching the positions and headings at rest.



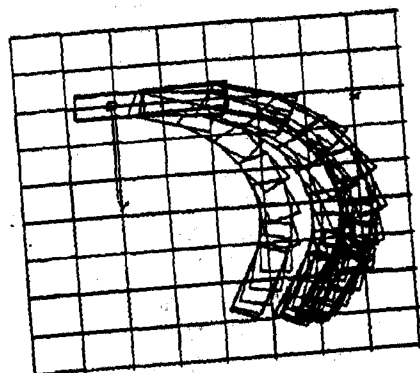
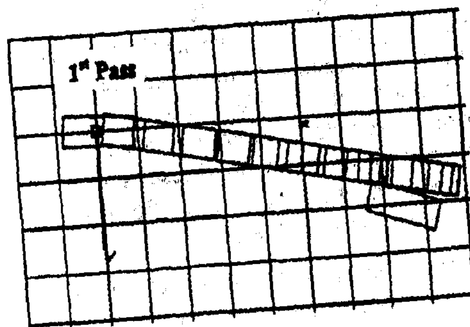


Figure 7 Sample of Revised CRASH3 Iterative Trajectory Simulation for RICSAC Test#4, Veh#1



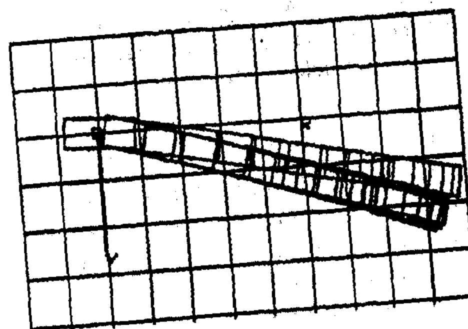


Figure 8 Sample of Revised CRASH3 Iterative Trajectory Simulation for RICSAC Test#7, Veh#1

Many different optimization and error minimization routines were investigated for convergence and optimization to a solution [32,33,34, 35]. A fundamental problem with the use of many of the investigated control algorithms was the inherent requirement that the functions must be continuous and/or linear. The spinout trajectory of a vehicle can be a highly non-linear event. Minor variations in starting conditions (i.e., speed, course angle, angular velocity) can produce major changes in the resulting rest positions (X, Y, PSI) and discontinuities in the calculated error evaluation terms. For example, during the various "step" changes in the decelerations of the linear and angular velocities as a vehicle travels from separation to rest, at any instant when the velocity vector aligns with the longitudinal axis the vehicle may "shoot off" tangentially in what has been described as a "dog-leg" type of trajectory. Traditional function minimization techniques tested as part of this research which normally require the evaluation of some form of derivatives (e.g., Cramer's rule, Newton's method) or include the assumption of a linear function (Powell's method, Broyden's method) were found to fail in many instances where step changes were produced in the "function" by minor alterations of the variables.

The choice of error terms and the relative magnitudes of the various calculated error terms can also cause iterative and optimization routines to fail. Within the reported research program error terms were developed which achieve a successful minimization of errors between predicted and actual positions of rest.

Problems encountered during the early formulation and comparisons of the revised CRASH trajectory simulation iteration procedure results (project USMAC2) with reported full-scale crash test results of RICSAC [8,9,10] revealed the need for an independent and separate task of evaluating the reported RICSAC test results. Most of the problems with comparisons were related to the values of separation conditions reported by Jones [10]. The reported values included some interpretation by Jones of the full-scale test results. In the full-scale

tests, the separation velocities were not directly measured and required some additional calculations which included interpretation of accelerometer information for accelerometers not located at the vehicle centers of gravity. Also, the accelerometers in some of the tests were located in or very near to the crush zone and/or on possible flexible components of the vehicle. Therefore a separate task was undertaken to check the reported results of RICSAC independent of any validation efforts for a particular computer program **[18]**.

The inputs for the ***refined CRASH3*** for the present project were derived from SMAC "test result" reconstructions (previously discussed in the introduction) and are contained in **Appendix 2**. The inputs consisted of the SMAC impact and rest positions and headings, the wheel steer angle and drag factors and the vehicle specifications (weights and dimensions).

A refined version of the original SPIN2 routine was utilized to determine starting values. The revised trajectory simulation routine, USMAC2, was then used to simulate the trajectory of each of the vehicles to positions of rest. Based on comparisons of the cited error terms, automatic adjustments were made to the separation conditions and the simulation was re-run. Once an acceptable "match" of the rest position and heading was achieved, the iterative procedure was stopped.

Table 3 contains SMAC generated separation conditions for the RICSAC full scale tests. The separation velocities are a direct output from the SMAC program and represent the first instance after the primary collision where the acceleration drops below 1 g-unit. **Table 4** contains a preliminary comparison of the USMAC2 with the SMAC generated separation conditions. Note that the comparison dataset does not include application of the angular momentum solution or external forces (these items are discussed in more detail below).

A review of the results reveal that most of the USMAC2 predicted separation speeds compare favorably with the SMAC "test result" generated separation speeds. However, some of the comparisons do not appear to agree very well. The reason for the differences is related to the assumptions regarding the time to separation. As a part of the ***refined CRASH*** trajectory solution procedure, a separation position is calculated. The calculation of a separation position requires assumption of a time to separation as well as reliance on the ***refined CRASH*** calculated impact speed. In order to provide for uniform interpretation in the ***refined CRASH*** results, a common assumption was made for the time to separation in the calculation procedure for the separation position and orientation. As more full scale tests become available (e.g., possibly the JARI crash test database which contains some 35 tests that are referenced in **[29]**) possible refinements of the technique will include modifications of the time to separation, etc. based on the impact configuration (e.g., a 90 degree side impact would be expected to have a longer time to separation than an offset frontal).

Another source of some discrepancy of the comparisons is related to the occurrence of secondary "side-slap" interactions. In tests 9 and 10 both the full-scale tests and SMAC reconstructions included secondary "side-slap" collisions. In test 10, a "side-slap" with magnitude greater than 5 mph on vehicle 1 occurred after the initial impact. The "side-slap" changed the magnitude and direction of the separation conditions. In test 9, subsequent to the initial contact the vehicles "side-slap" interacted at a lesser magnitude but for a longer time which produced re-direction of the initial separation velocities. The listed separation speeds in **Table 4** are the initial impact separation conditions which are different than the speeds and orientations after the occurrence of a "side-slap" collision.

The development of the trajectory *solution procedure*, discussed below, revealed that the cited comparison of SMAC and CRASH separation results, while not all within ±10%, were adequate when used with the refinements to the *trajectory solution* procedures to produce acceptable impact and ΔV predictions within ±10% of the corresponding SMAC results.

Table 3 SMAC generated separation conditions for 13 RICSAC tests

Test	Vehicle No. 1			Vehicle No. 2		
Test No.	SDOT1 in/sec	GAMS1 deg	PSID1 deg/sec	SDOT2 in/sec	GAMS2 deg	PSID2 deg/sec
1	125	-11.4	95	235	23.5	-12
2	210	-12	160	411	26	65
3	205	-6	-3	272	11	19
4	400	-6	3	432	13	40
5	435	-5	5	469	10	63
6	210	4	51	270	78	184
7	276	2	63	418	76	167
8	169	35	103	315	55	34
9	169	69	193	310	69	-44
10	282	64	357	488	67	-78
11	140	-29	-44	83	44	4
12	138	-30	-87	152	53	-27
14	282	49	-280	352	102	-200

Table 4: Comparison of USMAC2 predicted separation conditions with the SMAC generated separation results.

Test No.	Veh No.	SDOT In/sec	Error %	GAMS deg	Error deg	PSID deg/sec	Error deg/sec
1	1	108	-14	-11	0	103	8
2	1	227	8	-18	-6	147	-13
3	1	205	0	-12	-6	-1	2
4	1	439	10	-15	-9	29	26
5	1	431	-1	-5	0	7	2
6	1	198	-6	2	-2	36	-15
7	1	263	-5	1	-1	53	-10
8	1	138	-18	37	2	97	-6
9	1	274	62	57	-12	202	9
10	1	355	26	81	17	208	-149
11	1	118	-16	-34	-5	-42	2
12	1	129	-7	-35	-5	-93	-6
14	1	279	-1	46	-3	-270	10
1	2	257	9	25	2	7	19
2	2	422	3	29	3	59	-6
3	2	274	1	14	3	-32	-51
4	2	448	4	23	10	30	-10
5	2	508	8	8	-2	130	67
6	2	294	9	82	4	159	-25
7	2	419	0	78	2	159	-8
8	2	330	5	53	-2	42	8
9	2	269	-13	65	-4	6	50
10	2	494	1	63	-4	-30	48
11	2	82	-1	29	-15	11	7
12	2	157	3	46	-7	-23	4
14	2	322	-9	104	2	-198	2

REFINEMENT OF THE TRAJECTORY *SOLUTION PROCEDURE*

The primary task required to refine the trajectory *solution procedure* of CRASH3 was to implement an angular momentum solution. The implementation of the procedure required determination of separation positions and orientations. A secondary consideration was the implementation of effects of external tire forces during the impact phase.

The problems of approximation of positions and headings at separation are compounded by the fact that neither the impact nor the separation velocities are known. The ***refined CRASH*** trajectory *solution procedure* therefore had to be modified to include an iterative solution procedure.

A basic outline of the procedure used to determine the impact speeds for a combination linear and angular momentum solution was as follows:

1. The separation velocities are approximated on the basis of the vehicle travel from impact to rest.
2. The separation velocities are used with an application of conservation of linear momentum for an initial approximation of the impact velocities.
3. The separation positions and headings are approximated using the initial approximations of the impact and separation velocities.
4. With approximation of the separation positions and headings, the following steps are repeated to converge on a solution:

    4.1. The separation velocities are refined based on the vehicle travel from *separation* to rest.

    4.2. The refined separation velocity is used with an application of conservation of both linear and angular momentum for a refined approximation of the impact velocities.

An iterative procedure was developed to guide the iteration of step 4, until an "acceptable" solution was achieved. The task became both a choice of iterative schemes and determination of what constituted an "acceptable solution." It was hoped that once the solutions for the angular and linear solutions matched, an acceptable solution would result.

However, during the testing of various iteration schemes, the question became which solution procedure, linear or angular, should guide the iterative process. Some impact configurations are more suited to a linear momentum solution procedure and some are more suited to an angular momentum solution procedure. Also available to the CRASH program algorithm is a damage analysis procedure which can provide an additional discriminator to guide the iterative process. However, although included in initial testing, adequate approximations of the separation positions and orientations where achieved without the use of damage information.

Coincident with the investigation of iterative solution procedures was the implementation of terms to approximate the effects of external forces. Terms were developed and included in the momentum equations to approximate the forces and moments produced by the tire forces during the collision. The implementation included investigation of approximate values for the finite duration of the collision.

The successful inclusion of effects of external tire-forces and of a finite time for the collision momentum exchange in the ***refined CRASH*** solution procedure benefited from the outstanding resources afforded by modern day Pentium computers. Rather than being forced to perform iterations of individual runs for comparison tests of the effects of individual input variables, the whole collection of test cases could be iterated by control algorithms to quickly evaluate the effects of the assumptions on all of the validation comparison runs. A "suite" of tests was easily set up and the various variables tested. Some of the many assumptions and approximations tested and refined in this manner included:

- The approximation technique used for determining the separation positions and orientations.
- The approximate duration of the time to separation.
- The magnitude and duration of the forces and moments used to approximate the effects of external tire-forces during the collision.
- The various error evaluation terms and weighting functions used to guide the iteration and achieve an "acceptable" match.

RESULTS

A comparison of the ***refined CRASH3*** trajectory solution procedure with the time-forward SMAC simulation program for 13 of the RICSAC tests is presented in **Table 5** and in graphical form for ΔV in **Figure 9** and Impact Velocity in **Figure 10.** The results demonstrate that the "closed-form" ***refined CRASH*** program can reconstruct "open-form" SMAC generated "test results" within less than ±10% of the impact speeds and impact speed-changes.

When comparisons are made between a calculated value and a measured or "true" value, the magnitude of the error expressed as a percentage of the comparison with the "true" value of an estimated quantity appears to be the most meaningful engineering measure of accuracy. However, in the case of a very small or zero "true" value, the use of the actual magnitude of the error is more meaningful than a percentage. If it is assumed that ±1.0 MPH is an acceptable error range about a very small "true" value, then the transition from the actual magnitude of error to a percentage error can arbitrarily be made, for example, at a "true" value of 10 MPH. The error should be reported in both formats for "true" values in excess of 10 MPH. For lesser magnitudes, only the actual magnitude should be displayed (e.g., **Table 5,** note that "true" value for **Table 5** is the SMAC simulated "test result" reported in **Table 1.** The table gives a correlation of the predicted results of ***refined CRASH3*** with the "true" SMAC "test results")

In the validation of EDSMAC contained in [6] the authors used the combined speeds of the two colliding vehicles (i.e., the closing speed of the collision) as a basis for calculating the error in the reconstructed impact speed. This technique was used to avoid the case where one vehicle was initially at rest and therefore had a "true" value of zero for its impact speed. The technique used appears to serve no useful purpose other than making the error appear smaller.

The SMAC "open-form" simulation is a widely used program that has been demonstrated to be a generally valid and accurate model of motor vehicle collisions. The primary drawback of the SMAC program is that it requires initial estimates and multiple iterations by the user to converge on an "acceptable" solution. Variations of what constitutes an "acceptable" solution for the SMAC simulation program are to some extent dependent on the individual users.

The ***refined CRASH3*** procedure achieves essentially the same results as the "open-form" SMAC computer program. The ***refined CRASH3*** "closed-form" approach eliminates differences in results for a given reconstruction that are produced by variations in the capabilities and perseverance of individual SMAC users.

CONCLUSIONS AND RECOMMENDATIONS

With regard to measures of reconstruction accuracy by means of comparison of reconstruction results with reliable test data, the distinction must be recognized between applications to statistical samples of cases and an application to an individual case.

In a large number of applications, the intercept and slope of the linear regression fit for actual measured data plotted against reconstruction estimates may be compared with the intercept of zero and slope of 1.00 corresponding to perfect correlation (e.g., see **Figure 9, Figure 10**). In this manner, measures of the average error may be obtained.

In an application to an individual accident, the maximum error in a comparison of reconstruction estimates with reliable test data constitutes the most significant measure of accuracy since it indicates the error that is possible with good data in any given application.

Subsequent to the research related to a review of the RICSAC tests **[18]**, the ***refined CRASH3*** program will be compared directly with the refined RICSAC full-scale test results.

Further development of the ***refined CRASH*** program, which will follow the completion of work related to RICSAC **[18]** and the availability of additional full scale crash test results, should further improve the results of ***refined CRASH*** to be generally consistent with those of the SMAC simulation program. The preference for the CRASH program, particularly for case studies, is that it does not require any iterations by the user and, therefore, there should be little or no variation in reconstruction results obtained by different users.

Table 5 Summary of *refined CRASH3* reconstructions of RICSAC SMAC simulations:

RICSAC	Impact Speed						Delta-V					
TEST	Vehicle No. 1			Vehicle No. 2			Vehicle No. 1			Vehicle No. 2		
No.	Revised	Error		Revised	Error		Revised	Error		Revised	Error	
	crash	%	mph	crash	%	mph	crash	%	mph	crash	%	mph
1	19	-4	-.8	21.1	6.6	1.3	14.2	1.4	.2	19.9	-.5	-.1
2	33.4	6	1.9	32.1	1.9	.6	21.8	3.8	.8	31.5	5	1.5
3	20.8	-1	-.2	0		0	9.8	-2	-.2	15.6	4	.6
4	40.7	5.2	2	0		0	16.3	1.9	.3	25.5	2	.5
5	39.9	2	.8	0		0	15.9	6	.9	28.9	7	1.9
6	21	-2.3	-.5	22.8	6	1.3	9.8	8.9	.8	14.9	6.4	.9
7	28.7	-1.4	-.4	29.4	1	.3	13.8	6.2	.8	21.5	7.5	1.5
8	19.3	-7.2	-1.5	20.7	-.5	-.1	14.6	4.3	.6	12.1	.8	.1
9	21.9	3.3	.7	21.1	-.5	-.1	18.2	-9	-1.8	7.6	-1.8	-.1
10	34.1	2.4	.8	35.3	6	2	37	2.8	1	17.1	6.9	1.1
11	20	-2	-.4	19.9	-2.5	-.5	26.5	-5.4	-1.5	16.1	-8	-1.4
12	33.3	5.7	1.8	32.1	1.9	.6	39.1	-2.3	-.9	26.6	-1.5	-.4
14	25.9	-.4	-.1	36.5	-5.2	-2	19.6	3.7	.7	19.9	.5	.1

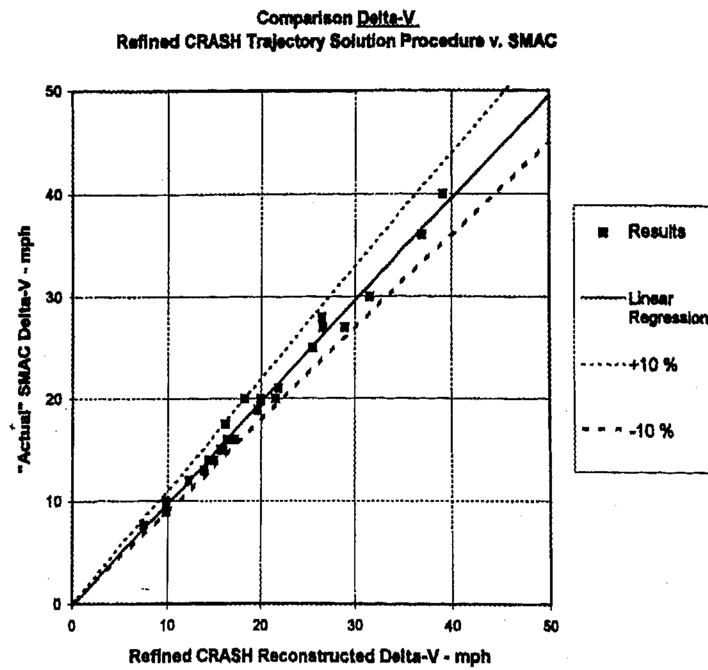


Figure 9 Comparison of reconstructed ΔV of *refined CRASH3* Trajectory Solution vs. SMAC

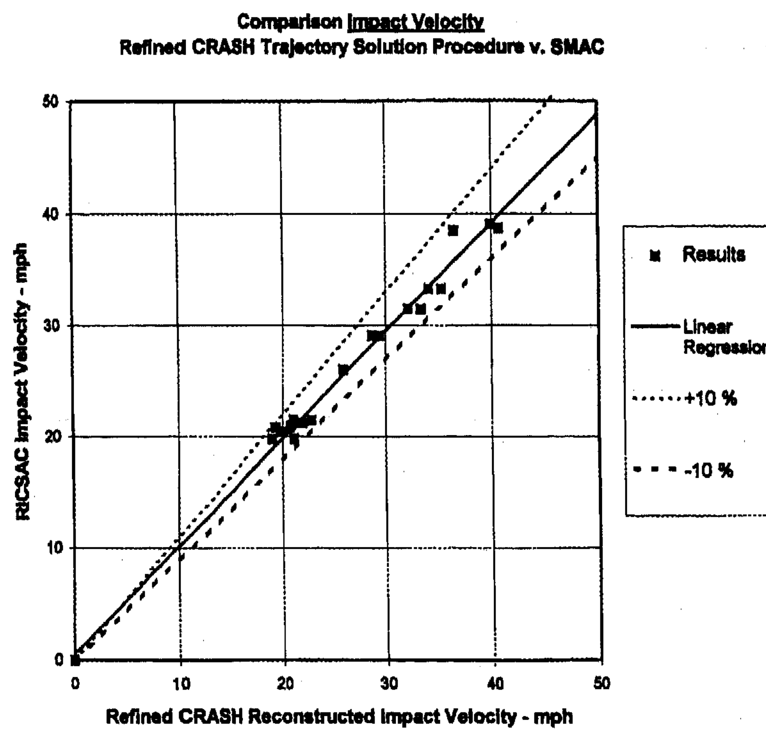


Figure 10 Comparison of reconstructed Impact Velocity of *refined CRASH3* Trajectory Solution vs. RICSAC Test Data

A secondary investigation is underway to determine internal tests and corresponding instructions that may be used to guide users if and when "bad" data are input. Since the results reported herein represent correlation of the predicted CRASH results with mathematically correct SMAC generated "test results", they indicate the 'best' correlation possible for the analytical technique. The next task is to test the procedure with "real world" data where problems may arise due to invalid or inconsistent inputs.

An auxiliary procedure developed during this project was a set of modifications for SMAC which automatically generates "evidence" for input to the ***refined CRASH3*** program. This routine provides a capability for further testing of the compatibility of results produced by the two programs. In tests performed to date, the ***refined CRASH3*** routine has consistently produced reconstruction results for the general case which are within ±15% of the speed inputs to the SMAC program and the corresponding impact speed-changes reconstructed by SMAC. Further testing and evaluation of the described routine and of the compatibility of the SMAC and CRASH program results will continue.

The final development of the ***refined CRASH3*** accident reconstruction procedure will incorporate previously defined refinements to the damage analysis procedure **[36,37]**. The utilization of damage information should assist in the analysis of real-world data and should dramatically improve the ability of the ***refined CRASH3*** to accurately reconstruct motor vehicle collisions while maintaining the original intent of the CRASH program to provide for *"uniform interpretation of accident evidence"*.

REFERENCES

1. McHenry, R.R., "The CRASH Program - A simplified Collision Reconstruction Program" Proceedings of the Motor Vehicle Collision Investigation Symposium, Calspan, 1975
2. McHenry, R.R. Lynch, J.P., "User's Manual for the Crash Computer Program" Calspan Report No. ZQ-5708-V-3, Contract No. DOT-HS-5-01124, Jan 1976
3. NHTSA,"CRASH3 User's Guide and Technical Manual", Revised Edition, National Highway Traffic Safety Administration, DOT-HS-805732, April 1982
4. Tsongos, N.G., "CRASH3 Technical Manual", U.S. DOT, NHTSA, NCSS, Accident Investigation Division, July 1986
5. Day, T.D., Hargens, R.L., "Differences Between EDCRASH and CRASH3", SAE Paper 85-0253
6. Day, T.D., Hargens, R.L., "Further Validation of EDCRASH Using the RICSAC Staged Collisions", SAE Paper 89-0740
7. Day, T.D., Siddall, D.E., "Validation of Several Reconstruction and Simulation Models in the HVE Scientific Visualization Environment", SAE Paper 96-0891
8. McHenry, R. R., Lynch, J. P., Segal, D. J., "Research Input for Computer Simulation of Automobile Collisions", Calspan Report ZQ-6057-V-1, Contract DOT-HS-7-01511, Jun 1977
9. Shoemaker, N. E., "Research Input for Computer Simulation of Automobile Collisions - Staged Collisions, Vol. II & Vol. III", Calspan Report ZQ-6057-V-4 & V-5, Contract DOT-HS-7-01511, December 1978
10. Jones, I. S., Baum, A. S., "Research Input for Computer Simulation of Automobile Collisions - Volume IV - Staged Collision Reconstructions", Calspan Report ZQ-6057-V-6, Contract DOT-HS-7-01511, December 1978
11. Smith, R.A., Noga, J.T., "Examples of Staged Collisions in Accident Reconstruction", presented at ASME Winter Meeting, Nov 1980 and contained in ASME publication "Highway Collision Reconstruction
12. Brach, R. M., "Impact Analysis of Two-Vehicle Collisions", SAE Paper 83-0468
13. Bundorf, R.T., "Analysis and Calculation of Delta-V from Crash Test Data", SAE Paper 96-0899
14. Woolley, R.L., "The "IMPAC" Computer Program for Accident Reconstruction", SAE Paper 85-0254
15. Wooley, R.L., Warner, C.Y., Tagg,M.D., "Inaccuracies in the CRASH3 Program", SAE Paper 85-0255
16. Cliff, E.C., Montgomery, D.T., "Validation of PC-CRASH - A Momentum Based Accident Reconstruction Program", SAE Paper 96-0885
17. Day, T. D., Hargens, R.L., "Further Validation of EDSMAC Using the RICSAC Staged Collisions", SAE Paper No. 90-0102
18. McHenry, B.G., McHenry, R. R., "RICSAC-97 - A Reevaluation of the Reference Set of Full Scale Crash Tests", SAE Paper 97-0961, to be published at the 1997 SAE International Congress
19. McHenry, R.R., "Development of a Computer Program to Aid the Investigation of Highway Accidents", Contract FH-11-7526, December 1971, Calspan Report VJ-2979-V-1,NTIS PB# 208537
20. McHenry, R.R.,"A Computer Program for Reconstruction of Highway Accidents", SAE Paper 73-0980, Proceedings of the 17th Stapp Car Conference, November 1973
21. The refinements of the SMAC inputs performed as part of this research are a preliminary step towards the creation of a RICSAC "test data" set. A more thorough check and refinement is to be performed as a part of other research **[18]**. At completion of that research a re-check of the ***refined CRASH3*** correlations will then be performed and published.
22. James, M.E., Hayes, E.R., Whittington, C., "Improvement of Accident Simulation Model and Improvement of narrow Object Accident Reconstruction", Texas Transportation Institute, Report RF3258,
23. Note that two other full-scale tests were run and reported in **[22]**. The other tests included secondary impacts (not side-slaps) between the collision partners. Secondary impacts will be addressed and included in future refinements of the ***refined CRASH3*** program

24. McHenry, R.R., McHenry, B.G., "National Crash Severity Study - Quality Contol, Task V: Analysis to Refine Spinout Aspects of CRASH", Calspan Field Services, Inc. ZP-6003-V-4; DOT-HS-6-01442, January 1981
25. McHenry, R.,R., Lynch, J.P., "CRASH 2 User's Manual", Contract DOT-HS-5-01124, Calspan Report ZQ-HS-5-01124, November 1976
26. McHenry, R.R., Lynch, J.P., "Revision of the CRASH2 Computer Program", US DOT HS-805-209, September 1979
27. Limpert, R., Andrews, D.F., "Linear and Rotational Momentum for Computing Impact Speeds in Two-Car Collisions (LARM)" , SAE paper 91-0123
28. Steffan, H., Moser, A.,"The Collision and Trajectory Models of PC-CRASH", SAE Paper 96-0886
29. Ishikawa, H., "Impact Center and Restitution Coefficients for Accident Reconstruction", Japan Automobile Research Institute, SAE Paper 94-0564
30. Fonda, A.G., "Nonconservation of Momentum During Impact", SAE Paper 95-0355
31. Based on personal experience of authors on MCAUTO time-share service in late 1970's
32. Hostetter, G.H., Santina, M.S., D'Carpio-Montalvo, P. Analytical Numerical and Computational Methods for Science and Engineering, Prentice Hall, Englewood Cliffs, NJ 1991, ISBN 0-13-026055-X
33. Forsythe, G.E., Malcolm, M.A., Moler, Computer Methods for Mathemetical Computations, C.B.,Prentice-Hall. Inc. Englewood Cliffs, NJ 1977, ISBN 0-13-165332-6
34. Etter, D.M., Fortran 77 with Numerical Methods for Engineers and Scientists, Benjamin/Cummings Publishing Company, Inc., 1994, ISBN 0-8053-1770-8
35. Press, W.H., Teukolsky, S.A.,Vetterling, W.T., Flannery, B.P., Numerical Recipes in Fortran. The Art of Scientific Computing ,Second Edition, Cambridge University Press, 1992 ISBN-0 521 43064 X
36. McHenry, R.R., McHenry, B.G., "A Revised Damage Analysis Procedure for the CRASH Computer Program", SAE Paper 86-1894
37. McHenry, R.R., McHenry, B.G., "Effects of Restitution in the Application of Crush Coefficients", SAE Paper 97-0960, to be published at the 1997 SAE International Congress and Exposition
38. Marquard, E.,"Progress in the calculations of Vehicle Collisions" Automobiletechnische Zeitschrift, Jarq. 68, Heft 3, 1966
39. Kahane, C.J, et al, "The National Crash Severity Study", Sixth International Technical Conference on Experimental Safety Vehicles (1976) 495-516
40. Fonda, A.G., Metz, L.D., "Post-Impact Spin, 1968-1993", SAE Paper 93-0653
41. Fonda, A.G., "Energy and Major Diversion in Accident Reconstruction", SAE Paper 96-0888

CONTACT INFORMATION

Questions or comments on the paper are welcomed and can be addressed to the authors by:

e-mail:	**mchenry@interpath.com**
Postal Service Mail:	**103 Brady Court** **Suite 200** **Cary, NC 27511 USA**
WWW:	**http://www.mchenrysoftware.com**

APPENDIX 1: DISCUSSION OF SPIN2

The SPIN2 procedure of the original CRASH program uses as a starting point the relationships developed by Marquard [**38**]. The Marquard procedure takes into account the fact that the linear and angular (i.e., yaw rotation) displacements of a four-wheeled vehicle subsequent to a collision each occur under conditions of intermittent deceleration when the wheels are free to rotate. By approximating the linear and angular deceleration rates of a vehicle with either (1) all wheels freely rotating or (2) all wheels locked during different phases of spinout motion, Marquard developed approximate relationships between the total linear and angular displacements during the travel from separation to rest and the corresponding linear and angular velocities of a vehicle at separation from its collision partner, for the two cited cases of rotational resistance.

In the CRASH program [1], the SPIN2 routine was developed to extend the relatively simple Marquard relationships to include the cases of partial braking and/or damage-locked individual wheels. Evaluations of the resulting, modified relationships by means of trial applications to spinout trajectories generated with SMAC [20] revealed several shortcomings of the initial SPIN2 relationships. First, a residual linear velocity frequently exists at the end of the rotational (i.e., yawing) motion. Next, the general shapes of plots of linear and angular velocity vs. time changed substantially as functions of the ratio of linear and angular velocity at separation from the collision. Finally, the transitions between the different deceleration rates of linear and angular motions were found to occur gradually rather than abruptly. Slope changes in the plots of linear and angular velocity vs. time were found to generally occur in the form of rounded "corners" in the curves.

To improve the accuracy of approximations of separation velocities, provisions for the introduction of a residual linear velocity at the end of the rotational motion and the development of empirical coefficients, in the form of polynomial functions of the ratio of linear to angular velocity at separation, were incorporated in the SPIN2 analytical relationships of the CRASH program. Since the separation velocity ratio is initially unknown, a solution procedure was

developed whereby several trial values of the ratio, based on an approximate equation, were used to test multiple solutions.

The cited analytical developments, reported in [2], involved only limited efforts which were aimed primarily at demonstrating the feasibility of the CRASH concept. Polynomial functions to generate empirical coefficients were developed, on the basis of 18 single-vehicle SMAC runs with relatively high linear and angular velocities for starting (i.e., separation) conditions. In the more common, real-life accident case, a relatively small rotation (i.e., yawing) velocity may exist at separation. In such a case the initial direction of the velocity vector with respect to the longitudinal axis of the vehicle will obviously affect the sequence and the duration of the linear and angular deceleration rates of the vehicle.

In consideration of known shortcomings of the SPIN2 aspect of the CRASH program, a subcontract to refine SPIN2 was undertaken in 1979 [24]. A representative sample of actual accident cases was selected from the NCSS [39] files for use in the study. A total of 50 cases were selected and then reconstructed with the SMAC computer program. For each of the SMAC reconstructions, separation information was used to formulate a basis for a refinement of the SPIN2 empirical coefficients.

A careful examination of the time-history plots of linear and angular velocities for all of the cases in the sample revealed a significant number of cases in which the SMAC-predicted behaviour deviated from the analytical assumptions upon which the SPIN2 routine is based. Attempts were undertaken within the research project to discriminate characteristics of separation conditions. Unfortunately, only partial success was achieved in the attempts to accommodate deviations by means of the use of logic and discriminators. As a result, a realistic appraisal of residual scatter in the empirical fits led to the conclusion in [24]:

"To achieve a general improvement in the reliability and accuracy of approximations of the angular and linear velocities at separation, a step-by-step time history form of trajectory solution should be implemented."

Subseqent work which has been performed on investigation and refinement of the SPIN empirical coefficients [30, 40, 41] and the corresponding modifications to CRASH is subject to the effects of 'scatter'. Any proposed refinements of the SPIN empirical coefficients and any reconstruction techniques which are based on the refinements of the SPIN empirical approach will ultimately fail in some applications to individual case reconstructions due to the possibililty that the particular case being investigated may be characteristic of a "scatter" point. The research cited in this paper strongly supports the conclusion from 1981 that implementation of a trajectory solution procedure should utilize an iterative time-history simulation.

APPENDIX 2: CRASH INPUTS

The following are the inputs for the ***refined CRASH3*** reconstructions used for comparison with the SMAC program. Note that most inputs (where applicable) are from the SMAC reconstructions.

The following questions are all answered the same and therefore not repeated for the inputs for the ***refined CRASH3*** test preliminary RICSAC test runs:

6. REST & IMPACT?.....................YES
9. ANY SLIP ANGLES?...................NO
11. SUSTAINED CONTACT?.................NO
13. SKIDDING STOP BEFORE REST?.........NO
15. CURVED PATH?.......................NO
18. MORE THAN 360 DEG? ....NO
20. SKIDDING STOP BEFORE REST?.........NO
22. CURVED PATH?.......................NO
25. MORE THAN 360 DEG? ....NO
26. TIRE-GROUND FRICTION............... .870
27. ROLLING RESISTANCE OPTION..........1
32. TRAJECTORY SIMULATION?.............YES
35. TERRAIN BOUNDARY?..................NO
38. DAMAGE DIMENSIONS?.................NO

In the input listings, the input specs were input to the ***refined CRASH3*** program in the following form:

SMAC: 4: A1,B1,TRW1,FMASS1,IZ1,XF1,XR1,YS1
SMAC: 5: A2,B2,TRW2,FMASS2,IZ2,XF2,XR2,YS2
SMAC 6: CSTF1, CSTF2, CSTF3, CSTF4
SMAC 7: CSTF5, CSTF6, CSTF7, CSTF8

where:

A,B = distances from CG to front/rear axles
TRW = track width
FMASS1 = vehicle mass, lb-sec**2/in
IZ1 = Total vehicle yaw moment of inertia, lb-sec-in
XF = distance from CG to front end
XR = distance from CG to rear end
YS = ½ width of vehicle
CSTF = cornering stiffness, lb/rad

The individual inputs are as follows:

File s1j95
1. TITLE..........sricsac1.bat test#1 7/15/95
2. CLASS/WEIGHTS...................... 4 4650. 2 3100.
3. CDC/PDOF # 1.......................12FDEW3 12.5
4. CDC/PDOF # 2.......................02RDEW3 71.9
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 2
7. REST COORDINATES.................. -.8 5.5 -.7 8.5 6.9 100.3
8. IMPACT COORDINATES................ -10.8 7.6 -30.0 .0 .0 90.0
12. SKIDDING OF # 1?..................YES
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?..................YES
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .019 .019 .340 .340
29. ROLL. RES, INDIV. WHEELS # 2....... .023 .023 .611 .611
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
35. TERRAIN BOUNDARY?.................NO
SMAC:4... 50.89 65.11 61.10 12.03 45019.0 90.99 -118.31 38.30
SMAC:5... 45.69 48.51 55.40 8.02 23676.0 83.19 -85.81 34.70
SMAC:6...-12338.0-12338.0-10991.0-10991.0
SMAC:7... -7547.0 -7547.0 -7108.0 -7108.0

File s2j95
1. TITLE.........sricsac2.bat test#2,7/26/95
2. CLASS/WEIGHTS...................... 4 4650. 2 3100.
3. CDC/PDOF # 1.......................11FDEW2 -17.3
4. CDC/PDOF # 2.......................02RDEW4 44.5
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 2
7. REST COORDINATES.................. 6.7 6.9 67.5 21.4 14.3 158.0
8. IMPACT COORDINATES................ -11.0 7.7 -30.0 .0 -.5 90.1
12. SKIDDING OF # 1?..................YES
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?..................YES
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .270 .019 .250 .250
29. ROLL. RES, INDIV. WHEELS # 2....... 1.000 .020 .400 .400
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
35. TERRAIN BOUNDARY?.................NO
SMAC:4... 52.45 63.55 61.10 12.03 45019.0 92.45 -116.05 38.30
SMAC:5... 46.75 47.45 55.40 8.02 23676.0 84.75 -84.45 34.70
SMAC:6...-12338.0-12338.0-10991.0-10991.0
SMAC:7... -7547.0 -7547.0 -9086.0 -9086.0

File s3j95.c15
1. TITLE.........sricsac3.bat test#3, 7/26/95
2. CLASS/WEIGHTS...................... 4 4980. 2 3140.
3. CDC/PDOF # 1.......................12FZEW1 6.2
4. CDC/PDOF # 2.......................06BZEW1 179.0
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 2
7. REST COORDINATES.................. 111.2 -5.0 -4.8 183.3 -10.0 -22.8
8. IMPACT COORDINATES................ .0 .0 .0 16.2 3.6 10.0
12. SKIDDING OF # 1?..................NO
17. ROTATION DIRECTION #1.............N/A
19. SKIDDING OF # 2?..................NO
24. ROTATION DIRECTION #2.............N/A
28. ROLL. RES, INDIV. WHEELS # 1....... .019 .019 .077 .077
29. ROLL. RES, INDIV. WHEELS # 2....... .020 .020 .107 .107
33. STEER ANGLES #1.................... -.5 -.5 .0 .0
34. STEER ANGLES #2.................... -1.5 -1.5 .0 .0
35. TERRAIN BOUNDARY?.................NO
SMAC:4... 55.65 62.32 63.45 12.89 58878.0 101.48 -116.32 39.65
SMAC:5... 37.32 56.78 55.40 8.13 22602.0 74.92 -94.08 34.70
SMAC:6...-12386.0-12386.0-11251.0-11251.0
SMAC:7... -8793.0 -8793.0 -5956.0 -5956.0

File s4j95.c15
1. TITLE.........sricsac4.bat test#4, 7/26/95
2. CLASS/WEIGHTS...................... 4 4980. 2 3190.
3. CDC/PDOF # 1.......................12FZEW2 8.5
4. CDC/PDOF # 2.......................06BDEW4 180.0
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 2
7. REST COORDINATES.................. 44.6 51.9 123.6 59.7 67.4 97.4
8. IMPACT COORDINATES................ .6 .0 .0 16.4 3.5 10.0
12. SKIDDING OF # 1?..................NO
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?..................NO
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .020 .019 .360 .360
29. ROLL. RES, INDIV. WHEELS # 2....... .020 .020 .460 .760
33. STEER ANGLES #1.................... 16.0 16.0 .0 .0
34. STEER ANGLES #2.................... 8.0 8.0 .0 .0
35. TERRAIN BOUNDARY?.................NO
SMAC:4... 55.90 62.10 63.45 12.89 56174.0 99.50 -115.90 38.40
SMAC:5... 35.70 58.80 55.40 8.26 22704.0 72.90 -96.10 33.25
SMAC:6...-12386.0-12386.0-11252.0-11252.0
SMAC:7... -8982.0 -8982.0 -5673.0 -5673.0

File s5j95.c15
1. TITLE.........sricsac5.bat test#5, 7/26/95
2. CLASS/WEIGHTS...................... 4 4600. 1 2530.
3. CDC/PDOF # 1.......................12FZEW1 .0
4. CDC/PDOF # 2.......................05BDEW8 170.0
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 1
7. REST COORDINATES.................. 257.4 7.6 2.0 67.2 42.4 275.0
8. IMPACT COORDINATES................ .6 .0 .0 15.0 3.1 10.0
12. SKIDDING OF # 1?..................NO
17. ROTATION DIRECTION #1.............N/A
19. SKIDDING OF # 2?..................YES
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .020 .020 .210 .210
29. ROLL. RES, INDIV. WHEELS # 2....... .000 .000 .410 .410
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 10.0 4.0
SMAC:4... 35.80 62.40 63.45 11.90 52177.0 99.20 -116.90 35.40
SMAC:5... 37.00 46.90 50.80 6.55 14000.0 68.00 -79.10 29.60
SMAC:6...-11615.0-11961.0 -8347.0 -8347.0
SMAC:7... -8560.0 -8560.0 .0 -6753.0

File s6j95
1. TITLE.........sricsac6.bat test#6, 7/26/95
2. CLASS/WEIGHTS...................... 4 4300. 2 2623.
3. CDC/PDOF # 1.......................11FZEW1 -20.0
4. CDC/PDOF # 2.......................01RDEW4 33.7
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 2
7. REST COORDINATES.................. 52.1 7.3 9.0 14.1 21.3 250.9
8. IMPACT COORDINATES................ .7 .0 .0 12.4 .8 120.0
12. SKIDDING OF # 1?..................NO
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?..................YES
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .010 .010 .200 .200
29. ROLL. RES, INDIV. WHEELS # 2....... .200 .200 .530 .020
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
SMAC:4... 50.50 60.50 62.00 11.20 35237.0 90.80 -113.70 38.30
SMAC:5... 40.30 54.60 56.00 6.83 15616.0 66.30 -88.67 30.50
SMAC:6...-11110.0-11110.0 -9266.0 -9266.0
SMAC:7... -7186.0 -7186.0 -5295.0 -5295.0

File s7j95
1. TITLE.........sricsac7.bat test#7, 7/26/95
2. CLASS/WEIGHTS...................... 4 4310. 2 2610.
3. CDC/PDOF # 1.......................12FDEW2 -3.0
4. CDC/PDOF # 2.......................02RDEW4 53.6
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 2
7. REST COORDINATES.................. 75.7 20.3 17.4 19.3 43.4 258.6
8. IMPACT COORDINATES................ .4 .0 .0 12.5 .4 120.0
12. SKIDDING OF # 1?..................NO
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?..................YES
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .010 .010 .230 .230
29. ROLL. RES, INDIV. WHEELS # 2....... .170 .170 1.000 .000
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
35. TERRAIN BOUNDARY?.................NO
SMAC:4... 50.50 60.50 62.00 11.20 46050.0 94.80 -113.70 38.30
SMAC:5... 39.70 55.30 56.00 6.75 15616.0 68.86 -90.44 30.50
SMAC:6...-11109.0-11109.0 -9265.0 -9265.0
SMAC:7... -7185.0 -7185.0 -5153.0 -5153.0

File s8j95
1. TITLE.........sricsac8.bat test#8, 7/26/95
2. CLASS/WEIGHTS...................... 4 4479. 4 4710.
3. CDC/PDOF # 1........................12FDEW2 -20.0
4. CDC/PDOF # 2........................03RYEW2 65.5
5. V#1 & V#2 STIFFNESS CATEGORIES..... 4 4
7. REST COORDINATES................... 1.7 7.8 28.5 1.0 23.0 145.0
8. IMPACT COORDINATES................. -10.8 1.7 .0 .0 .3 90.0
12. SKIDDING OF # 1?...................YES
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?...................YES
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .010 .010 .400 .400
29. ROLL. RES, INDIV. WHEELS # 2....... .010 .010 .350 .350
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
35. TERRAIN BOUNDARY?..................NO
SMAC:4... 50.50 60.50 62.00 11.59 47654.0 94.80 -113.70 38.30
SMAC:5... 50.50 60.50 62.00 12.20 50121.0 94.80 -113.70 38.30
SMAC:6...-11062.0-11062.0-10117.0-10117.0
SMAC:7...-12008.0-10590.0-11157.0-10779.0

File s9j95
1. TITLE.........sricsac9.bat test#9, 7/26/95
2. CLASS/WEIGHTS...................... 1 2256. 4 4900.
3. CDC/PDOF # 1........................11FDEW2 -31.0
4. CDC/PDOF # 2........................02RFEW2 59.5
5. V#1 & V#2 STIFFNESS CATEGORIES..... 1 4
7. REST COORDINATES................... .8 37.8 119.2 -1.0 48.2 136.5
8. IMPACT COORDINATES................. .0 .0 .0 8.5 -5.4 90.0
12. SKIDDING OF # 1?...................YES
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?...................NO
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .300 .300 .010 .010
29. ROLL. RES, INDIV. WHEELS # 2....... .120 .010 .180 .180
33. STEER ANGLES #1.................... 5.0 5.0 .0 .0
34. STEER ANGLES #2.................... 9.0 9.0 .0 .0
35. TERRAIN BOUNDARY?..................NO
SMAC:4... 37.17 49.83 50.80 5.88 12281.0 62.77 -84.13 29.60
SMAC:5... 54.56 63.94 63.45 12.76 58000.0 100.29 -117.50 39.65
SMAC:6... -5959.0 -5959.0 -5446.0 -5446.0
SMAC:7...-12193.0-12193.0-10406.0-10406.0

File s10j95
1. TITLE.........sricsac10.bat test#10,7/26/95
2. CLASS/WEIGHTS...................... 1 2306. 4 4720.
3. CDC/PDOF # 1........................10FDEW2 -65.0
4. CDC/PDOF # 2........................01RFEW2 25.0
5. V#1 & V#2 STIFFNESS CATEGORIES..... 1 4
7. REST COORDINATES................... 1.5 59.1 89.0 -2.2 117.3 126.3
8. IMPACT COORDINATES................. .0 .0 .0 8.4 -6.7 90.0
12. SKIDDING OF # 1?...................YES
17. ROTATION DIRECTION #1.............CW
19. SKIDDING OF # 2?...................NO
24. ROTATION DIRECTION #2.............CW
28. ROLL. RES, INDIV. WHEELS # 1....... .340 .340 .010 .010
29. ROLL. RES, INDIV. WHEELS # 2....... .180 .180 .240 .240
33. STEER ANGLES #1.................... -3.0 -3.0 .0 .0
34. STEER ANGLES #2.................... 5.0 5.0 .0 .0
35. TERRAIN BOUNDARY?..................NO
SMAC:4... 36.37 50.63 50.80 6.00 10000.0 61.42 -85.48 29.60
SMAC:5... 55.38 63.12 63.45 12.29 50000.0 101.80 -116.00 39.65
SMAC:6... -6168.0 -6168.0 -4807.0 -4807.0
SMAC:7...-11598.0-11598.0-11025.0-11025.0

File s11j95
1. TITLE.........sricsac11.bat test#11,7/26/95
2. CLASS/WEIGHTS...................... 2 3041. 4 4850.
3. CDC/PDOF # 1........................12FYEW3 -5.0
4. CDC/PDOF # 2........................12FYEW3 -13.0
5. V#1 & V#2 STIFFNESS CATEGORIES..... 2 4
7. REST COORDINATES................... 23.6 -7.5 167.0 7.1 .7 .4
8. IMPACT COORDINATES................. 15.4 -4.1 170.0 .3 .0 .0
12. SKIDDING OF # 1?...................YES
17. ROTATION DIRECTION #1.............N/A
19. SKIDDING OF # 2?...................YES
24. ROTATION DIRECTION #2.............N/A
26. TIRE-GROUND FRICTION............... .870
28. ROLL. RES, INDIV. WHEELS # 1....... .010 .010 .470 .470
29. ROLL. RES, INDIV. WHEELS # 2....... .010 .010 .185 .185
33. STEER ANGLES #1.................... -6.0 -6.0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
35. TERRAIN BOUNDARY?..................NO
SMAC:4... 49.45 47.50 55.20 7.92 22394.0 87.80 -84.40 32.70
SMAC:5... 54.40 63.60 63.45 12.63 56546.0 100.40 -117.40 39.65
SMAC:6... -6876.0 -6876.0 -7151.0 -7151.0
SMAC:7...-12056.0-12056.0-10314.0-10314.0

File s12j95
0 1. TITLE.........sricsac12.bat test#12,7/26/95
2. CLASS/WEIGHTS...................... 2 3130. 4 4512.
3. CDC/PDOF # 1........................12FYEW4 -7.0
4. CDC/PDOF # 2........................12FYEW4 -10.4
5. V#1 & V#2 STIFFNESS CATEGORIES..... 2 4
6. REST & IMPACT?.....................YES
7. REST COORDINATES................... 21.9 -9.4 140.6 7.6 2.4 -12.2
8. IMPACT COORDINATES................. 15.1 -4.4 170.0 .0 .0 .0
12. SKIDDING OF # 1?...................YES
17. ROTATION DIRECTION #1.............CCW
19. SKIDDING OF # 2?...................YES
24. ROTATION DIRECTION #2.............CCW
28. ROLL. RES, INDIV. WHEELS # 1....... .010 .440 .233 .233
29. ROLL. RES, INDIV. WHEELS # 2....... .010 .480 .240 .240
33. STEER ANGLES #1.................... 10.0 10.0 .0 .0
34. STEER ANGLES #2.................... 12.0 19.0 .0 .0
35. TERRAIN BOUNDARY?..................NO
SMAC:4... 47.20 49.58 55.20 8.15 23044.0 84.20 -88.00 32.70
SMAC:5... 55.62 62.38 63.45 11.75 52604.0 102.70 -115.10 39.65
SMAC:6... -7380.0 -7380.0 -7610.0 -7610.0
SMAC:7...-11002.0-11002.0-10628.0-10628.0

File s14j95
1. TITLE.........sricsac14.bat Texas A&M test#3 9/14/95
2. CLASS/WEIGHTS...................... 3 3820. 3 3820.
3. CDC/PDOF # 1........................11FDEW4 -20.0
4. CDC/PDOF # 2........................01FDEW5 40.0
5. V#1 & V#2 STIFFNESS CATEGORIES..... 3 3
7. REST COORDINATES................... 27.4 14.6-257.0 1.2 28.3 -80.0
8. IMPACT COORDINATES................. 5.2 -1.1 .0 8.4 -11.7 120.0
12. SKIDDING OF # 1?...................YES
17. ROTATION DIRECTION #1.............CCW
19. SKIDDING OF # 2?...................YES
24. ROTATION DIRECTION #2.............CCW
26. TIRE-GROUND FRICTION............... .750
28. ROLL. RES, INDIV. WHEELS # 1....... .020 .020 .400 .130
29. ROLL. RES, INDIV. WHEELS # 2....... .130 .250 .130 .130
33. STEER ANGLES #1.................... .0 .0 .0 .0
34. STEER ANGLES #2.................... .0 .0 .0 .0
35. TERRAIN BOUNDARY?..................NO
SMAC:4... 43.50 68.50 60.00 9.89 30255.0 84.00 -119.00 36.50
SMAC:5... 43.50 68.50 60.00 9.89 30255.0 84.00 -119.00 36.50
SMAC:6...-10200.0-10200.0-10200.0-10200.0
SMAC:7...-10200.0-10200.0-10200.0-10200.0

SAE #970949
Reviewer's Discussion
by Ric D. Robinette, Fay Engineering Corp.
CRASH 97 Refinement of the Trajectory Solution Procedure
Brian G. McHenry, Raymond R McHenry, Authors

The paper is a part of a long term continuous effort by McHenry Consultants, Inc. to develop improvements in the CRASH and SMAC family of programs. The basic intent is to modify CRASH by incorporating a post impact trajectory to rest iterative solution analysis methodology similar to the process currently utilized within SMAC. The objective is to remove at times misused assumptions and data selection questions from the hands of a user, and to replace the process with a more refined analysis process. The methodology is a reasonable extension to the CRASH and SMAC family of programs, and shows promise to potentially improve the CRASH based results. The primary question is how does this improved program version become available to the accident reconstruction community as a whole, because unless such improvements are made widely available they will be of a limited value.

SAE #970949
Reviewer's Discussion
by J. Rolly Kinney
CRASH-97 Refinement of the Trajectory Solution Procedure
Brian G. McHenry, Raymond R McHenry, Authors

This paper outlines helpful technical concepts that can improve the accuracy of the momentum collision model and the trajectory model contained in CRASH3 and possibly other algorithms that reconstruct an incident by beginning with the rest conditions and computing back in the event sequence to the impact conditions. However, the authors conclude the trajectory cannot be reconstructed with sufficient accuracy and reliability for some collision types and that the best generalized estimate for separation conditions is obtained by a trial and error simulation of the trajectory, for which they have developed an efficient iteration algorithm. That conclusion suggests reconstruction of some collision types is unreliable and that trajectory simulation should be used when improved accuracy for all type collisions is desired.

They also conclude that momentum reconstruction collision models must include angular momentum, tire forces, and movement during engagement in order to reasonably match the SMAC collision simulation model, which itself reasonably matches RICSAC results. Those who reconstruct collisions should incorporate those elements if they desire accurate results for all collision types.

For a real world collision the accuracy of the separation conditions established by any algorithm would be dependent on the wheel drag and steering angle values assumed. This paper did not discuss that aspect of the accuracy issue which probably cannot be evaluated until either the AIRP Trajectory Algorithm Comparison task group produces applicable results or until the refinements in this paper are applied in a blind analysis to instrumented tests.

The data shown in this paper demonstrate improved accuracy for the RICSAC test configurations but the rationale and values for some of the algorithm variable assumptions are not stated. Hopefully, the authors will detail those values and rationale in future presentations when their development is complete.

970951

A Comparison of Moment of Inertia Estimation Techniques for Vehicle Dynamics Simulation

Duane D. MacInnis, William E. Cliff, and Kurt W. Ising
MacInnis Engineering Associates

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The moments of inertia, in yaw, pitch, and roll, as well as the center of gravity height are necessary to successfully model the 3D dynamic behavior of vehicles before, during and after collision. A number of vehicle parameter estimation techniques have been developed and are currently in use in North America and Europe. Many parameters have been measured by NHTSA and others. The estimation techniques are compared to the available measured values, and recommendations are made for best estimating the parameters when measured values are not available. The sensitivity of 3D vehicle collision dynamics and trajectory simulation to variance in the moment of inertia is demonstrated.

INTRODUCTION

Early 2D collision and spin-out models required only the mass and the yaw moment of inertia. Modeling the 3D dynamic behavior of a vehicle requires a minimum of the vehicle geometry, the center of gravity height and the triaxial moments of inertia. Because moments of inertia are difficult to measure for a given vehicle, there is a need for estimation techniques that use easily obtained vehicle measurements.

More complex models will require data concerning the sprung and unsprung vehicle parameters, as well as more detailed suspension and tire modeling data. The work of this paper is limited to discussion of the whole vehicle roll, pitch, and yaw moments of inertia relative to the total vehicle centroidal axes, and the height of the whole vehicle center of gravity.

This paper briefly addresses the theory, and presents a number of estimation techniques that are available in the literature. The values measured by Garrott of the NHTSA are compared to the values determined by the estimation techniques in five different vehicle categories. Correction factors were applied to each of the estimation equations to eliminate the average error. The adjusted equations were ranked by standard deviation, and the best equations, the ones with the smallest standard deviation, were chosen and presented. The nomenclature used is described in Appendix A.

THEORY

In simple linear Newtonian mechanics, the mass of a vehicle is assumed to be concentrated at the center of gravity location. This assumption may be appropriate for the simplest dynamic models, pure central collisions, or for very coarse estimation purposes. When the vehicles are involved in eccentric collisions with rotation, as is the most common case, then the simple models are inadequate, and the distribution of the vehicle's mass about the centroid must be considered.

The resistance of an object to rotation is a function of the mass of the object, and the location of the mass with respect to the center of rotation, and is known as its moment of inertia, I.

An idealized motor vehicle can be considered a solid homogeneous slab as shown in Figure 1.

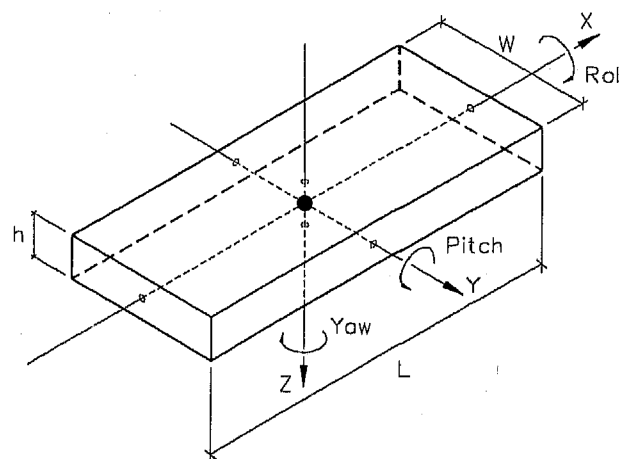


Figure 1 - The vehicle as a solid homogeneous slab or prism

The moments of inertia about the three axes of the homogeneous slab would be:[1]

$$I_{xx} = \frac{1}{12} m\left(W^2 + h^2\right) \qquad (1)$$

$$I_{yy} = \frac{1}{12} m\left(L^2 + h^2\right) \qquad (2)$$

$$I_{zz} = \frac{1}{12} m\left(L^2 + W^2\right) \qquad (3)$$

This is termed the prism method. The yaw moment of inertia for a solid slab is a function of the length and width as well as mass, not mass alone.

A motor vehicle is not a solid homogeneous slab. It has concentrated masses at the engine, transmission, suspension, etc., as shown in Figure 2.

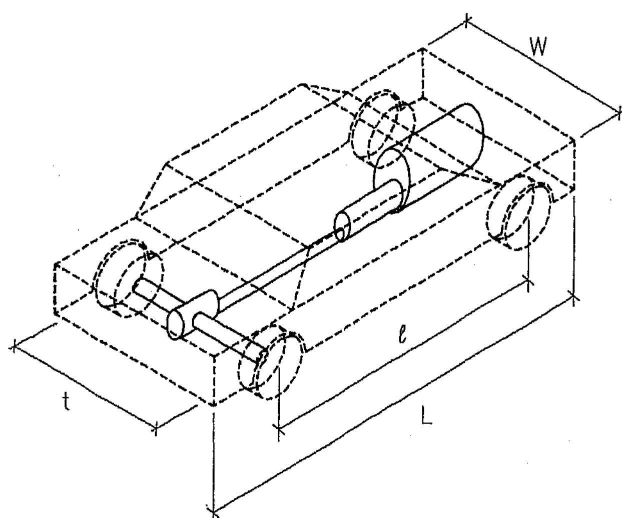


Figure 2 - The vehicle with concentrated masses

The moment of inertia for any object can be calculated, if enough is known about these distributed masses. However, the moment of inertia of a motor vehicle is more practically measured by testing.

A simplified top view of a vehicle is shown in Figure 3.

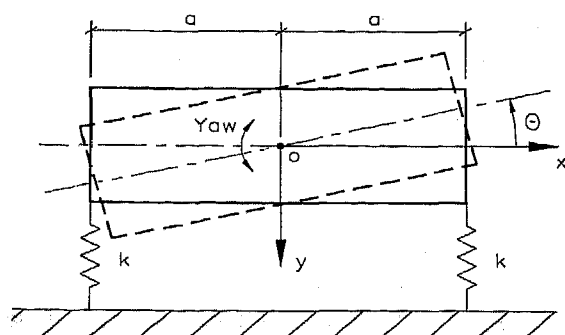


Figure 3 - Top view of a vehicle being tested for the yaw moment of inertia

The vehicle is allowed to pivot at its center of gravity, on a massless undamped frictionless frame. The fore/aft mass distribution is 50/50. The moments sum as follows:

$$\sum M = I\alpha \qquad (4)$$

$$-2(ka\theta)a = I\ddot{\theta} \qquad (5)$$

$$\ddot{\theta} + \left(\frac{2ka^2}{I}\right)\theta = 0 \qquad (6)$$

The period of this oscillating structure is: [2]

$$\tau = 2\pi\sqrt{\frac{I}{2ka^2}} \qquad (7)$$

In this idealized example, one measures the period, the geometry, and the spring constants, and the moment of inertia is readily determined:

$$I = \frac{1}{4\pi^2} 2ka^2\tau^2 \qquad (8)$$

The actual measurement of the inertial properties of vehicles is far more complex than presented here, and the details and complexities have been developed by Garrott.[3,4,5]

MEASURED VALUES

The NHTSA has measured the moments of inertia for a number of vehicles, as described by Garrott (1993). These values are published and available on disk.

Burg[6] (1982) in Germany tested a number of estimation methods against the measured values for 56 contemporary European vehicles. The original data has been lost.[7] The work of Rasmussen[8] (1970) at General Motors has been superseded by later researchers, and will not be included in this paper. Reide[9] (1984) tested a large number of passenger vehicles sold commercially "somewhere in the world." The data was not included in the published work. Curzon[10] (1991) measured the inertial properties of more than 100 passenger cars (not used in his study), sport utility vehicles, pickup trucks, and vans, but has not published the data. Bixel et al.[11] (1996)) measured the parameters for a total of 313 vehicles. The data was not included with their paper.

There has been work done by others as well, but the NHTSA work is the best available current source for measured values.

ESTIMATION TECHNIQUES

A number of methods have been proposed to estimate the moments of inertia, and in some cases the height of the center of gravity. These are listed here in chronological order.

Grime[12] (1969) proposed the prism method described previously. The work of Rasmussen[13] (1970) at General Motors has been superseded by later researchers, and will not be included in this paper.

mab -- The most commonly cited estimation of the yaw moment has been with us for some time. The original

source is uncertain. It has been recently cited and derived by Bastow[14]:

$$I_{zz} = mab \tag{9}$$

This method is intuitively satisfying in that it assumes the vehicle mass is concentrated at the center of the axles.

Burg -- Burg (1982) tested a number of methods against the measured values for 56 European vehicles and found that the method with the best correlation coefficient for the yaw moment for his sample was:

$$I_{zz} = 0.1269mlL \tag{10}$$

The Burg method has also been used by Reimpell.[15]

Reide -- Researchers at General Motors (1984) built on the earlier work by Rasmussen, testing a number of passenger vehicles "sold commercially somewhere in the world." No truck or bus types were included. The author describes the data as being applicable to moderate payload and driving conditions.

Body styles ranged from 2-passenger sports cars to station wagons, masses from 575 to 1761 kg, and wheelbases from 2025 to 2953 mm. Nine were front engine-rear drive, six were front engine-front drive, and two were rear engine-rear drive, for a total of 17 cars. They were tested in curb condition, no passengers and full gas tank.

The best fit for the whole vehicle moments of inertia in roll, pitch, and yaw, relative to the total vehicle centroidal axes, were:

$$I_{xx} = 0.37m - 86.4\,,\ R^2 = 0.893 \tag{11}$$

$$I_{yy} = 2.56m - 1{,}103\,,\ R^2 = 0.919 \tag{12}$$

$$I_{zz} = 2.86m - 1{,}315\,,\ R^2 = 0.920 \tag{13}$$

The sprung mass moments of inertia relative to the sprung mass centroidal axes were also given. No techniques for estimation of center of gravity heights were given.

Allen – Allen (1987) et al[16] proposed the rule of thumb values for the yaw moment as follows:

$$I_{zz} = mab \pm 10\% \tag{14}$$

The method proposed by Allen for the roll moment is a variant of the prism method:

$$I_{xx} = \frac{1}{12}m\left(W^2 + h^2\right) \tag{15}$$

Where,

$$h^2 = \alpha_1 h_1^{\,2} + \alpha_2 h_2^{\,2} \tag{16}$$

$$\alpha_1 + \alpha_2 = 1 \tag{17}$$

The values α_1 and α_2 shown on Figure 4 are weighting coefficients which consider the proportions of the vehicle at various heights. Allen suggests that this method of determining the roll moment is accurate to within ± 20%. However, the dimensional data to use this technique, α_1, α_2, h_1, and h_2 are not usually available in the tested or measured databases. These data are only available if the case vehicle or an exemplar is available to measure.

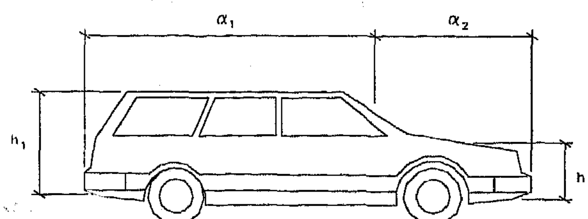


Figure 4 - The vehicle with Averaged Height

The center of gravity height is given as:

$$h_{cg} = 0.38h_{roof} \pm 5\% \tag{18}$$

Allen has offered no recommendations for the pitch moment.

Garrott -- The most complete listing of measured parameters was published by Garrott in 1988. In addition to the values measured, the authors presented a number of historic "rules of thumb", and proposed a set of revised rules of thumb (all units have been converted to metric) which are listed in Table 1.

Rules of Thumb:

$$h_{cg} = 533.4\,mm \tag{19}$$

$$h_{cg} = 0.40h_{roof} \tag{20}$$

$$I_{xx} = \frac{mt_{av}h_r}{4} \tag{21}$$

$$I_{yy} = mab \tag{22}$$

$$I_{zz} = I_{yy} \tag{23}$$

Table 1 - Garrott's Revised Rules of Thumb

Pass Cars	Light Trucks	
$h_{cg} = 540.8 \pm 38.1\,mm$	$h_{cg} = 678.4 \pm 101.6\,mm$	(24)
$h_{cg} = 0.395h_{roof} \pm 2.6\%$	$h_{cg} = 0.387h_{roof} \pm 3.5\%$	(25)
$I_{xx} = (0.73 \pm 0.13)\frac{mt_{av}h_r}{4}$	$I_{xx} = (0.67 \pm 0.16)\frac{mt_{av}h_r}{4}$	(26)
$I_{yy} = (1.07 \pm 0.17)mab$	$I_{yy} = (1.04 \pm 0.22)mab$	(27)
$I_{zz} = (1.03 \pm 0.08)I_{yy}$	$I_{zz} = (1.00 \pm 0..1)I_{yy}$	(28)

Garrott also fitted curves through the measurement data, and presented the moment of inertia parameters for passenger cars and light trucks as a function of vehicle mass.

Passenger Cars:

$$I_{xx} = 0.538m - 203,\ R^2 = 0.80 \qquad (29)$$

$$I_{yy} = 2.96m - 1{,}558,\ R^2 = 0.89 \qquad (30)$$

$$I_{zz} = 3.08m - 1{,}635,\ R^2 = 0.88 \qquad (31)$$

Light Trucks:

$$I_{xx} = 0.66m - 319,\ R^2 = 0.70 \qquad (32)$$

$$I_{yy} = 3.35m - 2{,}247,\ R^2 = 0.70 \qquad (33)$$

$$I_{zz} = 3.08m - 1{,}821,\ R^2 = 0.73 \qquad (34)$$

Curzon -- Later researchers developed the work of Garrott further. Curzon and colleagues (1991) measured the inertial properties of sport utility vehicles, pickup trucks, and vans, and developed a number of moment of inertia estimation techniques applicable to most light trucks. The authors devised a number of estimation equations and tested them against the sum total of light trucks tested by their company and also those tested by Garrott.

They found the prism method was the best estimate for the yaw moment for sport utility vehicles, using the track width instead of the vehicle width.

$$I_{zz} = \frac{1}{12} m\left(L^2 + t^2\right) \qquad (35)$$

For the yaw moment of pickup trucks, the rule of thumb was best:

$$I_{zz} = mab \qquad (36)$$

For the yaw moment for vans, they recommend either the prism method as for sport utility vehicles, or Garrott's value for light trucks:

$$I_{zz} = 3.08m - 1{,}821 \qquad (37)$$

The roll moment was best approximated by a sum of prisms, similar to the method first proposed by Allen, that account for the varying cross sections of the vehicles, as shown in Figures 5, 6, and 7.

The recommended relationship for the roll moment for sport utility vehicles is as follows:

$$I_{xx} = \frac{1}{12} m\left(\frac{L_r h_r^2}{L} + W^2\right) + \frac{1}{12} m\left(1 - \frac{L_r}{L}\right)(h_w - h_a)^2 \qquad (38)$$

Like the earlier work by Allen, the roof length is required to use this technique. The roof length is not usually available in the tested or measured databases. This data is only available if the case vehicle or an exemplar is available to measure.

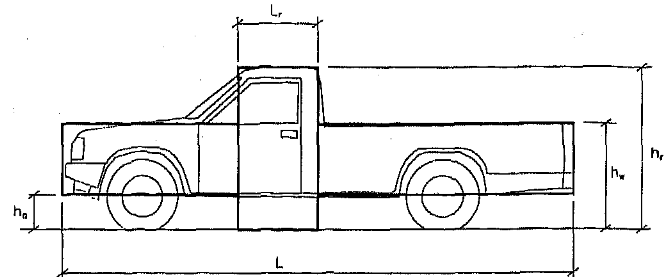


Figure 5 - Prism Models for Pickup Truck

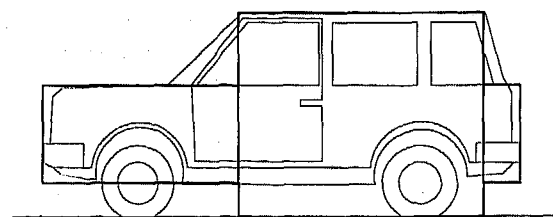


Figure 6 - Prism Models for Sport Utility

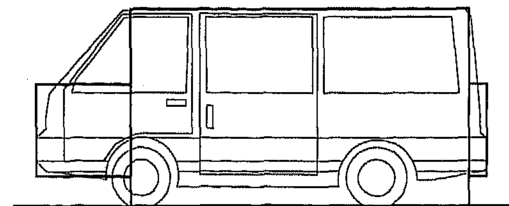


Figure 7 - Prism Models for Van

For the roll moment for vans, Curzon recommends either the method above for sport utility vehicles, or Garrott's value for light trucks:

$$I_{xx} = 0.66m - 319 \qquad (39)$$

Curzon offered no recommendations for the pitch moment or the center of gravity height.

Noon -- Noon[17] (1994) has proposed three methods to estimate the yaw moment. These methods for convenience have been called Noon-1, Noon-2, and Noon-3. The first method has the moment being a function of the mass and the overall length:

$$I_{zz} = \frac{mL^2}{12} \quad \text{Noon-1} \qquad (40)$$

The second has the moment being a function of the mass and the wheelbase alone, assuming the masses are concentrated at half the wheelbase from the geometric center:

$$I_{zz} = \frac{ml^2}{4} \quad \text{Noon-2} \quad (41)$$

Noon's third method, considering the engine to be the most significant concentrated mass affecting the application of the prism method, accounts for the effect of the engine mass with a combination of methods one and two:

$$I_{zz} = 2(1-x)\frac{mL^2}{12} + (2x-1)m\frac{l^2}{4} \quad \text{Noon-3} \quad (42)$$

where,

x = fraction of vehicle mass borne by front axle

This third method assumes that the difference between the weights on the axles is due to a point mass at the heavier axle, and that the centroid is at the geometric center. Although contradictory, these assumptions are justified for an estimation technique.

Noon offered no recommendations for the pitch or roll moments or the center of gravity height.

Bixel – Bixel (1996) measured the parameters for 104 passenger cars, 84 multi-purpose vehicles, 82 pickup trucks, and 43 vans,.and recommended a number of estimation techniques. They found the roll moment for all vehicles can be estimated as:

$$I_{xx} = \frac{m(h_{roof} + h_{cg})t}{K} \quad (43)$$

Where K is an approximation constant determined for each class of vehicle.

The pitch moment for passenger cars can be estimated as:

$$I_{yy} = \frac{m(h_{roof} + h_{cg})L}{K} \quad (44)$$

The pitch moment for multi-purpose vehicles, pickup trucks and vans can be estimated as:

$$I_{yy} = \frac{m(h_{roof} + h_{cg})l}{K} \quad (45)$$

The yaw moment for all vehicles can be estimated as:

$$I_{zz} = \frac{mtl}{K} \quad (46)$$

All of the methods proposed by Bixel except yaw use the center of gravity height, a quantity that is not generally available on published databases.

The Bixel K factors are summarized in Table 2. Bixel notes that these approximation constants must evolve with evolving vehicle design.

Table 2 - Bixel's Approximation Constants (K)

Vehicle Class	Pitch	Roll	Yaw
Passenger Cars	5.2901	7.9846	2.1942
Sport Utility, l<2.41 m	4.2193	9.4212	2.2048
Sport Utility, l>2.41 m	3.4510	9.4212	2.2048
Pickups	3.3783	9.4738	2.1858
Vans	3.4734	7.8854	2.2168

COMPUTER SIMULATION PROGRAMS

A number of computer programs are available to simulate the behavior of motor vehicles in collision and spin-out. The most common products are the crush-energy based progeny of the original Crash III program, EDCRASH[18] and SLAM[19], and EDSMAC[20] based on the early work of McHenry.[21] More recently PC-Crash[22], a collision and trajectory program developed by Dr. Steffan,[23] has become available.

EDCRASH and EDSMAC use a default value for the yaw moment of inertia derived from the vehicle class categories. These values can be altered by the user.

WinCRASH, the Windows version of the SLAM program, uses a default value for the radius of gyration, k, that can be used to calculate the yaw moment:

$$k_{zz} = (1.023 \pm 0.085)\sqrt{ab} \quad \text{Slam-1} \quad (47)$$

Alternatively WinCRASH proposes:

$$k_{zz} = (0.298 \pm 0.030)L \quad \text{Slam-2} \quad (48)$$

When expressed as the moment of inertia, the WinCRASH values are used as follows:

$$I_{zz} = mk^2 \quad (49)$$

The WinCRASH default value can be altered by the user.

PC-Crash uses the method of Burg for the default yaw moment of inertia for passenger cars:

$$I_{zz} = 0.1269mlL \quad (50)$$

PC-Crash uses the roll and pitch moments as functions of the Burg yaw moment:

$$I_{xx} = 0.3I_{zz} \quad (51)$$

$$I_{yy} = I_{zz} \quad (52)$$

For trucks and trailers, PC-Crash recommends the prism method. Because the height of a truck is similar to the width, PC-Crash reduces the roll, pitch, and yaw moments to:

$$I_{xx} = \frac{2}{12}mW^2 \quad (53)$$

$$I_{yy} = \frac{1}{12} m(L^2 + W^2) \quad (54)$$

$$I_{zz} = I_{yy} \quad (55)$$

The PC-Crash default values can be altered by the user.

COMPARISON OF MEASURED VALUES WITH ESTIMATED VALUES

The NHTSA has measured the moment of inertia for a large number of vehicles, as described by Garrott. These values, for about 414 tests, are published and available on disk. Many of the tests were performed with multiple occupants and significant ballast to load the vehicles to the GVW. About 215 of the tests with multiple occupants or ballast appear to be repeats of tests done on the same vehicle with one or no occupants. Therefore, vehicles loaded with more than one occupant and those carrying loads have been eliminated from the samples used here. Also eliminated were multiple tests involving identical vehicles so as not to weight the sample set toward any particular vehicle.

The comparisons have been done by dividing the vehicles into five categories, as shown on Table 3, where n is the number of vehicles in a category sample. In each case the measured pitch, roll, and yaw moments of inertia and the center of gravity height are compared with the estimated values. The estimation methods compared are given in Table 4.

Table 3 - Vehicle categories for comparison

Vehicle Category	n
Passenger Car - front wheel drive	46
Passenger Car - rear wheel drive	24
Sport Utility (Multi-Purpose)	22
Pickup Truck	25
Van	17

The estimated/measured values for the 46 front wheel drive passenger cars was tested to confirm it had a nearly normal distribution. Although some of the data is skewed, the mean, median, and mode are generally in agreement.

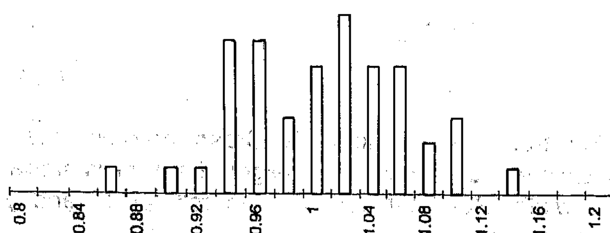


Figure 8 - Histogram of Estimated/Measured Values for FWD Passenger Cars

A histogram for the prism pitch data is as shown on Figure 8. This data set had a mean of 1.00, mode of 0.94, median of 1.01 and a skew of -0.04.

The distribution of the errors about the median shown on Figure 8 is approximately bell-shaped, thus standard statistical tests of the quality of the data will be used.

Table 4 - Estimation Methods Compared

Estimation Method	I_{xx}	I_{yy}	I_{zz}	h_{cg}
Prism	✓	✓	✓	
Burg/PC Crash	✓	✓	✓	
Reide	✓	✓	✓	
Allen				✓
Garrott	✓	✓	✓	
Rule of Thumb	✓	✓	✓	✓
Garrott's Revised Rule of Thumb	✓	✓	✓	✓
Rule of Thumb 2				✓
Garrott's Revised Rule of Thumb 2				✓
Noon 1			✓	
Noon 2			✓	
Noon 3			✓	
Bixel	✓	✓	✓	
Slam 1			✓	
Slam 2			✓	

For each vehicle in the comparison category, the estimated value has been compared to the measured value in the following way. First the individual error, E, of each estimated value compared to its measured value is calculated as:

$$E = \frac{V_e}{V_m} - 1 \quad (56)$$

Where:

V_e = estimated value

V_m = measured value

The average error, E_{ave}, is:

$$E_{ave} = \frac{\pm \sum_{1}^{n} E}{n} \quad (57)$$

Where:

n = number of vehicles in category sample

The range, R, of the errors is:

$$R = E_{max} - E_{min} \quad (58)$$

Where:

E_{max} = maximum positive error in sample category

E_{min} = minimum positive error in sample category

The standard deviation, SD, is:

$$SD = \sqrt{\frac{1}{n-1}\sum\left(E - E_{ave}\right)^2} \quad (59)$$

The 68% confidence limit, L_{68}, is

$$L_{68} \cong E_{ave} \pm SD \quad (60)$$

The 95% confidence limit, L_{95}, is;

$$L_{95} \cong E_{ave} \pm 2SD \quad (61)$$

All the estimation methods were adjusted with a correction factor as follows:

$$V_{Correct} = CV_e \quad (62)$$

Where the correction factor, C, is:

$$C = \frac{1}{E_{ave} + 1} \quad (63)$$

The adjusted equations were then compared, and the best estimation equations were chosen based on the smallest standard deviation. In the event of a tie, the equation with the narrowest range was chosen.

Figure 9 shows the unadjusted values for the yaw moment of inertia for rear wheel drive passenger vehicles. The equation numbers are shown in brackets.

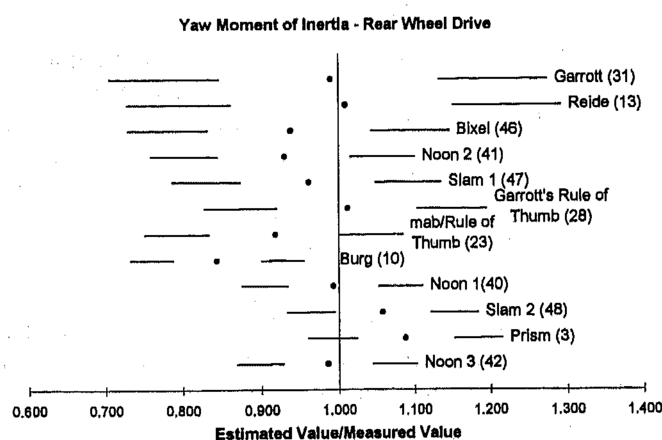


Figure 9 - Unadjusted Estimates

Figure 10 shows the adjusted values for the yaw moment of inertia for rear wheel drive passenger vehicles. The error bands on each graph connect the standard deviation to the 95th percent confidence limit, as on Figure 9. Note that the error bands for each equation have been adjusted as well.

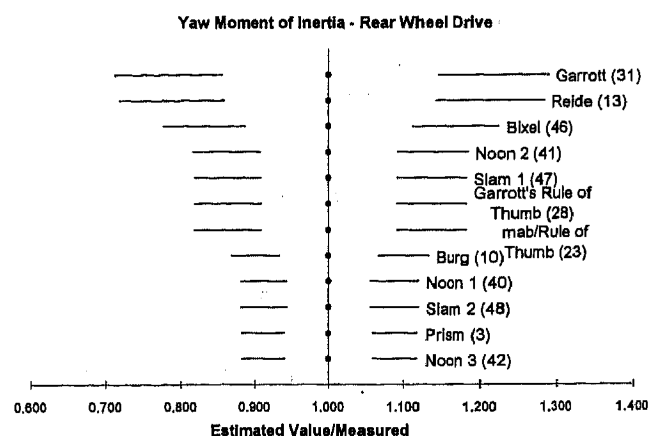


Figure 10 - Adjusted Estimates

RECOMMENDED ESTIMATION TECHNIQUES

The tables in Appendix B provide the correction factors applied to each of the estimation equations as well as statistical information indicating their accuracy. The equation numbers for each estimation technique are shown in brackets.

Each table is accompanied by a chart which graphically ranks the equations as described previously. The equations providing the smallest range of standard deviation error (68% confidence limit) are given in Tables 5 to 9. Note that these equations have been multiplied through by their individual correction factors.

If the information required to use a recommended equation is not available, another equation may be used with its correction factor taken from the tables in Appendix B. The user must then be aware of the slightly greater error being introduced.

When adjusted, it was found that Garrott's revised rules of thumb provided identical results to the original rules of thumb. This was expected as Garrott's revised rules are the result of multiplying the original rules by a constant. Allen's adjusted equation for center of gravity height also provided identical results to those of the second rule of thumb (equation 20) for the same reason.

The inertia estimation methods proposed by Garrott and Reide had greater deviations than the other methods. Both the Reide and Garrott methods consider the moments of inertia to be linearly related to the vehicle mass and do not consider the vehicle geometry.

The only available body of moment of inertia data for European vehicles has been given in the Burg study. No guidance can be offered for those European vehicles not included in the NHTSA data used for this comparison study.

Table 5 - Best Equations for Front Wheel Drive Passenger Cars

	Adjusted Equation
I_{xx}	$0.1274m(h_{roof} + h_{cg})t \pm 6.5\%$
I_{yy}	$0.1425mlL \pm 5.7\%$
I_{zz}	$0.1478mlL \pm 4.8\%$
h_{cg}	$535.0\ mm \pm 3.0\%$

Table 6 - Best Equations for Rear Wheel Drive Passenger Cars

	Adjusted Equation
I_{xx}	$0.120m(h_{roof} + h_{cg})t \pm 9.5\%$
I_{yy}	$0.0733m\left(L^2 + h^2\right) \pm 5.2\%$
I_{zz}	$1.015\left[2(1-x)\frac{mL^2}{12} + (2x-1)m\frac{l^2}{4}\right] \pm 5.9\%$
h_{cg}	$0.390h_{roof} \pm 4.9\%$

Table 7 - Best Equations for Sport Utility Vehicles

	Adjusted Equation
I_{xx}	$0.145mt_{av}h_r \pm 7.6\%$
I_{yy}	$0.1561mlL \pm 5.5\%$
I_{zz}	$0.4622mtl \pm 6.7\%$
h_{cg}	$0.390h_{roof} \pm 5.4\%$

Table 8 - Best Equations for Pickup Trucks

	Adjusted Equation
I_{xx}	$0.138mt_{av}h_r \pm 7.4\%$
I_{yy}	$0.952mab \pm 5.2\%$
I_{zz}	$0.958mab \pm 4.6\%$
h_{cg}	$0.376h_{roof} \pm 3.5\%$

Table 9 - Best Equations for Vans

	Adjusted Equation
I_{xx}	$0.1212m(h_{roof} + h_{cg})t \pm 8.0\%$
I_{yy}	$0.1508mlL \pm 7.4\%$
I_{zz}	$0.1525mlL \pm 7.4\%$
h_{cg}	$0.381h_{roof} \pm 2.8\%$

EFFECT OF VEHICLE LOADING

The comparisons have been done with the NHTSA tests of unloaded vehicles (no more than one occupant). If the passenger and loading mass and distribution is significant, the estimated values must be adjusted.

Occupant positioning in a passenger car is relatively close to the radius of gyration about the centroid, thus the increase in the moments of inertia due to the passenger loading will generally be proportional to the increase in the mass.

If there is a known concentrated mass in a vehicle, due to occupant positioning or cargo loading remote from the radius of gyration, then the moments can be summed using the parallel axis theorem:

$$I' = I + mr^2 \qquad (64)$$

Arndt[24] has identified the effect of passenger and cargo loading on the inertial properties, and provided methods of adjusting the estimated (or measured) moments for specific loading conditions.

SENSITIVITY ANALYSIS

The pitch and roll behavior of the vehicles in 3D simulation will be affected by the respective moments during collision and once the vehicles have separated. The inertial parameter with the greatest potential to influence a simulation is the yaw moment. Variations in the yaw moment of inertia will affect the forces developed between the two vehicles in collision and the immediate post-impact spin rate. When the vehicles have separated, the angular acceleration will be governed to a great extent by the yaw moment. A greater yaw moment of inertia will reduce the yaw angular acceleration when compared to the case of lesser yaw moment.

A test case has been selected to demonstrate the effect of variations in the yaw moment of inertia. The case uses a highly eccentric collision to maximize the sensitivity of the model to the yaw moment. The case uses a 1991 Ford Mustang and a 1989 Honda Accord, both arbitrarily chosen. The vehicle parameters used in PC-Crash are shown on Table 10.

The eastbound Mustang struck the right side at the rear of the southbound Accord, with both vehicles traveling at 50 km/h. The baseline run arbitrarily uses the moments of inertia calculated with the Burg method.

Both vehicles were permitted to spin out to rest. For the Mustang, the front wheels were locked at the instant of collision, and the rear wheels set to a drag of 5%. The Accord rear wheels were locked at impact, and the front wheels set at 5% drag. There was no steering introduced to simulate rolling damaged wheels. The force location at maximum engagement was set at the center of the overlap area (45 ms after contact), and the contact plane was set parallel to the longitudinal axis of

the Accord. The collision coefficient of restitution was set at 0.1, and the inter-vehicle coefficient of friction was set at 1.00.

Table 10 - Test Vehicles

Parameter	Description	V1	V2
	Year & Make	91 Ford	89 Honda
	Model	Mustang	Accord
L	Length, m	4.56	4.57
W	Width, m	1.73	1.71
l	Wheelbase, m	2.55	2.60
h	Height, m	1.32	1.35
t	Track width, m	1.42	1.48
m	mass, kg	1270	1210
a	CG to front axle, m	1.07	0.96
I_{xx}	roll, kg-m^2	562	547
I_{yy}	pitch, kg-m^2	1874	1824
I_{zz}	yaw, kg-m^2	1874	1824
h_{cg}	CG Height	0.53	0.54
V	Speed, km/h	50	50

These simulation parameters are not intended to be representative of any particular vehicle or collision circumstance. The case was selected to illustrate the sensitivity of the collision to variance of the one variable, the yaw moment of inertia.

The yaw moment was varied from 70% to 130% of the default value. The altered values for the yaw moment of inertia are as shown in Table 11.

Table 11 - Altered Yaw Moment Values, kg-m^2

Description	Mustang	Accord
70% Value	1312	1277
80% Value	1499	1459
90% Value	1687	1642
100% Value	1874	1824
110% Value	2061	2006
120% Value	2249	2189
130% Value	2436	2371

First, the simulation was run with the default values. This run is shown graphically in Figure 11. The tire marks are drawn to simulate the actual tire marks assuming tire marks are visible if the tires are 95% saturated.

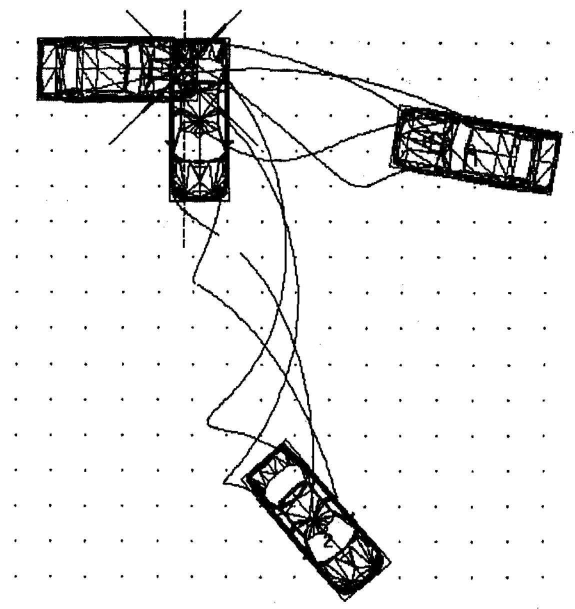


Figure 11 - Simulation with Default Yaw Moments

Next, runs were done varying only the yaw moment of inertia. The run with the yaw moment set to 130% is shown in Figure 12. The DXF drawing showing the rest positions and the tire marks for the run at 100% is plotted for comparison. The 100% run rest positions are shown as rectangles beneath the vehicles. The remainder of the runs are shown in Appendix C.

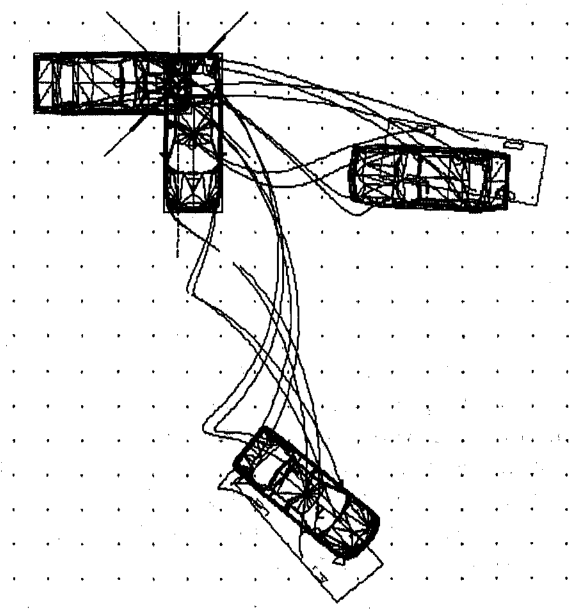


Figure 12 - Simulation with Yaw Moments increased to 130%

Finally, the 130% run was redone, varying the impact speeds and the orientation at impact to duplicate as closely as possible the rest positions for the default moments. The impact speeds were adjusted to 51.5 km/h (+3%) on the Mustang and 51 km/h (+2%) for the Accord and the Mustang was repositioned about 22 cm north at impact. This run is shown on Figure 13.

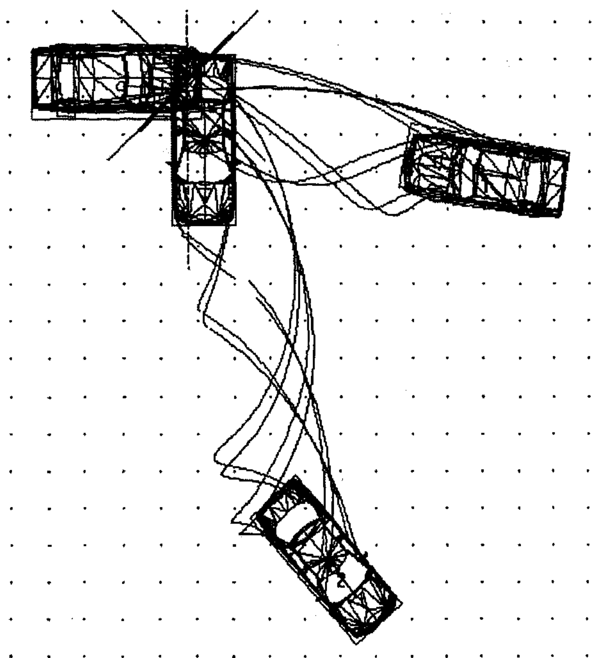


Figure 13 - 130% Simulation varied to match rest positions for 100% case

Note that although the rest positions are similar to the default case in Figure 11, the tire traces are quite different.

Runs were done to test the sensitivity of the simulation to variations in the pitch and roll moments, and the centroid height. As expected, the simulation was not very sensitive to variation in these parameters.

The simulations were very sensitive to variations in the wheel lockup factors, and steer angles on individual tires. The rest positions were more sensitive to these factors than to the yaw moments.

SUMMARY

The values measured by Garrott of the NHTSA have been compared to the values determined by the estimation techniques in five different vehicle categories. Correction factors were applied to each of the estimation equations to eliminate the average error. The adjusted equations were ranked by standard deviation, and the best equations, those with the smallest standard deviation, were chosen and presented.

The recommended equations will predict the moments of inertia and the centroid height parameters of the Garrott sample to better than ±10% with 68% confidence. The 95% confidence limit for the yaw moment was less than ±10% for front wheel drive cars and pickups, and less than ±15% for the other vehicle categories.

In some cases the differences between the recommended equations and the next best equations were very small. Where the dimensional data necessary to estimate the parameters with a recommended equation are not available, alternative equations are presented, with the expected errors given.

The inertia estimation methods proposed by Garrott and Reide consider the moments of inertia to be linearly related to the vehicle mass and do not consider the vehicle geometry. These methods generally produced greater errors than the other estimation methods.

Simulations of a highly eccentric collision and spin-out with PC-Crash have shown that if the estimated values for the yaw moment of inertia are within about ±10% of the actual values, the impact speeds in the simulation will vary less than about 1%. Variations in pitch and roll moments have been shown to have little influence on the calculated vehicle speeds for the simulated 3D planar cases.

If the selected yaw moment of inertia is greater than the actual value by about 30%, the simulation will tend to overestimate the collision speed of the vehicles by as much as 3%. If the selected yaw moment of inertia is less than the actual value by about 30%, the simulation will tend to underestimate the collision speed.

If the estimation method has 95% confidence limits greater than ±30% for the yaw moment, as could occur if the estimation method is based on mass alone, then the simulation errors could become significant.

The simulations indicate the rest positions of vehicles are more sensitive to wheel lockup factors and steer angle than to variations in moments of inertia.

The idealized vehicle prism method and "mab" rule of thumb have often yielded very good results.

This work is based on the measurement data of Garrott and the NHTSA. Others, notably Curzon and Bixel, have measured a number of vehicles. If these data were available, the results may vary.

APPENDICES

The following appendices are attached:

A. Nomenclature

B. Tables and charts of adjusted equations

C. PC-Crash runs

REFERENCES

[1] Meriam, J. L., Dynamics, John Wiley & Sons, 1966

[2] Thompson, William T.; Vibration Theory and Applications, Prentice Hall, 1965.

[3] Garrott, W. Riley,; "Inertial Parameters of Selected 1988 Four Wheel Drive Utility Vehicles", Project VRTC-88-0087, Report #1 NHTSA

[4] Garrott, W. Riley, et al, "Vehicle Inertial Parameters - Measured Parameters and Approximations", SAE881767

[5] Garrott, W. Riley; "Measured Vehicle Inertial Parameters - NHTSA's Data Through September 1992", SAE930897

[6] Burg, Heinz, Approximation von Tragheitsmomenten bei Personenkraftwagen, March 1982

[7] Private communication with Dr. H. Steffan, Linz, Austria, who has recently contacted Mr. Burg on this subject.

[8] Rasmussen, R. E.; et al.; "Typical Vehicle Parameters for Dynamic Studies", General Motors Proving Ground, A-2542, April 1970

[9] Reide, Peter M., et al; "Typical Vehicle Parameters for Dynamic Studies Revised for the 1980's",SAE840561

[10] Curzon, A. M., et al, "Light Truck Inertial Properties", SAE910122

[11] Bixel, Ronald A.;Heydinger, Gary J.; et al; "Developments in Vehicle Center of Gravity and Inertial Parameter Estimation and Measurement" SAE960183

[12] Grime, G.; Jones, I. S.; "Car Collisions - The Movement of Cars and their Occupants in Accidents", Proc. I. Mech. E., 1969-70, Vol 184 Pt 2A No 5,

[13] Rasmussen, R. E.; et al.; "Typical Vehicle Parameters for Dynamic Studies", General Motors Proving Ground, A-2542, April 1970

[14] Bastow, Donald, Car Suspension and Handling, Pentech Press, London, 1993

[15] Reimpell, Jornsen, Stoll, Helmut, The Automotive Chassis: Engineering Principles, Society of Automotive Engineers, 1996

[16] Allen, R. Wade, et al, "Steady State and Transient Analysis of Ground Vehicle Handling", SAE870495

[17] Noon, Randall K., Engineering Analysis of vehicle Accidents, CRC Press, 1994

[18] © Engineering Dynamics Corporation, Beaverton, OR.

[19] © Trantech Corporation, Bellevue, WA.

[20] © Engineering Dynamics Corporation, Beaverton, OR.

[21] McHenry, Raymond R., "Computer Program for Reconstruction of Highway Accidents, SAE730980

[22] PC-CRASH is a Windows©-based accident-reconstruction program developed at the Technical University of Graz, Austria, by Hermann Steffan, Andreas Moser, and Wolfgand Neubauer.

[23] Steffan, Hermann; Moser, Andreas; "The Collision and Trajectory Models of PC-Crash", SAE960886

[24] Arndt, Mark W., et al, "Effects of Passenger and Cargo Loading on a Motor Vehicle's Mass Properties", SAE952676

APPENDIX A -- NOMENCLATURE

The following nomenclature will be used throughout:

I = moment of inertia
L = overall vehicle length
l = wheelbase
W = overall vehicle width
h = height
t = track width
a = distance center of mass to front axle
b = distance center of mass to rear axle
T = torque
ω = angular velocity
α = angular acceleration
τ = period of oscillation
k = spring constant
or radius of gyration
r = radius

Subscripts have been used as appropriate, and the meaning will generally be self-evident. The nomenclature of some cited authors has been altered for consistency. Where necessary, the units used in older cited works have been converted to metric.

Table A1 - Metric/US Conversions

1 lb force	4.448 N
1 in	25.4 mm
1 lb-ft-sec^2	1.356 kg-m2

The SAE frame of reference shown in Figure A1 is used:

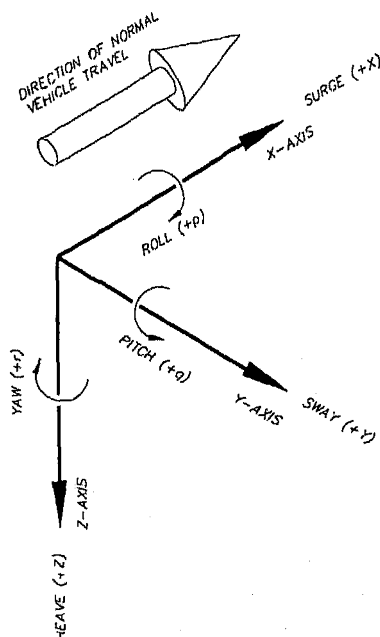


Figure A1 - The SAE frame of reference

Another common frame of reference, shown in Figure A2, has the positive direction for the Z axis directed upwards. The standard instruction in engineering schools, for example, has the X axis pointing to the right with angles about Z measured counterclockwise from the X-axis.

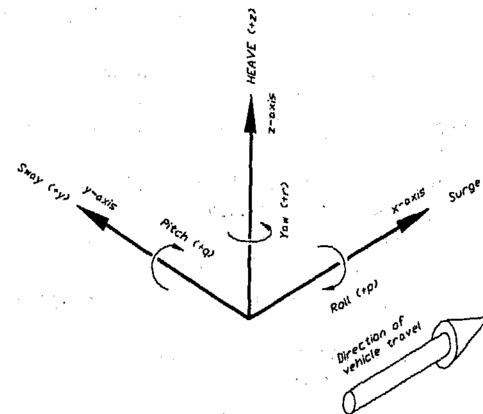


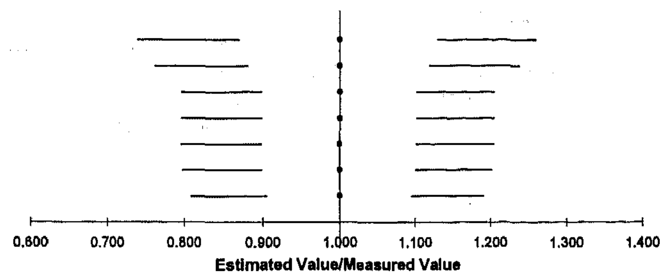
Figure A2 - A Common frame of reference

For both frames the positive angles are defined using the right hand rule.

APPENDIX B - ACCURACY OF CORRECTED EQUATIONS

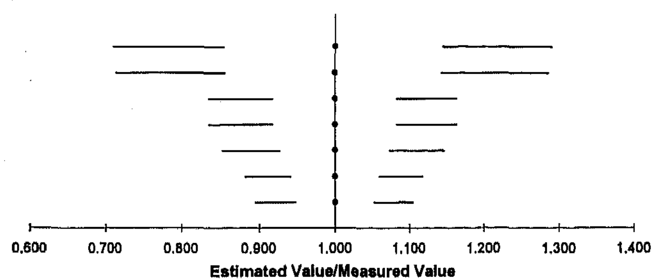
Rear Wheel Drive Passenger Cars

A) Roll Moment of Inertia



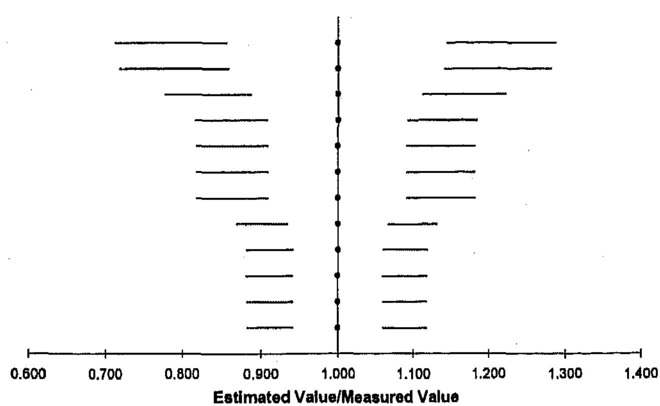
Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (11)	1.092	0.520	0.130	0.260
Garrott (29)	0.872	0.486	0.119	0.238
Prism (1)	1.000	0.497	0.102	0.204
Garrott's Rule of Thumb (26)	0.916	0.490	0.102	0.204
Rule of Thumb (21)	0.669	0.490	0.102	0.204
PC Crash (51)	0.717	0.440	0.101	0.202
Bixel (43)	0.960	0.433	0.095	0.191

B) Pitch Moment of Inertia



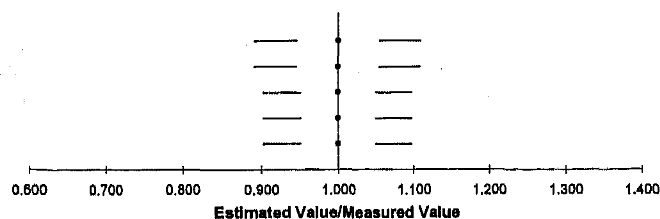
Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (12)	1.015	0.620	0.145	0.290
Garrott (30)	0.992	0.592	0.143	0.286
Garrott's Rule of Thumb (27)	0.968	0.273	0.082	0.164
Rule of Thumb (22)	1.035	0.273	0.082	0.164
Bixel (44)	1.026	0.250	0.073	0.147
PC Crash (52)	1.126	0.256	0.059	0.118
Prism (2)	0.880	0.201	0.052	0.105

C) Yaw Moment of Inertia



Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott (31)	1.010	0.608	0.144	0.288
Reide (13)	0.991	0.583	0.141	0.282
Bixel (46)	1.066	0.474	0.112	0.223
Noon 2 (41)	1.076	0.319	0.092	0.184
SLAM 1 (47)	1.041	0.315	0.091	0.182
Garrott's Rule of Thumb (28)	0.988	0.315	0.091	0.182
mab/Rule of Thumb (23)	1.089	0.315	0.091	0.182
Burg/PC Crash (10)	1.185	0.264	0.066	0.132
Noon 1 (40)	1.007	0.242	0.059	0.119
SLAM 2 (48)	0.945	0.242	0.059	0.119
Prism (3)	0.919	0.252	0.059	0.118
Noon 3 (42)	1.015	0.221	0.059	0.118

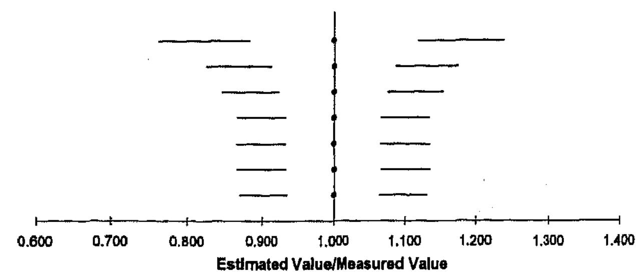
D) Center of Gravity Height



Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott's Rule of Thumb 1 (24)	0.996	0.202	0.054	0.109
Rule of Thumb 1 (19)	1.010	0.202	0.054	0.109
Allen (18)	1.027	0.265	0.049	0.098
Garrott's Rule of Thumb 2 (25)	0.988	0.265	0.049	0.098
Rule of Thumb 2 (20)	0.976	0.265	0.049	0.098

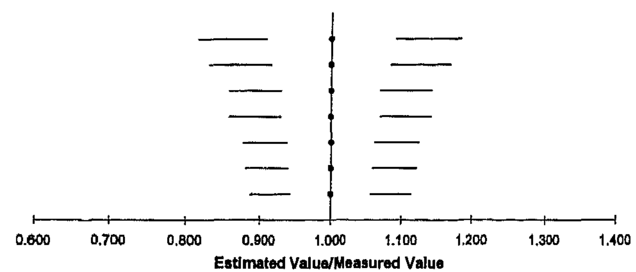
Front Wheel Drive Passenger Cars

A) Roll Moment of Inertia



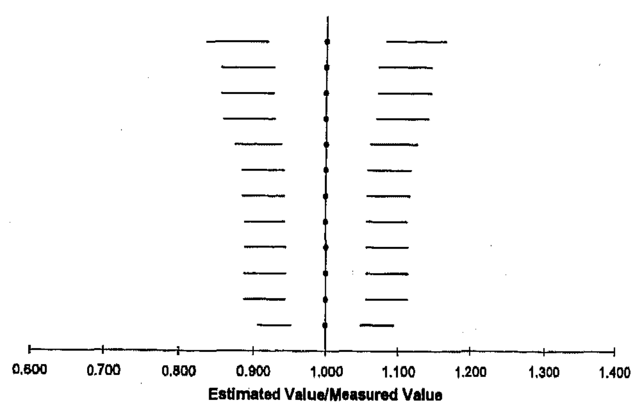
Method	Correction Factor	Range	Standard Deviation	95% Limit
PC Crash (51)	0.822	0.425	0.118	0.237
Garrott (29)	0.961	0.336	0.087	0.174
Reide (11)	1.177	0.309	0.076	0.153
Garrott's Rule of Thumb (26)	0.971	0.331	0.067	0.134
Rule of Thumb (21)	0.709	0.331	0.067	0.134
Prism (1)	1.061	0.329	0.067	0.133
Bixel (43)	1.017	0.321	0.065	0.130

B) Pitch Moment of Inertia



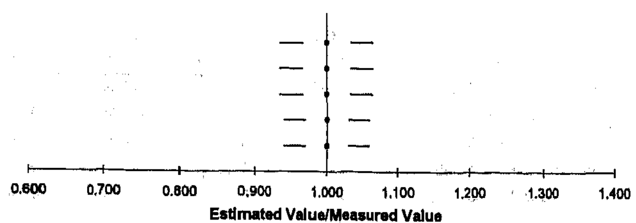
Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott (30)	0.985	0.377	0.091	0.183
Reide (12)	0.984	0.362	0.084	0.168
Garrott's Rule of Thumb (27)	0.986	0.297	0.070	0.141
Rule of Thumb (22)	1.055	0.297	0.070	0.141
Bixel (44)	0.998	0.289	0.062	0.123
Prism (2)	0.896	0.293	0.060	0.120
PC Crash (52)	1.123	0.247	0.057	0.113

C) Yaw Moment of Inertia



Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott (31)	0.989	0.316	0.084	0.167
Noon 1 (40)	1.021	0.319	0.074	0.147
SLAM 2 (48)	0.958	0.319	0.074	0.147
Reide (13)	0.953	0.301	0.072	0.143
Prism (3)	0.922	0.279	0.063	0.127
Bixel (46)	0.991	0.331	0.059	0.118
Noon 3 (42)	1.023	0.267	0.058	0.117
Noon 2 (41)	1.025	0.233	0.057	0.113
SLAM 1 (47)	1.046	0.233	0.056	0.113
Garrott's Rule of Thumb (28)	0.993	0.233	0.056	0.113
mab/Rule of Thumb (23)	1.095	0.233	0.056	0.113
Burg/PC Crash (10)	1.165	0.204	0.048	0.095

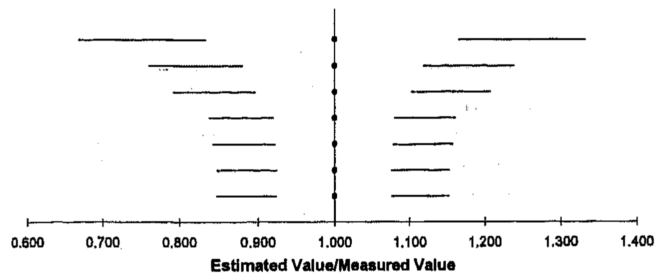
D) Center of Gravity Height



Method	Correction Factor	Range	Standard Deviation	95% Limit
Allen (18)	1.027	0.136	0.033	0.065
Garrott's Rule of Thumb 2 (25)	0.988	0.136	0.033	0.065
Rule of Thumb 2 (20)	0.976	0.136	0.033	0.065
Garrott's Rule of Thumb 1 (24)	0.989	0.146	0.030	0.060
Rule of Thumb 1 (19)	1.003	0.146	0.030	0.060

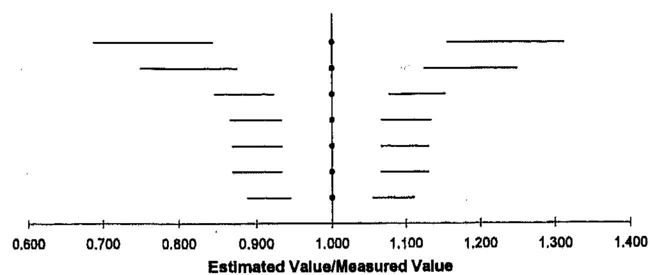
Sport Utility Vehicles

A) Roll Moment of Inertia



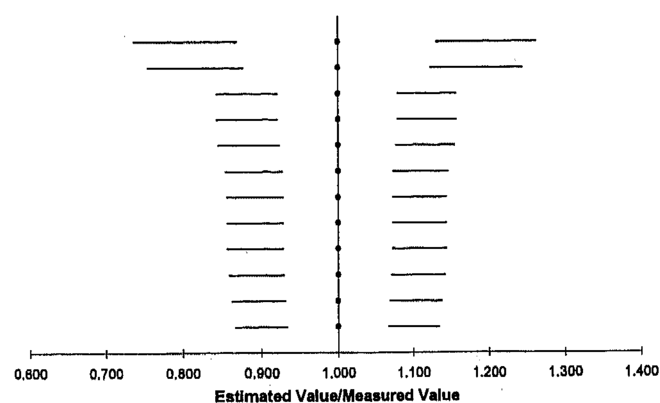
Method	Correction Factor	Range	Standard Deviation	95% Limit
PC Crash (51)	0.893	0.637	0.166	0.332
Garrott (32)	0.786	0.425	0.119	0.239
Reide (11)	1.143	0.379	0.103	0.207
Prism (1)	0.858	0.283	0.080	0.161
Bixel (43)	0.983	0.293	0.078	0.157
Garrott's Rule of Thumb (26)	0.867	0.273	0.076	0.152
Rule of Thumb (21)	0.581	0.273	0.076	0.152

B) Pitch Moment of Inertia



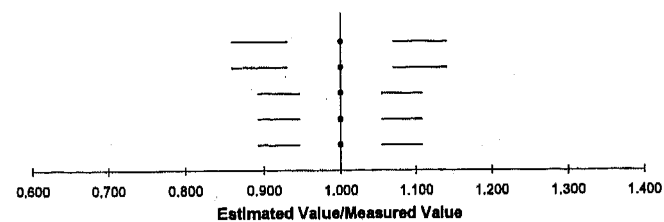
Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (12)	0.870	0.720	0.156	0.312
Garrott (33)	0.850	0.560	0.125	0.250
Bixel (45)	1.017	0.308	0.077	0.153
Prism (2)	0.959	0.238	0.067	0.134
Garrott's Rule of Thumb (27)	1.002	0.269	0.066	0.131
Rule of Thumb (22)	1.042	0.269	0.066	0.131
PC Crash (52)	1.228	0.182	0.055	0.111

C) Yaw Moment of Inertia



Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (13)	0.828	0.601	0.131	0.263
Garrott (34)	0.875	0.523	0.123	0.245
Noon 1 (40)	1.158	0.274	0.079	0.157
SLAM 2 (48)	1.087	0.274	0.079	0.157
Noon 3 (42)	1.153	0.270	0.077	0.155
Noon 2 (41)	1.071	0.289	0.073	0.146
SLAM 1 (47)	1.028	0.280	0.072	0.144
Garrott's Rule of Thumb (28)	1.034	0.280	0.072	0.144
mab/Rule of Thumb (23)	1.076	0.280	0.072	0.144
Prism (3)	1.032	0.237	0.071	0.142
Burg/PC Crash (10)	1.267	0.255	0.069	0.138
Bixel (46)	1.019	0.267	0.067	0.134

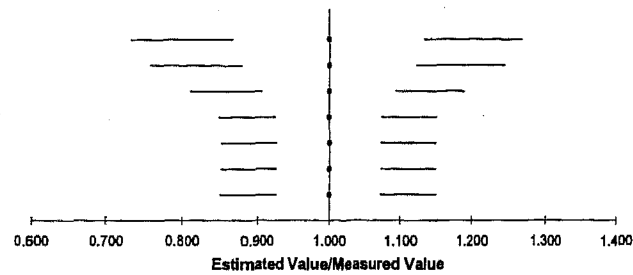
D) Center of Gravity Height



Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott's Rule of Thumb 1 (24)	0.982	0.249	0.070	0.141
Rule of Thumb 1 (19)	1.249	0.249	0.070	0.141
Allen (18)	1.025	0.187	0.054	0.108
Garrott's Rule of Thumb 2 (25)	1.006	0.187	0.054	0.108
Rule of Thumb 2 (20)	0.974	0.187	0.054	0.108

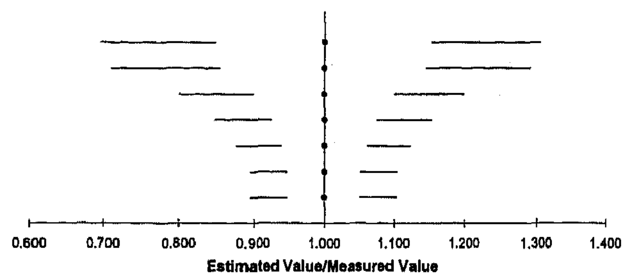
Pickup Trucks

A) Roll Moment of Inertia



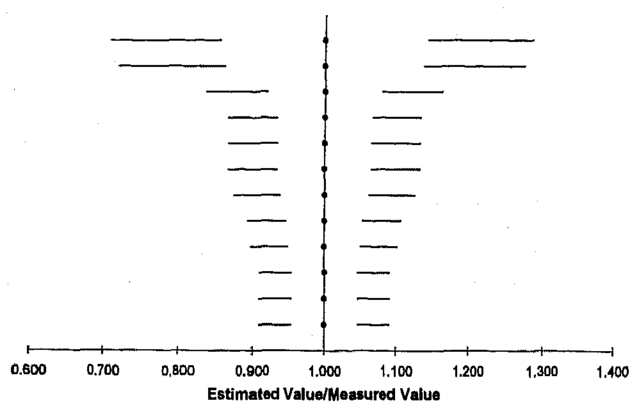
Method	Correction Factor	Range	Standard Deviation	95% Limit
PC Crash (51)	0.613	0.502	0.133	0.267
Reide (11)	1.138	0.392	0.121	0.243
Garrott (32)	0.777	0.328	0.094	0.189
Bixel (43)	0.947	0.317	0.075	0.150
Prism (1)	0.821	0.304	0.074	0.149
Garrott's Rule of Thumb (26)	0.822	0.306	0.074	0.148
Rule of Thumb (21)	0.551	0.306	0.074	0.148

B) Pitch Moment of Inertia



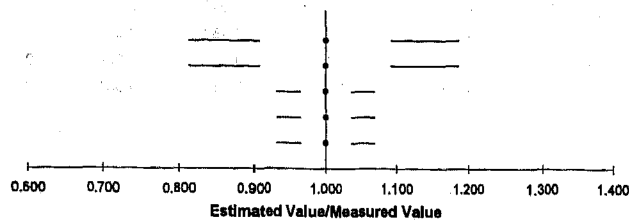
Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (12)	1.125	0.517	0.151	0.303
Garrott (33)	1.074	0.495	0.144	0.289
Bixel (45)	1.005	0.315	0.100	0.199
Prism (2)	0.898	0.274	0.075	0.151
PC Crash (52)	1.094	0.250	0.061	0.122
Garrott's Rule of Thumb (27)	0.916	0.253	0.052	0.104
Rule of Thumb (22)	0.952	0.253	0.052	0.104

C) Yaw Moment of Inertia



Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (13)	1.037	0.465	0.145	0.290
Garrott (34)	1.085	0.443	0.139	0.279
Bixel (46)	1.004	0.283	0.082	0.165
Prism (3)	0.922	0.289	0.068	0.135
Noon 1 (40)	1.007	0.285	0.067	0.134
SLAM 2 (48)	0.945	0.285	0.067	0.134
Noon 3 (42)	0.993	0.255	0.063	0.127
Burg/PC Crash (10)	1.100	0.211	0.054	0.108
Noon 2 (41)	0.928	0.198	0.051	0.103
SLAM 1 (47)	0.915	0.174	0.046	0.092
Garrott's Rule of Thumb (28)	0.921	0.174	0.046	0.092
mab/Rule of Thumb (23)	0.958	0.174	0.046	0.092

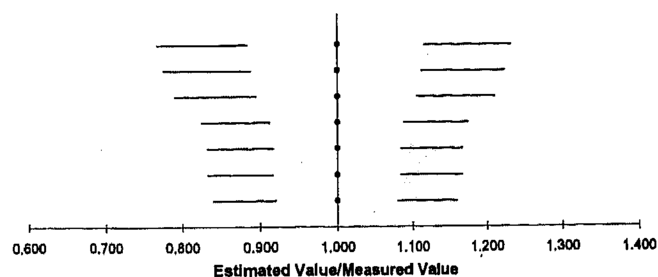
D) Center of Gravity Height



Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott's Rule of Thumb 1 (24)	0.948	0.326	0.093	0.186
Rule of Thumb 1 (19)	1.206	0.326	0.093	0.186
Allen (18)	0.990	0.124	0.035	0.069
Garrott's Rule of Thumb 2 (25)	0.972	0.124	0.035	0.069
Rule of Thumb 2 (20)	0.940	0.124	0.035	0.069

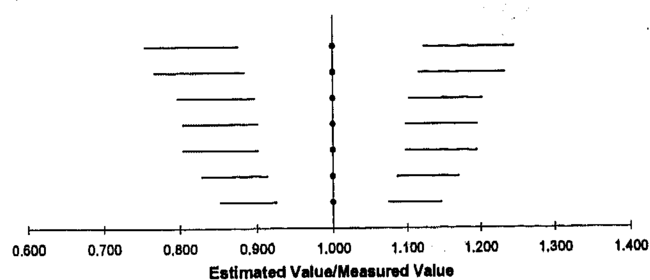
Vans

A) Roll Moment of Inertia



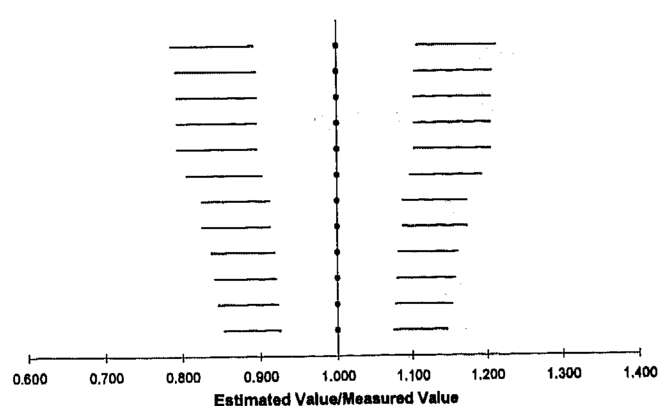
Method	Correction Factor	Range	Standard Deviation	95% Limit
PC Crash (51)	0.896	0.411	0.116	0.232
Reide (11)	1.465	0.335	0.112	0.224
Garrott (32)	0.981	0.318	0.105	0.210
Prism (1)	0.996	0.317	0.088	0.175
Garrott's Rule of Thumb (26)	1.000	0.305	0.084	0.167
Rule of Thumb (21)	0.670	0.305	0.084	0.167
Bixel (43)	0.956	0.291	0.080	0.160

B) Pitch Moment of Inertia



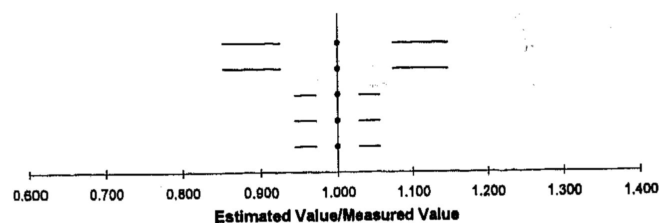
Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (12)	1.067	0.461	0.123	0.245
Garrott (33)	0.993	0.418	0.116	0.233
Bixel (45)	0.988	0.338	0.102	0.203
Garrott's Rule of Thumb (27)	0.934	0.319	0.098	0.196
Rule of Thumb (22)	0.972	0.319	0.098	0.196
Prism (2)	0.999	0.373	0.086	0.171
PC Crash (52)	1.188	0.304	0.074	0.147

C) Yaw Moment of Inertia



Method	Correction Factor	Range	Standard Deviation	95% Limit
Reide (13)	0.989	0.399	0.107	0.214
Garrott (34)	1.021	0.375	0.104	0.208
SLAM 1 (47)	0.939	0.328	0.103	0.206
Garrott's Rule of Thumb (28)	0.945	0.328	0.103	0.206
mab/Rule of Thumb (23)	0.982	0.328	0.103	0.206
Noon 2 (41)	0.960	0.340	0.097	0.194
Noon 1 (40)	1.160	0.333	0.087	0.174
SLAM 2 (48)	1.088	0.333	0.087	0.174
Prism (3)	1.040	0.318	0.081	0.162
Bixel (46)	0.998	0.257	0.079	0.158
Noon 3 (42)	1.126	0.321	0.077	0.154
Burg/PC Crash (10)	1.202	0.309	0.074	0.147

D) Center of Gravity Height



Method	Correction Factor	Range	Standard Deviation	95% Limit
Garrott's Rule of Thumb 1 (24)	1.001	0.214	0.074	0.148
Rule of Thumb 1 (19)	1.273	0.214	0.074	0.148
Allen (18)	1.003	0.118	0.028	0.056
Garrott's Rule of Thumb 2 (25)	0.985	0.118	0.028	0.056
Rule of Thumb 2 (20)	0.953	0.118	0.028	0.056

APPENDIX C - PC-CRASH RUNS

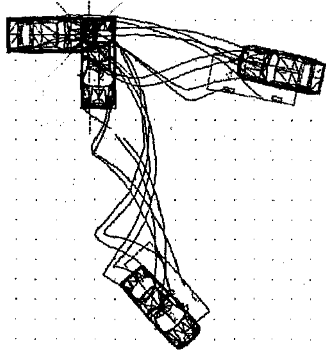


Figure C1 - Simulation with Yaw Moments decreased to 70%

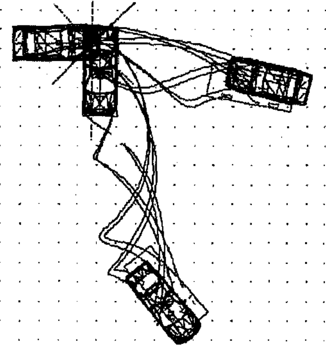


Figure C2 - Simulation with Yaw Moments decreased to 80%

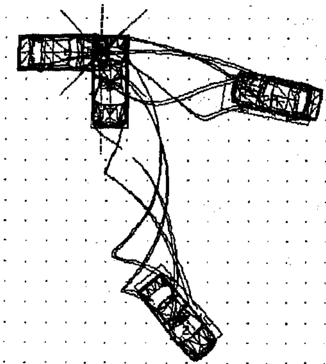


Figure C3 - Simulation with Yaw Moments decreased to 90%

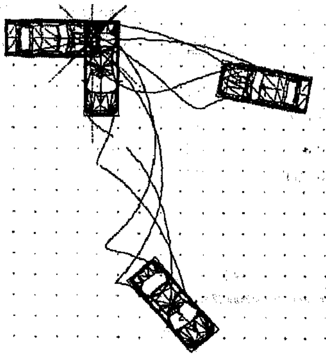


Figure C4 - Simulation with Yaw Moments increased to 100%

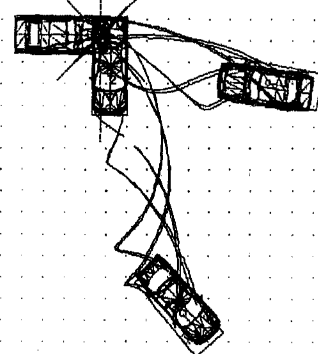


Figure C5 - Simulation with Yaw Moments increased to 110%

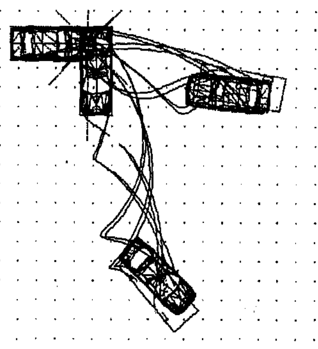


Figure C6 - Simulation with Yaw Moments increased to 120%

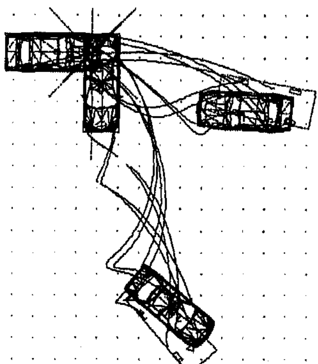


Figure C7 - Simulation with Yaw Moments increased to 130%

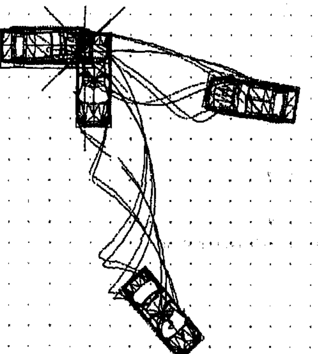


Figure C8 - Simulation with Yaw Moments increased to 130%, and corrected.

SAE # 970951
Reviewer's Discussion
by Bruce Main

A Comparison of Moment of Inertia Estimation Techniques for Vehicle Dynamics Simulation

William E. Cliff, Duane D. MacInnis, Craig A. Brown, Colin W. Blair, Authors

In "A Comparison of Moment of Inertia Estimation Techniques for Vehicle Dynamics Simulation", MacInnis et. al. provide a useful summary of the various vehicle parameter estimation techniques developed over the years. Their review of the vehicle dynamics models assists readers to compare the various techniques and methods used. The authors present the distribution for one vehicle category which is approximately normal. The use of standard statistical tests appear appropriate. However, the reader is left to assume that the other vehicle categories are also approximately normally distributed.

The authors' analysis indicate that earlier estimation techniques for the yaw moment of inertia were quite robust. It is encouraging to find that early and perhaps less sophisticated estimation efforts provide robust results when compared with current analytical capabilities.

MacInnis et. al. accurately use the term "corrected" in the sense that the work of prior authors has been adjusted to minimize the deviation from a calculated value to the measured value. This should not be interpreted as a criticism of the prior work or that the prior work is "incorrect". Rather, the current authors present an analysis to reduce the level of variation between the estimated and measured values.

MacInnis et. al. provide a concise summary of the estimation techniques for moments of inertia for vehicle dynamics simulations. Their analyses of the techniques which provide the smallest error will provide engineers a useful resource for reconstruction analyses.

970952

An Evaluation of Rectified Bitmap 2D Photogrammetry with PC-Rect

William E. Cliff, Duane D. MacInnis, and David A. Switzer
MacInnis Engineering Associates

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Without good-quality measurements taken at the time of an accident the analyst is faced with the need to extract measurement data from incident scene photographs. This paper discusses the history and development of the mathematical model for two-dimensional (2D) single exposure analytical photogrammetry, presents the software PC-Rect, and compares the analytical results obtained with PC-Rect to survey results. The sensitivity of the analytical results to the variation in such parameters as subject distance, camera height, digital photograph resolution, and bitmap density is discussed. The concept of using the directly rectified scanned photograph in the reconstruction task is introduced, and the utility of performing the dynamic simulation directly on the rectified photograph is discussed.

INTRODUCTION

Photogrammetry is the science of producing real world measurements from photographs. 3D photogrammetry requires at least two photographs, both containing all the points of interest, and will yield measurements in three dimensions. Relatively simpler 2D photogrammetry, the subject of this study, can be done from a single photograph and is all that is required for planar surfaces.

Photogrammetry as we know it began in the latter half of the nineteenth century and a good synopsis of its general history is given in the ASP Manual[1]. The history and techniques of non-topographical photogrammetry have been summarized by Karara[2].

The use of photogrammetry in accident reconstruction is not new. The Zurich City Police[3] have been conducting 3D photogrammetric accident reconstruction since at least 1933, over 60 years ago. Their technique, using a pair of "metric" cameras and an optical/mechanical stereo plotter, is discussed in a paper by Lillesand[4]. Bryner[5] published a study in 1974 which used stereo cameras and the "parallax" method of data reduction.

Developments in 3D photogrammetry for use in accident reconstruction have been reported by Sherz[6] and Gillen[7], both using a MACO (AMI) 35/70 analytical stereo plotter developed by H. Dell Foster. Wester-Ebbinghaus[8] has proposed 3D photogrammetry for vehicle crush measurements, and Wolf[9] has presented a general discussion of the use of photogrammetry in forensic work. More recent developments have included the camera reverse projection technique discussed by Main[10], and 3D methods reported by Pepe[11,12] and Fay[13]. A good summary of the accuracies expected was offered by Tumbas[14].

Graphical 2D methods include the "perspective grid" approach, discussed by Baker[15], Whitnall[16] and Kerkoff[17], and the less well-known "paper strip" method[18]. A good summary of graphical photogrammetry techniques is given by Williamson[19]. Although graphical methods can yield useful information when carefully applied, they are time-consuming and prone to error.

In the early 1970's, Bleyl[20,21] proposed an analytical technique using the projectivity equations and published a Fortran IV algorithm. Since then, a number of 2D software packages for microcomputers, including those by Niederer[22], Brelin[23], Kinney[24], and Smith[25], have been developed and published. 2D analytical photogrammetry can be done by rectifying the entire bitmap image rather than calculating the coordinates for discrete points, presented by Pepe[26] and Rohde[27].

For most accident scenes, 2D photogrammetry is the preferred method, because of its relative simplicity and because the accident reconstructionist often has only one photograph of a particular area of the road surface.

PC-Rect[28] is a Windows©-based 2D photogrammetry program. PC-Rect will rectify an oblique photograph of a planar accident scene into a scaled plan view "photographic" image. This image contains all the skid marks, road lines and other road markings that appear in the original oblique photograph, as well as discrete points of interest.

Even when no measurements have been taken at the time of an accident, the reconstructionist can obtain enough dimensional data to rectify photographs by going to the scene and taking measurements of painted lines or other long-lasting roadway features that appear in the original photographs.

The purpose of this study was to evaluate PC-Rect's accuracy by comparing scene images rectified with PC-Rect to survey results. Camera lens focal length, camera location and the number and type of input dimensions have been varied to assess the effect of these factors on the accuracy of the results.

THEORY

The typical case encountered in accident reconstruction is a scene photograph taken by a police officer of the roadway and incident artifacts. In this case, the film plane, or image plane, is nearly vertical with respect to the roadway or object plane, and the photograph is simply a projection of the object plane onto the image plane. The process of reversing this projection, i.e., creating a map from the photograph, is known as rectification.

The theory of 3D and 2D rectification is quite similar. The fundamental assumption of rectification is that the camera focal point, a point on the negative and the corresponding surface point are collinear (there is no bending of the rays between object and image). In 2D rectification there is an additional assumption that the surface to be rectified is planar.

The theory of 2D and 3D rectification has been included as Appendix C. The 3D and 2D mathematics rely on the published works by Moffit[29], Wolf[30], Strang[31], and the ASP Manual. The theory outlines methods for developing the solution for 3D rectification and then shows the simplification for 2D rectification with PC-Rect.

ACCURACY

The accuracy presently expected when reconstructing an accident is less than that required in many "high-tech" applications. Accuracies of 0.3 m for the scene information may be adequate to reconstruct collisions where speed from skid distance is the task. Greater precision is required in some cases, as discussed later in this paper.

An analyst's equipment can be calibrated by staging a scene, carefully surveying it, mapping it by photogrammetry, and then comparing the two results. This check has been performed by Kerkoff, Niederer, Smith, Grimes[32], Tumbas and others for the 2D case.

Kerkoff reported an error of 0.05 m from an 8"x10" enlargement of a photograph taken through an inexpensive zoom lens 11.0 m from the subject. Niederer reported an accuracy of 0.2 m using 2D photogrammetry from the negative compared to measurements obtained from routine 3D photogrammetry performed by the Zurich Police. Using Brelin's program, Grimes found a maximum error of 0.64 m at a distance of about 33 m from both 8"x10" enlargements and color slides. Tumbas provided a good summary of a number of 2D and 3D methods, and the accuracy expected.

The potential sources of error in 2D rectification include:

CAMERA IMPERFECTIONS - There is no perfect camera. The "metric" camera traditionally used in photogrammetry has a fixed focal length, and the variations from the ideal optical axis and principal point are known. The distortion of the lens has been determined and can be accommodated in the analysis.

Standard 35 mm cameras commonly used by police agencies, insurance adjusters and analysts are "non-metric" cameras. However, these cameras are thought to produce adequate accuracy for accident reconstruction purposes. Hatzopoulos[33] tested a common high-quality non-metric camera (Olympus OM-2N) and found that it provided accuracies of at least 1:7000.

IMPERFECTIONS IN CREATING THE BITMAP - There are several methods available to create a bitmap from a negative. Likely the most common method used is scanning a print made from the negative, which involves the accuracy of the printing process and the accuracy of the scanner. The accuracy of the printing process, which depends on the equipment used, would often be difficult to obtain from the print shop. The accuracy of the scanning process depends on the scanner, and is thought to be significantly better for flat bed scanners than for hand held scanners.

NON-PLANARITY OF THE ROAD SURFACE - The greatest potential source of error in 2D rectification of photographs taken from oblique angles is any vertical variance of the ground points with respect to the plane defined by the control points. This condition, which has been discussed by Kinney, Niederer and Smith, can be illustrated in an example where the measured points are all at the pavement edges and the road is crowned as shown in Figure 1.

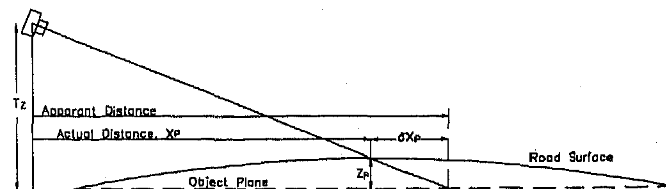


Figure 1 - Non-planarity error

When rectified, points on the road surface which lie above the object plane will appear farther away from the camera than they actually are. Similarly, points on the road surface which lie below the object plane would appear closer to the camera when rectified. The amount of error in the rectification of a point at some height

above or below the object plane can be derived from similar triangles on Figure 1, as follows:

$$\delta X_P = \frac{Z_P X_P}{T_Z - Z_P} \quad (1)$$

where

T_Z = camera height
Z_P = height of point P
X_P = distance on road plane from camera
δX_P = apparent distance - actual distance

The non planarity error increases linearly by the horizontal distance from the camera, although the error also depends on the camera height and the point height. The error increases as the height of the point above the object plane increases, and decreases as the camera height increases. For example, if a 14.6 m wide road with a crown having a two percent cross slope is defined by measurements along each curb edge, a point at the centerline of the road (Z_P= 0.146m) that is 15 m from the photographer with a camera height of about 1.63 m would appear 1.48m further away than it actually is.

This problem can, however, be overcome with a piece-wise analysis in which each side of the road is rectified separately. In this case the maximum error can be reduced significantly as shown in Figure 2 below. This can be done with PC-Rect, as will be shown later.

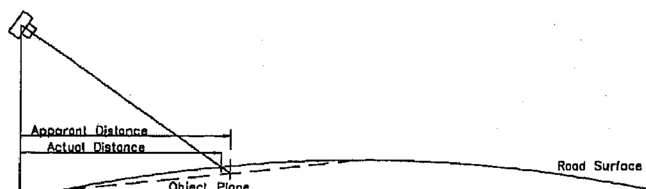


Figure 2 - Non planarity error for half of road surface

CONTROL POINT ACCURACY AND LOCATION - The accuracy in the measurement of the control points will also influence the accuracy of the rectification process. In general for nominally flat surfaces, control points with long distances between them will result in greater accuracy than for control points with short distances between them, since any uncertainty in the position of the points will be averaged over the longer distance.

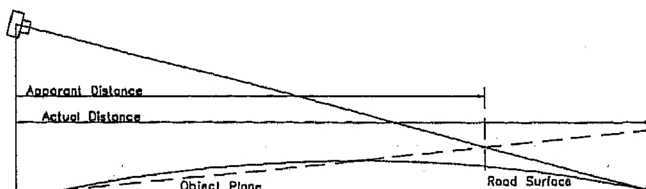


Figure 3 - Error in far side based on near side control points

Also, it is important to choose control points which surround the area of interest. In Figure 3, even though the error on the near side of the road surface may be low, the error on the far side could be great, if the rectified image is based on a plane defined by near side curb and road centerline control points, as shown.

BITMAP RESOLUTION - The resolution of the bitmap will influence the accuracy of the rectification, especially at increasing distances from the camera.

In a typical scene photograph, moving from the bottom of the photograph upwards, the road surface occupies less and less of the photograph width (Figure 4). In the upper portion of the photograph, each pixel represents a greater length and width of the road surface. Therefore, when a bitmap is rectified the resolution will diminish as the distance from the camera increases. The final resolution of the rectified bitmap will depend largely on the pixel density of the original bitmap and the distance from the camera to the area of interest, but will also depend on factors such as camera height, focal length, and photograph print size.



Figure 4 - Typical oblique photograph of road surface

In PC-Rect the user must input the desired resolution (in pixels/meter) of the rectified bitmap. The resolution of a particular point of interest on the rectified bitmap will depend on both the original bitmap resolution and the rectified bitmap resolution, with the original bitmap resolution affecting the quality largely for points far from the camera and the rectified bitmap resolution affecting the quality largely for points near the camera. Images 15 to 22 in Appendix D show the effects of varying the original bitmap resolution and the rectified resolution. It is clear from these images that scanning the photograph at a higher resolution will give superior results, even when rectifying at a lower resolution.

PROCEDURE

Two scenes were photographed from various positions and surveyed. Camera height was kept constant (at normal standing height) at one scene (Scene A) and was varied at the other (Scene B).

Scene A was a flat, slightly sloped parking lot with parallel painted divider lines at right angles to a concrete curb. Two different SLR cameras were used to take photographs from various positions. Camera 1 had a fixed nominal 50mm focal length lens, while Camera 2, which was equipped with a 28-70mm zoom lens, was used to take photographs at 35 and 70mm zoom settings. The photographs were all taken by a person

standing on the ground, resulting in a lens height of about 1.56 meters.

For Scene A, color 100 ASA film was used, and the negatives were developed onto 4" x 6" prints. These prints were scanned into bitmap files at 150 dpi with a flat bed scanner. This resulted in 600 x 900 pixel image with a file size of about 1.5 MB (color image) or 0.5 MB (256 shade black & white image). After confirming the black & white images would show all the required information, they were used in all cases presented. The images were rectified in PC-Rect to a resolution of 30 pixels/meter, and then overlayed on a DXF scale drawing of the incident scene so that the accuracy of the rectified image could be easily judged.

When scanned prints are used, it is important to scan the full print, such that the center of the scanned image is near the center of the negative. A distorted rectified image will result if only one side of a print is used. If the print is cropped so that the new center is the same as the original one, the rectified image will be correct, but the calculated focal length and camera height will be incorrect. This is also true if the specified film size is incorrect.

Scene B was a straight length of 12 meter wide two lane roadway with numerous tire marks. Scene B had a significant crown of approximately constant curvature, with the center of the road being about 16 cm higher than the outside edges. At this scene, photographs were taken from two heights: by a person standing on the ground (nominal lens height of about 1.56 meters) and a person standing on a step ladder (nominal lens height of about 2.42 meters). An SLR camera with a fixed lens focal length of 50mm was used for this scene.

Color 100 ASA film was again used, and the negatives were digitized onto a Kodak Photo CD. The digitized photographs were rectified from 2048 x 3072 pixel bitmaps. This resulted in a file size for each picture of about 6 MB (256 shade black & white image). The images were rectified in PC-Rect to a resolution of 30 pixels/meter, and then overlayed on a DXF scale drawing in the same manner as Scene A.

RESULTS

The results of the rectifications are shown in bitmap image form in Appendix A. Cases 1 to 10 are from Scene A and Cases 11 to 14 are from Scene B.

SCENE A - Cases 1 to 10 are presented in groups of three images (Image 1a, 1b, and 1c, etc.), one group per page. All images with the subscript "a" (and Image 5b) show the original oblique photographs with the reference lines (known dimensions) used to rectify the photographs. All images with the subscript "b" (except Image 5b) show the rectified photographs overlayed on the scale drawing of the scene. All Images with the subscript "c" (except Image 5c) show close-up views of the area of the scene within about 12*h* (where *h* is the camera height above the scene surface) of the camera position. A 12*h* radius is shown with a dotted line, as are each photograph's field of view limits running from the camera position.

Case 5 shows the results of combining two rectified images in PC-Rect. This is done by selecting two common points in each of the rectified images. The overall accuracy of the resulting joined image depends primarily on how accurately the common points were selected, and the distance between them. The larger this distance is, the less will be the angular error between the two images.

For the connected Image 5c, the error is not reported. The distance errors in each image were similar to the errors reported for the other images. There was little or no increase in these values when the images were joined (this depends only on how accurately the connection points are chosen), and so the overall error of the joined image, as a percentage of the combined scene length, was lower than for the individual images.

As the distance from the camera increases, the clarity of the rectified image decreases, as expected. A radius of about 20*h* is about the maximum practical distance for viewing the location of the white lines in the images shown, which, as stated previously, were all rectified at 30 pixels/meter from a 150 dpi original.

The maximum errors between points in the rectified images and the actual scene are shown in Table 1. The scene depth in each case is defined as the distance between the camera and the farthest (unmeasured) point from it in the portion of the scene under consideration (12*h* to 22*h* for the given cases).

Table 1 indicates that errors of less than 1% of the scene depth occurred where the scene depth was less than about 12*h*, except for Case 6, where an error of about 2% occurred. In most cases, the maximum error was lengthwise in the photograph, as expected.

The 2% error in Case 6 occurred at the two parking lot divider lines closest to the camera (on the near side of the raised curb divider). It is not necessarily true that the least amount of error in the scene will be closest to the camera, because the location of the area of maximum error in the rectified image is also dependent on where the measurements have been taken in the scene.

It can be seen in Image 6a that the measurements used to rectify the photograph are in the middle portion of the photograph, with none in the foreground. When two foreground reference lengths were added to this photograph (see Case 7), the error in the rectified image decreased from 2% to 0.6%.

Generally, if foreground measurements are included, this area will contain the least error. This is primarily because of the increased angle (from horizontal) between a straight line from the camera to a foreground

point over one running to a point further in the distance. This increased angle will have the same effect as an increased camera height, in that it will reduce the error effects of any unevenness in the scene.

Table 1 - Scene A Errors

Case No.	Focal Length	Camera Ht. (m)	Scene Depth (m)	Scene Depth / Camera Ht.	Maximum Error cm	Maximum Error % Dist.
			20	12	13	0.7
1	50mm	1.71	26	15	18	0.7
			31	18	46	1.5
			20	13	13	0.7
2	50mm	1.56	26	17	25	1.0
			31	20	25	0.8
			20	13	9	0.5
3	50mm	1.56	26	17	13	0.5
			31	20	42	1.4
			20	13	17	0.9
4	50mm	1.56	27	17	17	0.6
			34	22	91	2.7
5	50mm	1.56	------	------	-----	-----
			20	12	40	2.0
6	35mm	1.71	26	15	48	1.8
			31	18	120	3.9
			20	12	12	0.6
7	35mm	1.71	26	15	30	1.2
			31	18	80	2.6
			20	13	10	0.5
8	35mm	1.56	27	17	30	1.1
			34	22	120	3.5
			20	12	11	0.6
9	70mm	1.71	26	15	22	0.8
			31	18	42	1.4
			20	13	16	0.8
10	70mm	1.56	27	17	25	0.9
			34	22	102	3.0

Even at scene depths of 15*h* - 17*h*, the errors for all but Case 6 were still in the 1% range. Errors increased markedly at greater scene depths, however, to the 2% to 4% range at scene depths of 18*h* - 22*h*.

PC-Rect reports the individual errors for the entered reference lengths, as well as a weighted total error, defined as follows:

$$\text{Weighted total error} = \frac{\sqrt{\sum_{i=1}^{n}(w_i \cdot \delta L_i)^2}}{\sum_{i=1}^{n} w_i} \quad (2)$$

where

n = Number of the reference distances
w_i = Weight of the distance i (in %)
δL_i = Relative error of the distance i

The weighted total errors for the reported cases were less than 0.5% except for Case 5 (Image 5b) and Case 6, which were 1.25% and 0.94%, respectively.

Table 1 shows that, for this scene and camera combination anyway, the error differences between lens focal lengths was relatively minor. The 35mm focal length rectified images were, however, less clear than the 50mm or 70mm ones, especially at the end of the scene furthest from the camera.

The focal lengths reported by PC-Rect were close to the actual camera focal length settings. For the 50mm fixed lens, PC-Rect gave values of 52 to 55mm. For the zoom lens, PC-Rect gave values of 35 to 38mm at the 35mm setting and 68 to 71mm at the 70mm setting. Thus, assuming the film negative size (commonly "35mm") is known, the approximate lens focal length can be determined with PC-Rect.

The camera heights reported by PC-Rect were also close to the actual heights. Values of 1.59 to 1.70m were returned for the 1.71m height (photographer on a 0.15m high curb) and 1.51 to 1.60m for the 1.56m height. The approximate elevation of the camera used to take a scene photograph can also thus be determined with PC-Rect.

If the camera focal length and/or height are known for a photograph, these values can be entered and fixed in PC-Rect. This is especially useful in cases where only a few scene dimensions are known. Normally at least four scene dimensions are required, but this can be reduced if one or more camera parameters are known. For all the cases referred to in this paper, none of the four camera parameters (focal length, height, rotation or inclination) were fixed in PC-Rect.

Note that there are diagonal reference lengths in all the reported cases. When only longitudinal and lateral reference lengths are included (Figure 5), the rectified picture can become skewed slightly, as shown in Figure 6 (compare with Image 6c). However, the overall error in the area of the rectified image shown in Figure 6 was still less than +/-1%, indicating the skewing was not that significant.

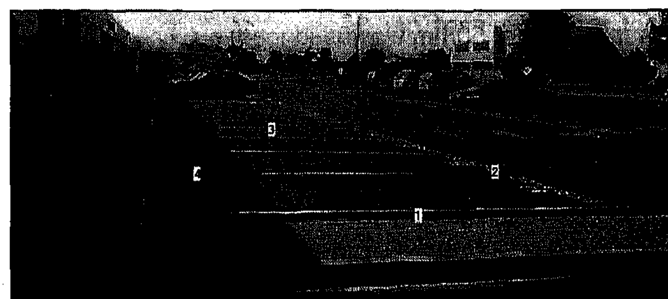


Figure 5 - Image 6a without diagonal reference lengths

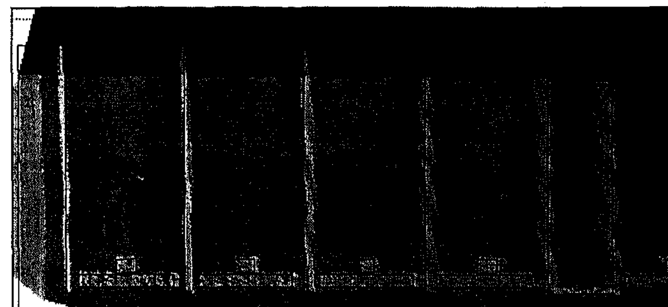


Figure 6 - Rectified without diagonal reference lengths

SCENE B - Cases 11 to 14 are presented in groups of three images in the same manner as Scene A, with the exception of Case 13 which has a fourth image showing

two bitmaps joined together. The images from Scene B are shown in Appendix B.

The photographs for Cases 11-14 were taken from opposite ends of several skid marks. Cases 11 & 12 show photographs taken from approximately the same location but at different heights. Case 11 was taken from a camera height of about 2.42 meters, and Case 12 was taken from a camera height of about 1.56 meters. Cases 13 & 14 show the same section of road surface taken from the opposite direction and show both the near and far sides of the road. As for Cases 11 and 12, Case 13 was taken from a camera height of about 2.42 meters, and Case 14 was taken from a camera height of about 1.56 meters. For each photograph pair (Cases 11&12, and Cases 13&14), the lengths between four points on the near lane of the road surface were chosen as the known distances. Also, for each pair the same four points were used.

The maximum errors between points in the rectified images and the scaled scene drawing are shown in Table 2. The maximum errors shown are for the near lane of the road surface. The errors in the far lane of the road surface are much higher. The weighted total errors for Cases 11-14 were less than 0.5% except for Case 11 (Image 11b) where the weighted total error was 0.67%

For Cases 11 and 12 the maximum errors were all 1% of the scene depth or less. The errors in Case 11, with the higher camera height were slightly greater than in Case 12, with the lower camera height. This is opposite to what was expected. The higher camera height should have produced less error based on reducing the effect of non-planarity of the road surface. Since all the errors were low for these cases, the road surface was likely quite flat and non-planarity was probably not a large factor in the error. The weighted total errors for Case 11 and Case 12 were 0.67% and 0.47% respectively, so the overall lower error in Case 12 may be due to the difference in the weighted total error.

The errors on the near side of the road from Cases 13 and 14 were also quite similar. The errors on the far side of the road surface were larger for the lower camera height (Case 14), than for Case 13. For both cases the far side of the road is reduced in width considerably (Image 13b, c and Image 14b, c). This demonstrates the error on the far side of a road surface based on near side measurement points, as illustrated in Figure 3. Once again, the error can be reduced by rectifying both sides of the road separately and then joining them. Image 13d shows the rectified bitmap with both sides of the road rectified separately and connected together along the centerline of the road.

Cases 11-14 did not show any appreciable increase in accuracy for the higher camera heights, with the exception of the far lane of the road in Cases 13 and 14. In each of Cases 11-14 the long curved skid mark on the near lane was reproduced quite accurately, both on overall length and in radius of curvature.

Table 2 - Scene B Errors

Case No.	Focal Length	Camera Ht. (m)	Scene Depth (m)	Scene Depth / Camera Ht.	Maximum Error cm	Maximum Error % Dist.
			20	8	10	0.5
11	50mm	2.42	26	11	15	0.6
			30	12	30	1.0
			19	12	10	0.5
12	50mm	1.56	24	15	10	0.4
			28	18	10	0.4
			16	7	10	0.6
13	50mm	2.42	24	10	20	0.8
			30	12	40	1.3
			18	12	8	0.4
14	50mm	1.56	23	15	5	0.2
			25	16	25	1.0

DISCUSSION

It is beneficial to know what effect the errors in a rectified photograph can have on the accuracy of the reconstruction of an accident. A 1% error in the overall length of a skid will only cause a 0.5% error in the calculated slide to stop speed, since this is a function of the square root of the skid distance. However, a 1% overall scene depth error does not mean the skid length error is also 1%, unless the skid length is as long as the scene depth. This is because the stated error could be between any two points in the photograph, due mainly to local irregularities in the road surface.

For example, with a scene depth of 20m and a rectified image error of 1%, or 20cm, the error of a 5m long skid in that scene could be 5m +/-20cm, or +/-4%. Table 3 shows what effect this can have on the calculated speed of a vehicle skidding at 0.75g to a stop from distances of 5 to 20m.

Table 3 - Slide to Stop Speed Errors

Skid Dist. (m)	Error in Skid Dist. (cm)	Error in Skid Dist. (%)	Speed (km/h) Min.	Speed (km/h) Max.
5	+/-20	+/-4.0	30.3	31.5
10	+/-20	+/-2.0	43.2	44.1
15	+/-20	+/-1.3	53.1	53.8
20	+/-20	+/-1.0	61.4	62.1

The effect of speeds calculated from yaw marks is more complex. The speed V of a vehicle along a curved path is as follows:

$$V = \sqrt{aR} \qquad (3)$$

where

a = centripetal acceleration
R = radius of vehicle's path

If there is a 1% error in R, this will only cause a 0.5% error in the calculated speed, similar to the skid case. However, the radius of yaw marks is usually not measured directly, but is instead calculated from measured chords and offsets using the formula:

$$R = \frac{C^2}{8m} + \frac{m}{2} \quad (4)$$

where

C = chord length
m = offset at center of chord

If there is a 1% error in C, the calculated speed error will also be 1%, a direct relationship. For most accident reconstruction cases, the $m/2$ term is insignificant and a 1% error in m will cause only a 0.5% error in speed.

The offset m is, however, usually a very small dimension. For example, a vehicle cornering at 100 km/h at a centripetal acceleration of 0.75g in pure yaw (no longitudinal drag from the wheels) can leave yaw marks with an average radius R of about 105m. If the measured chord length C of a 105m radius yaw mark is 10m, the offset m will be only about 12cm. A +/-20cm error (1% of a 20m long scene) in the offset m would thus produce an unacceptable error in the calculated speed.

However, it is the relative error between the end points of the chord and the offset at its center, not the overall scene error, that determines the accuracy of the calculated radius.

Good results will generally be obtained if the photograph to be rectified is a view looking along the yaw mark. In this case there will be little or no relative error in the curved mark between the endpoints of the chord, in the direction across the photograph (in line with the offset m). Although this will also give the greatest amount of possible error in C, a +/-20cm scene length error will result in only a maximum error of +/-2% in a 10m long chord.

If there was a +/-20cm overall error in the rectified photograph (1% of a 20m long scene), there should be only a +/- 1 or 2% error in both the 10m chord length and the offset m. If the chord and offset were both either +2% or -2%, the calculated speed error would be about +/-1%. If the chord error was +2% and the offset error was -2%, the calculated speed error would be about +3%.

However, the error in the offset m could be higher than this, depending on the resolution of the scanned image and also the rectified image. In practice the error would likely be on the order of +/-1cm, or about +/-10% for the stated case with m = 12cm. This error could be higher for uneven scenes where the view is not along the length of the curved mark. A +/-10% error in m will result in speed calculation errors of +/-3 to 8% when combined with a +/-2% chord length error, compared to 1% to 3% with a +/-2% error in both m and the chord length.

The analyst must be cautious when measuring chord offsets in a rectified photograph when a wide angle lens has been used. Figures 7 and 8 each show overhead photographs of a grid of 1x1 unit squares along with 105 unit radius curves, with a chord and offset of 10 units and 0.12 units, respectively. Figure 7 was taken with a 28mm wide angle lens and Figure 8 was taken with a 50mm lens. It can easily be seen that the distortion is so great near the edge of the 28mm photograph in Figure 7 that the radius of the actual curve would be calculated incorrectly.

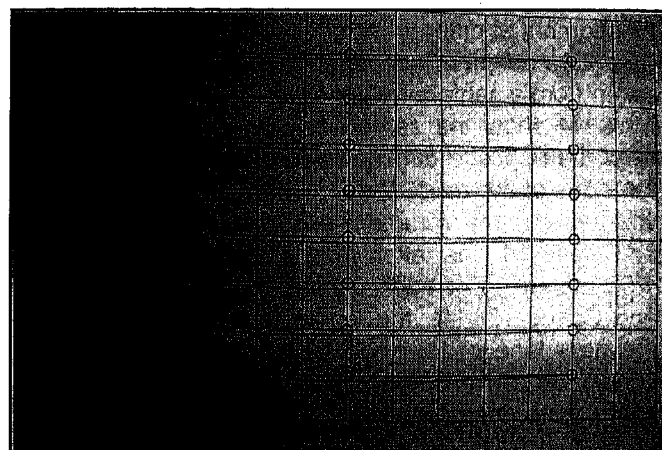


Figure 7 - Plan view of grid with 28mm lens

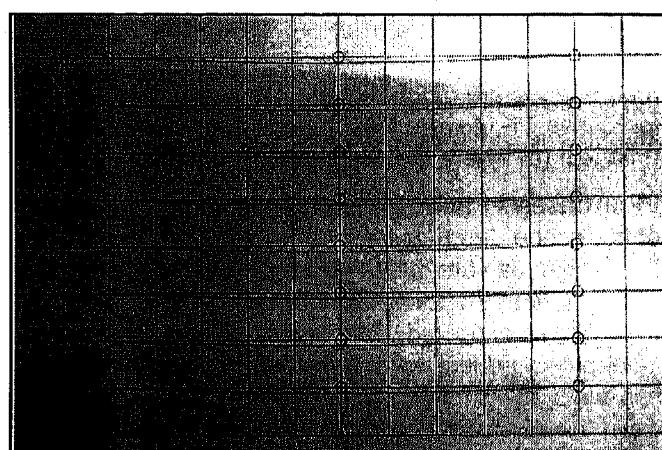


Figure 8 - Plan view of grid with 50mm lens

Thus, if a wide angle lens was used as the basis of a rectified photograph, curve radius calculations should not be relied on unless the curve is near the center of the photograph. The 50mm photograph in Figure 8 is relatively distortion free, and would give good results even in areas near the edge of the photograph.

To illustrate this, an oblique photograph of the grid in the previous two figures was taken with a 50mm lens from a point about two units above the grid plane (Figure 9). This photograph was rectified using the reference lengths shown, resulting in the image in Figure 10.

Chord and offset lines were drawn for all the 105 unit radius curved lines in Figure 10, as shown. The error in the offset m (which was 0.12 units long) was within +/-0.01 units, regardless of where the curved line was in the photograph. Note that the rectified line width in Figure 10 is about 0.02 units, which could be the main source of the error.

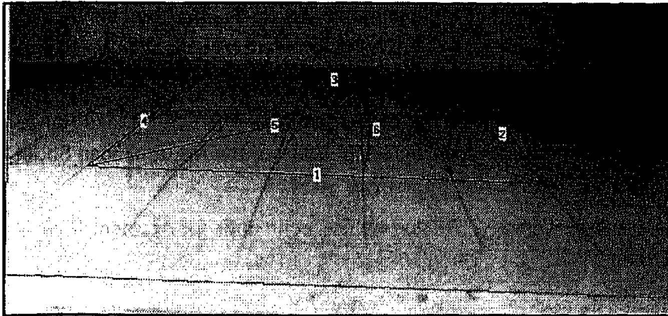


Figure 9 - Oblique view of grid, 50mm lens

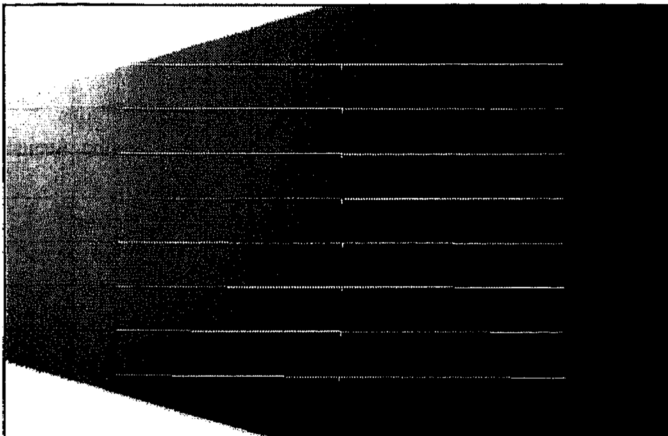


Figure 10 - Rectified view of grid from Figure 9

CASE EXAMPLE

Figure 11 shows an accident scene photograph taken by the attending police. In this accident, a Ford Mustang was struck on its left side by the front of a GMC van. Although the police measured the impact point and rest positions at the scene, no measurements were made of the Mustang's post-impact tire marks. In order to do an accurate reconstruction, the positioning of these marks was required.

PC-Rect was used to rectify this photograph, which was then exported to PC-Crash[34], along with a DXF drawing including the police measurements. The accident was reconstructed with PC-Crash over the drawing and the rectified image, as shown in Figure 12.

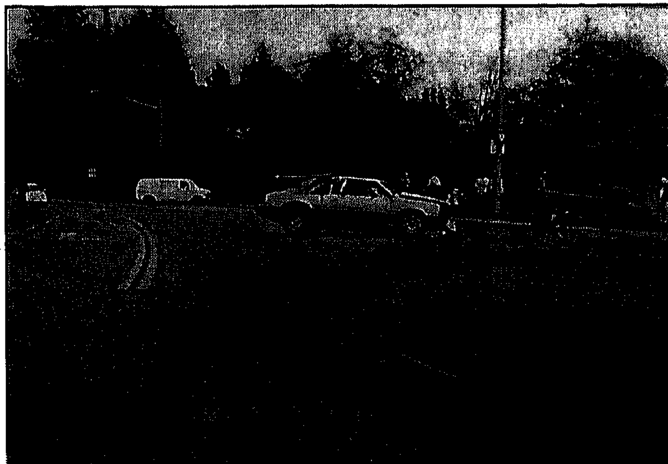


Figure 11 - Police scene photograph

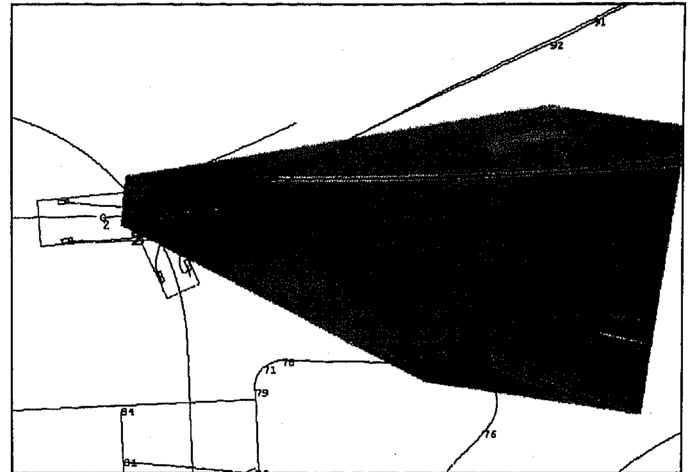


Figure 12 - Simulation over drawing and rectified image

The post-impact tire marks in Figure 12 allowed the reconstructionist in this case to vary the pre-impact speeds until both simulation vehicles matched the incident vehicles' post-impact trajectories, providing a unique solution for the speeds of each. The resulting pre-impact speeds were approximately 16 km/h for the left-turning Mustang (which was accelerating from a stop sign) and 81 km/h for the striking GMC van, which was travelling straight through the intersection in a 50 km/h speed zone.

CONCLUSIONS

1. PC-Rect can be used to rectify photographs of flat accident scenes, resulting in a high degree of accuracy (less than 1% error for scene depth / camera height ratios less than about 12, with no significant increase in error until this ratio was over the 15 to 17 range).

2. Accuracy differences between the 50mm fixed lens and the zoom lens at 35mm and 70mm were relatively minor.

3. PC-Rect can be used to determine the approximate height and lens focal length of the camera used to take a photograph.

4. It is desirable to have diagonal reference lengths in the photograph as well as lateral and longitudinal ones, to avoid skewing of the rectified image. However, when no diagonal reference lengths were used in one of the examples, the resulting small amount of skewing did not cause a significant increase in error.

5. PC-Rect can be used to rectify photographs of non-flat scenes, although lower accuracies can be expected than for flat scenes. For better accuracy, non-flat scenes can be divided into smaller areas of relatively small unevenness, which are each rectified into a separate image. These can then be joined together into one image which has the same absolute accuracy as each individual image.

6. Two different methods were used for generating the photograph bitmaps for Scenes A & B, and there was no clear difference in accuracy between the two. Thus,

scanning the photograph print with a flat bed scanner, and digitizing the negatives onto a Kodak Photo CD both appear to be acceptable methods for generating the photograph bitmap for accident reconstruction purposes. The Kodak Photo CD is, however, likely the most accurate as it eliminates any errors in printing.

7. Caution should be exercised when calculating the radius of curved marks from rectified photographs. Only views looking along the curved marks should be used. When a wide angle lens has been used to take the photograph, the calculated radius of a mark not near the center of the photograph would be suspect.

REFERENCES

[1] "Manual of Photogrammetry", Fourth Edition, American Society of Photogrammetry, 1980.

[2] "Non-Topographic Photogrammetry", Second Edition, American Society of Photogrammetry and Remote Sensing, 1989.

[3] "1933 - 1983 50 Jahre Fotogrammetrie im Dienste der Stadtpolizei Zurich ("50 Years of Stereophotogrammetry") City Police of Zurich, 1983.

[4] Lillesand, Thomas M.; Clapp, James L., "The Utility of Stereometric Systems in Traffic Accident Investigation", American Society of Photogrammetry, Symposium on Close-Range Photogrammetry, January 1971.

[5] Bryner, Clifford G., "A Photogrammetric System for Motor Vehicle Accident Investigation", NHTSA PB-231562, DOT HS-801-098, 1974.

[6] Scherz, James P.; Gillen, Larry, "Potential Use of 35 mm Cameras and the 35/70 Analytical Stereoplotter for Mapping Accident Scenes, Crime Scenes, and Other Subjects". Close-Range Photogrammetry and Surveying: State-of-the Art, American Society of Photogrammetry, September 1984.

[7] Gillen, Larry G., "Photogrammetric Mapping of Vehicle Deformations", SAE 861421.

[8] Wester-Ebbinghaus, Wilfrid; Wezel, Ulrich E., "Photogrammetric Deformation Measurement of Crash Vehicles", SAE 860207.

[9] Wolf, Paul R., "A Summary of The Use of Close-Range Photogrammetry in Crime and Accident Investigations", Close-Range Photogrammetry and Surveying: State-of-the Art, American Society of Photogrammetry, September 1984.

[10] Main, Bruce W., Knopf, Eric A., "A New Application of Camera Reverse Projection in Reconstructing Old Accidents", SAE 950357.

[11] Pepe, M.D., Sobek, J.S., Huett, G.J. "Three Dimensional Computerized Photogrammetry and its Application to Accident Reconstruction", SAE 890739.

[12] Pepe, M.D., Sobek, J.S., Zimmerman, D.A. "Accuracy of Three-Dimensional Photogrammetry as Established by Controlled Field Tests", SAE 930662.

[13] Fay, Richard J., Gardner, Judith, "Analytical Applications of 3-D Imaging in Vehicle Accident Studies", SAE 960648.

[14] Tumbas, Nicholas S., Kinney, J.R., Smith, G.C. "Photogrammetry & Accident Reconstruction: Experimental Results" SAE 940925.

[15] Baker, J. Stannard, "Photogrammetry for Traffic-Accident Investigation", Topic 830, Traffic Accident Investigation Manual, Northwestern University, 1985.

[16] Whitnall, Jack, "Unimpeachable Witness: The Grid", Close-Range Photogrammetry and Surveying: State-of-the Art, American Society of Photogrammetry, September 1984.

[17] Kerkhoff, John F., "Photographic Techniques for Accident Reconstruction", SAE 850248.

[18] Moffitt, Francis H.; Mikhail, Edward M., "Photogrammetry", Third Edition, Harper & Row, 1980.

[19] Williamson, James R., Brill, Michael H., "Dimensional Analysis Through Perspective - A Reference Manual", Kendall/Hunt Publishing Company, 1990.

[20] Bleyl, Robert L., "Traffic Analysis of Time-Lapse Photographs Without Employing a Perspective Grid", Traffic Engineering, August 1972.

[21] Bleyl, Robert L., "Using Photographs to Map Traffic Accident Scenes: A Mathematical Technique", Journal of Safety Research, June 1976.

[22] Niederer, Peter; Birchler, Bernhard; Mesqui, FranÅois; Lehareinger, Yves, "Computer Assisted Single-View Photogrammetry for Accident Scene Documentation", SAE 850067.

[23] Brelin, Janet M.; Cichowski, William G.; Holcomb, Michael P., "Photogrammetric Analysis Using the Personal Computer", SAE 861416.

[24] Kinney, J. Rolly; Magedanz, Bill, "TRANS4- A Traffic Accident Photogrammetric System, Description of the System and Its Inherent Errors", SAE 861417.

[25] Smith, Gregory C.; Allsop, Douglas "L", "A Case Comparison of Single-Image Photogrammetry Methods", SAE 890737.

[26] Pepe, M.D., Grayson, R., McClary, A. "Digital Rectification of Reconstruction Photographs", SAE 961049.

[27] Rohde, Ron., "An Introduction to Desktop Photogrammetry", Accident Investigation Quarterly, pgs.16-23, Summer 1995.

[28] PC-Rect is a Windows©-based stand-alone photogrammetry program developed by Andreas Moser and Louis Hohl at the Technical University of Graz, Austria.

[29] Moffitt, Francis H.; Mikhail, Edward M.; "Photogrammetry", Harper & Row, New York, 3rd Ed., 1980.

[30] Wolf, Paul R.; "Elements of Photogrammetry", McGraw-Hill, 2nd Ed., 1983.

[31] Strang, Gilbert, "Linear Algebra and its Applications", Academic Press, 2nd Ed., 1980.

[32] Grimes, Wesley D.; Culley, Charles H.; Cromack, J. Robert, "Field Application of Photogrammetric Analysis Techniques: Applications of the FOTOGRAM Program", SAE 861418.

[33] Hatzopoulos, John N., "An Analytical System for Close-Range Photogrammetry", Photogrammetric Engineering and Remote Sensing, Vol. 51, No. 10, Oct 1985, pp. 1583-1588.

[34] PC-CRASH is a Windows©-based accident-reconstruction program developed by Dr. Hermann Steffan, Andreas Moser and Wolfgang Neubauer at the Technical University of Graz, Austria.

APPENDIX A

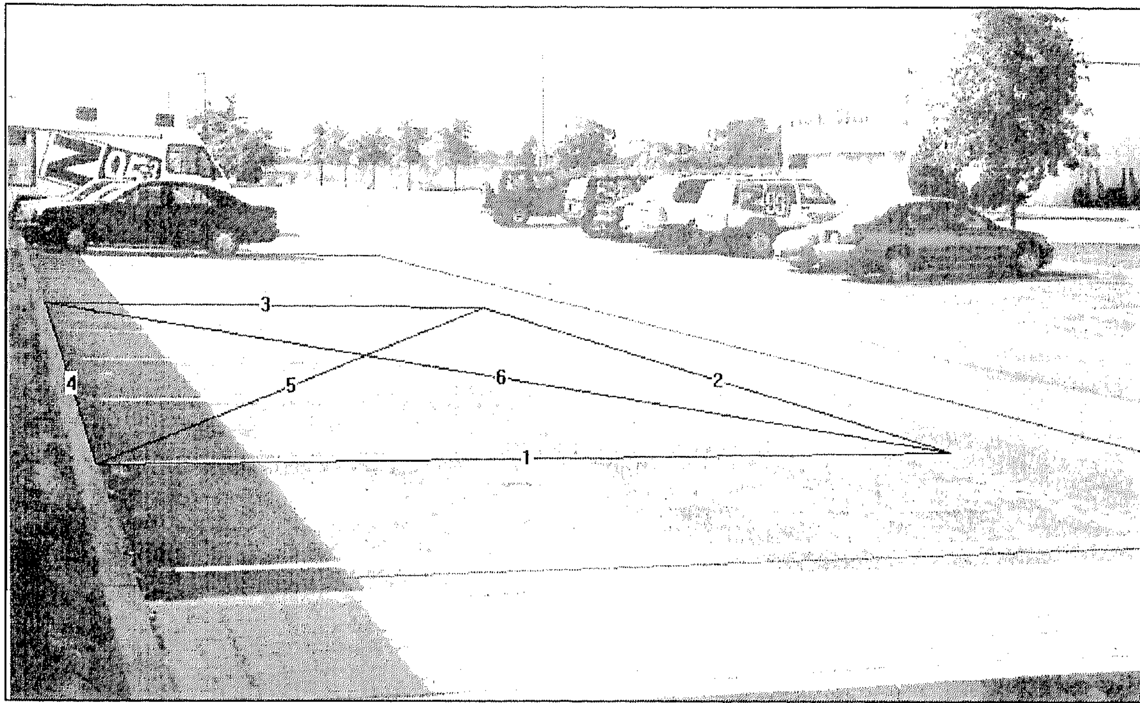
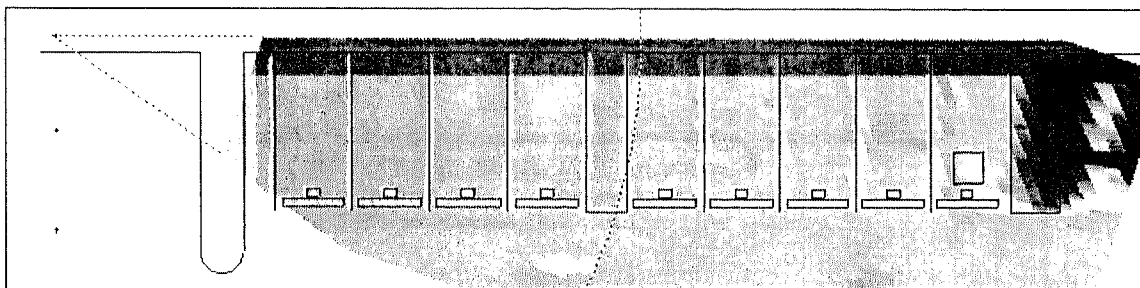


Image 1a


Image 1b

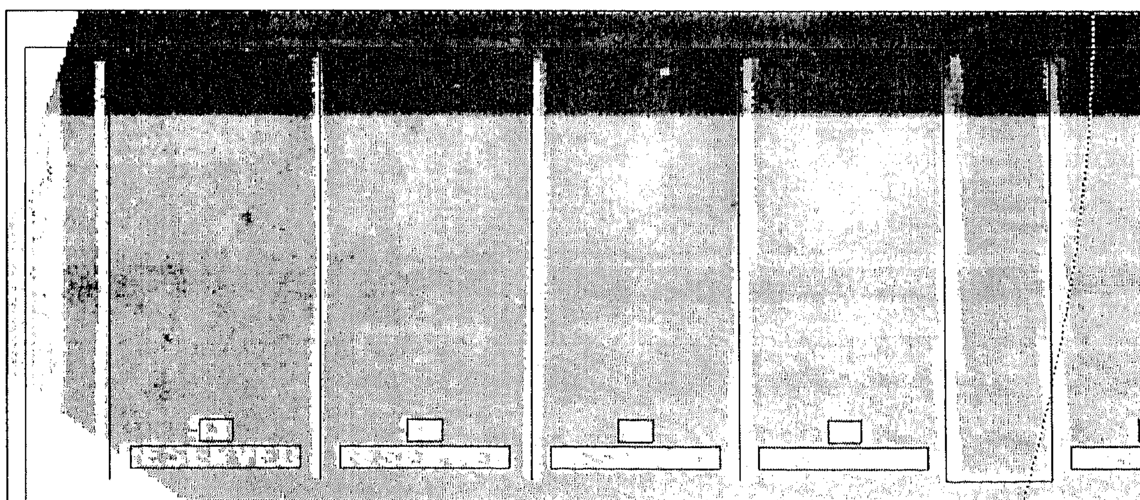
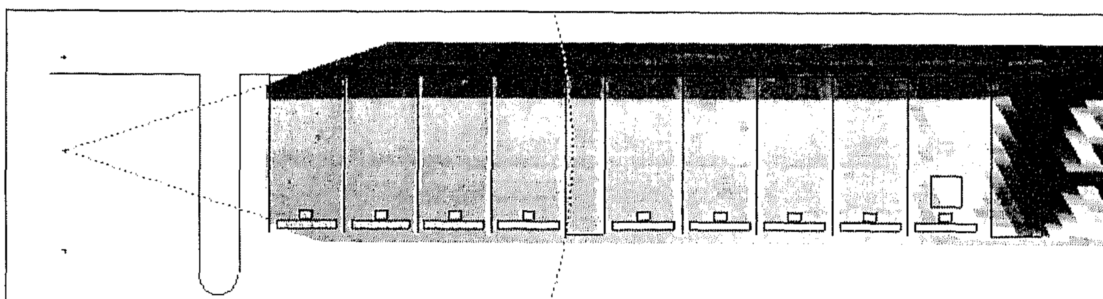

Image 1c



Image 2a


Image 2b

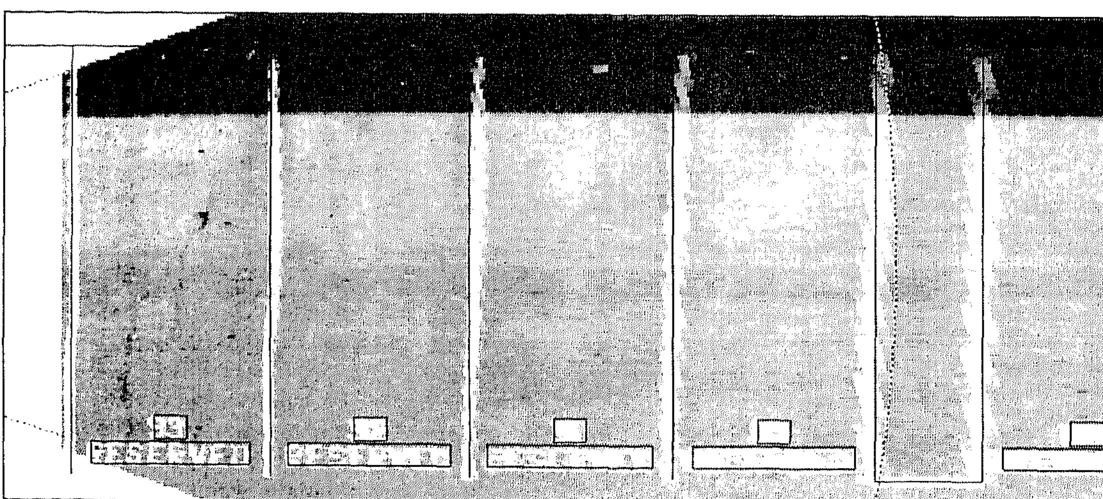
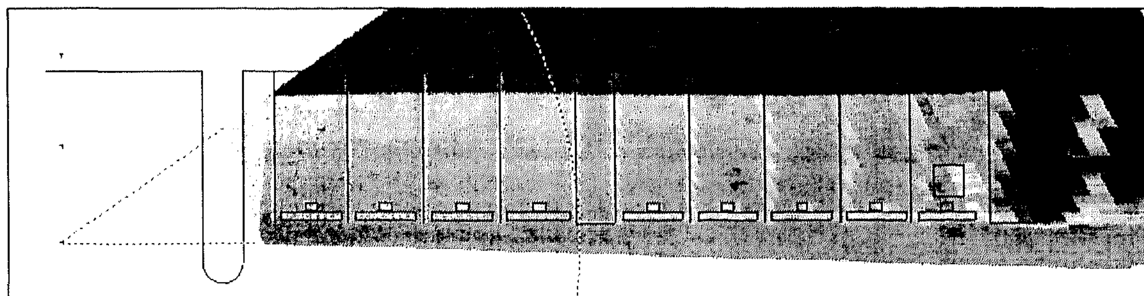

Image 2c



Image 3a


Image 3b

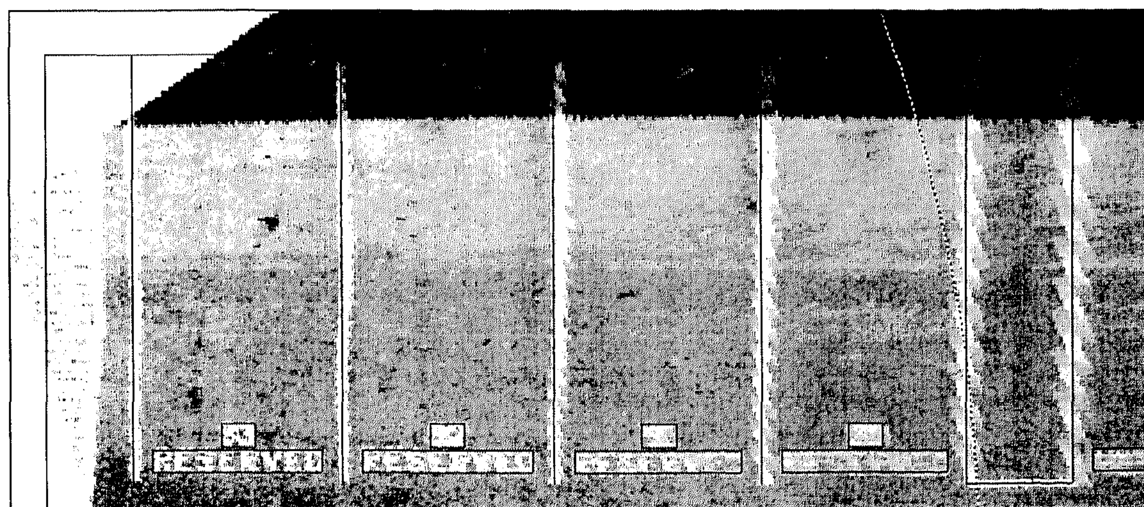
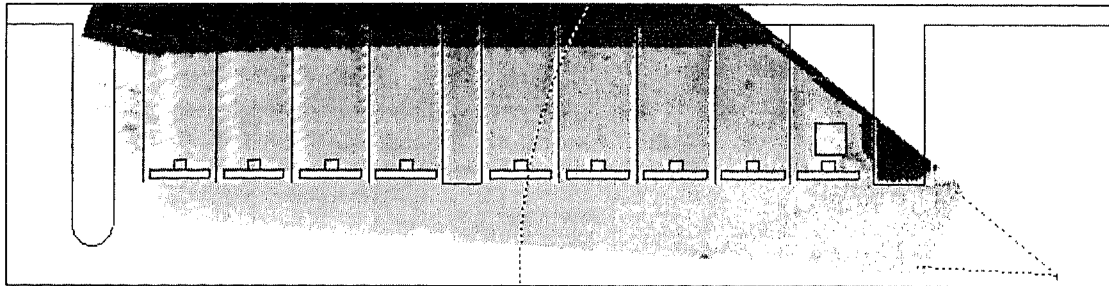
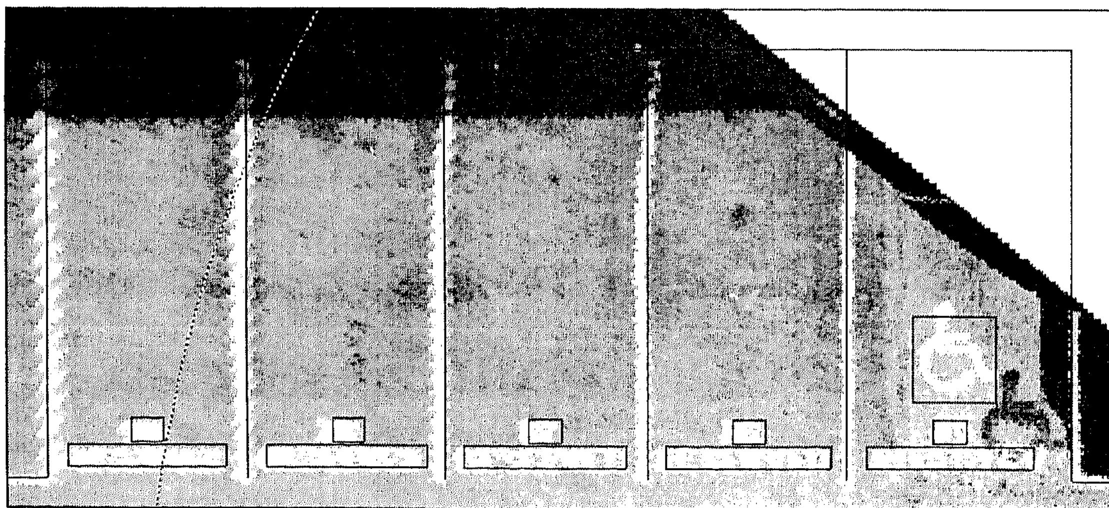

Image 3c



Image 4a


Image 4b


Image 4c

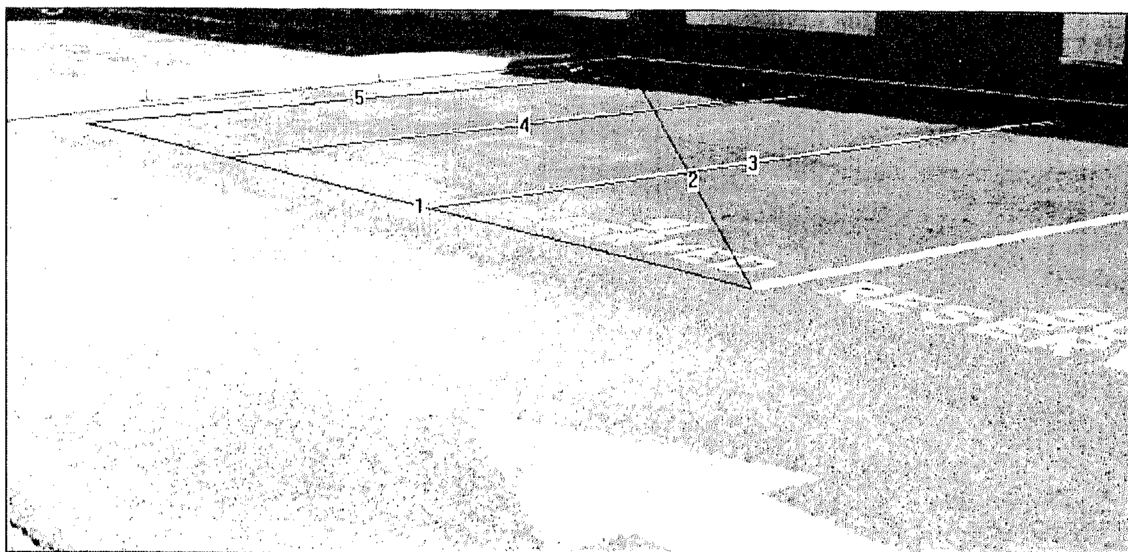


Image 5a

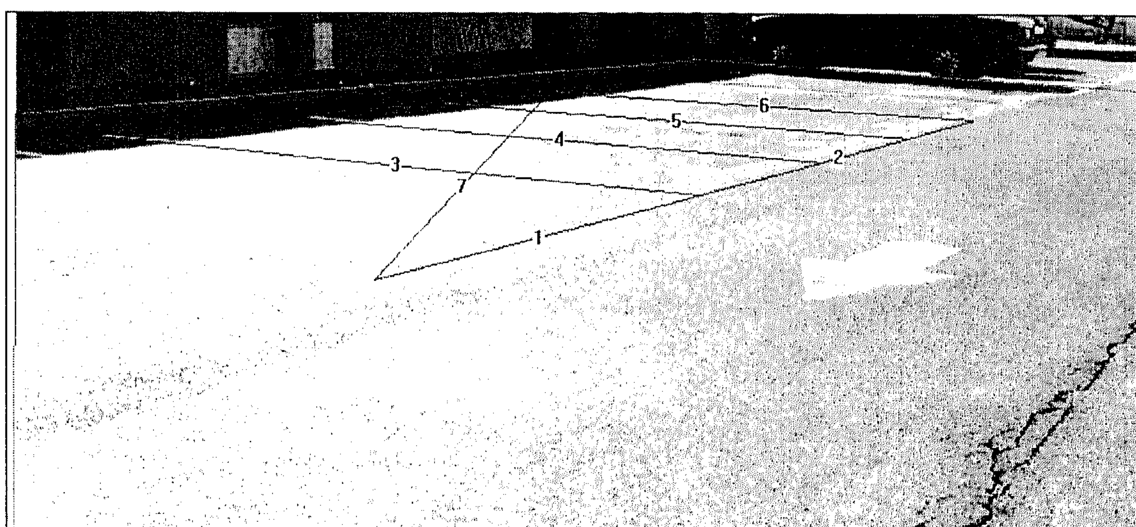
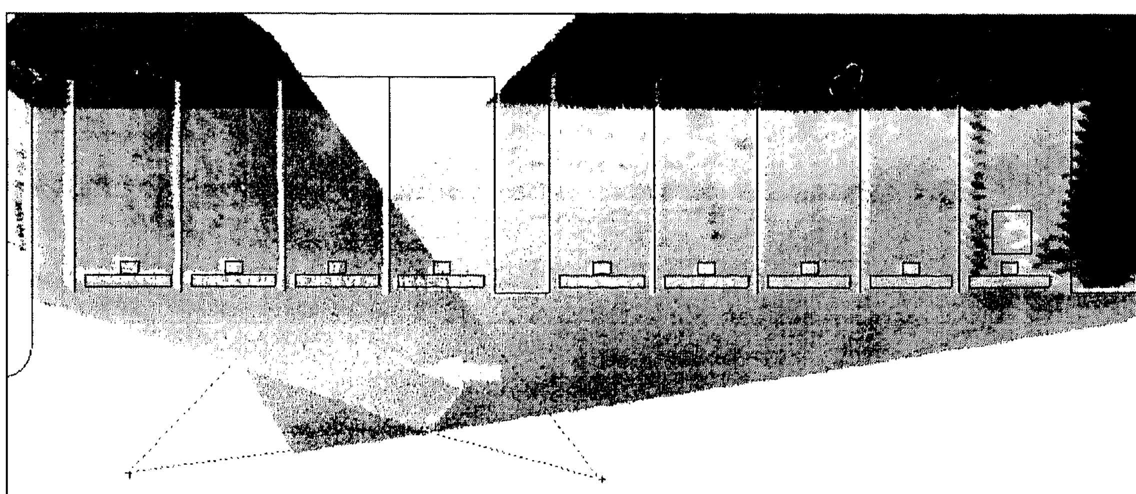


Image 5b


Image 5c

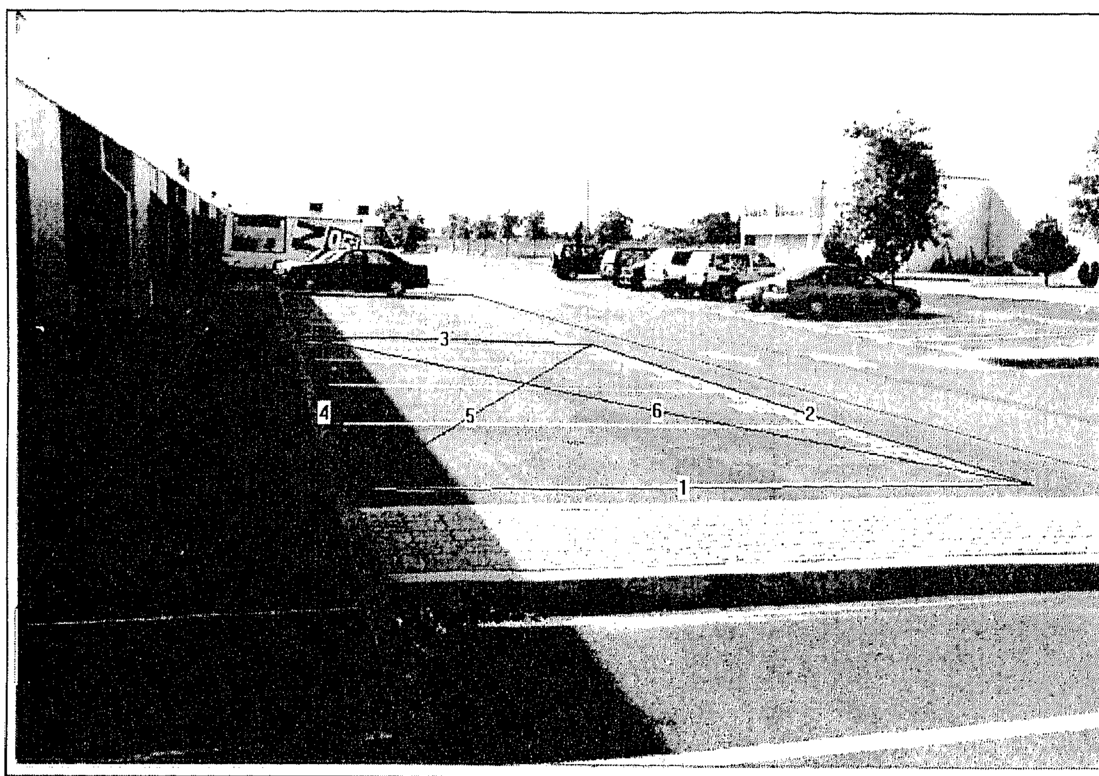
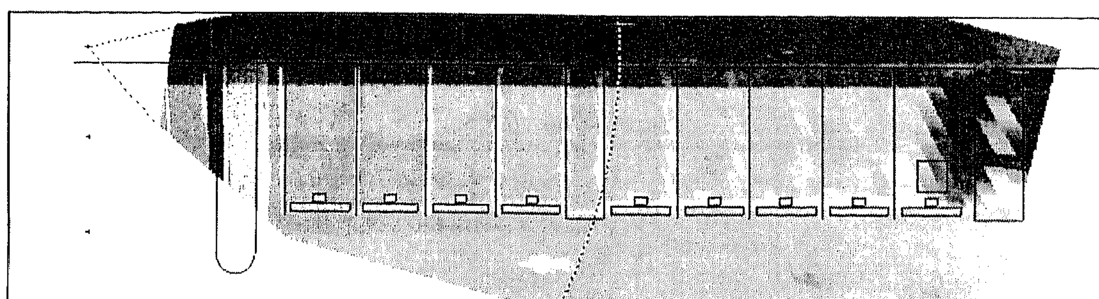
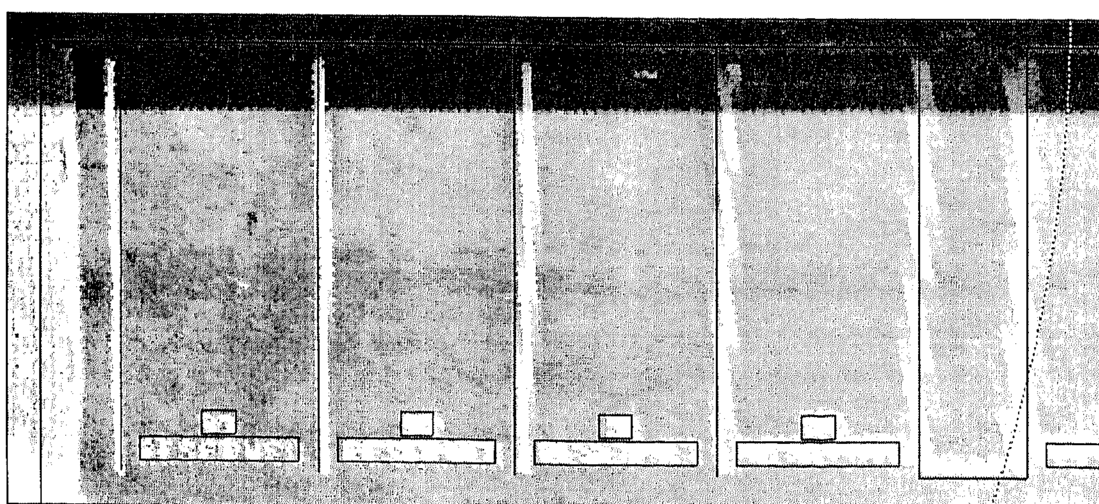


Image 6a


Image 6b


Image 6c

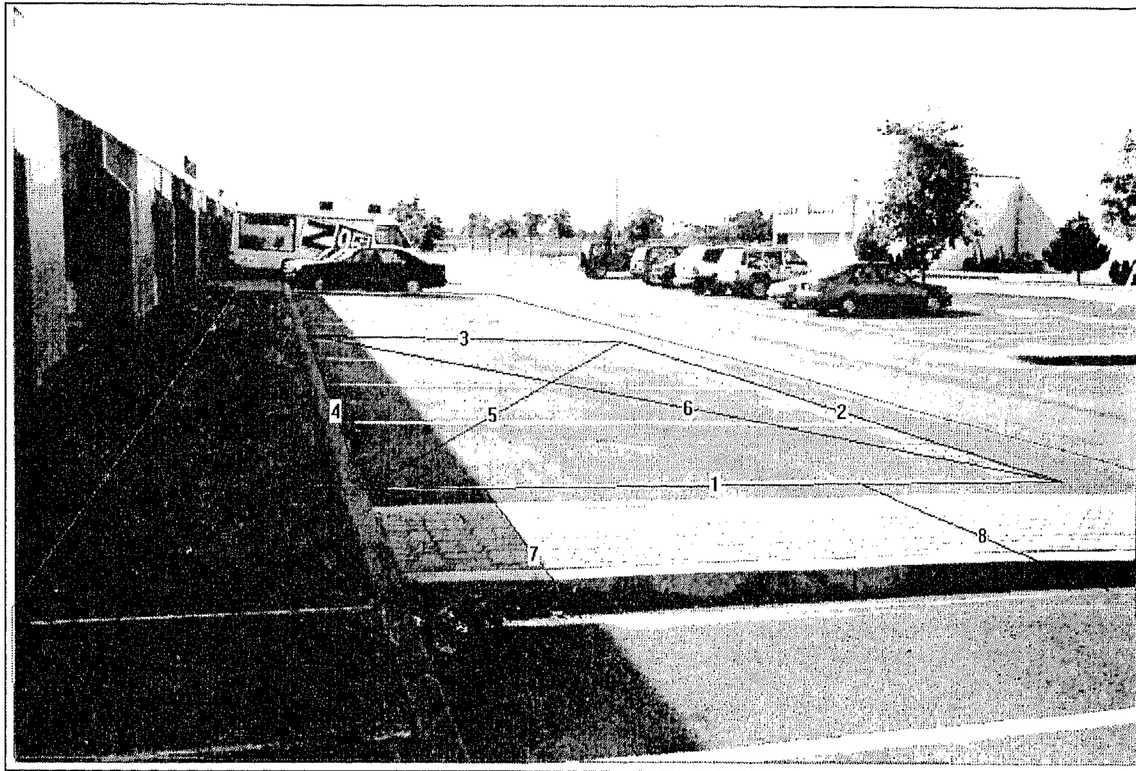
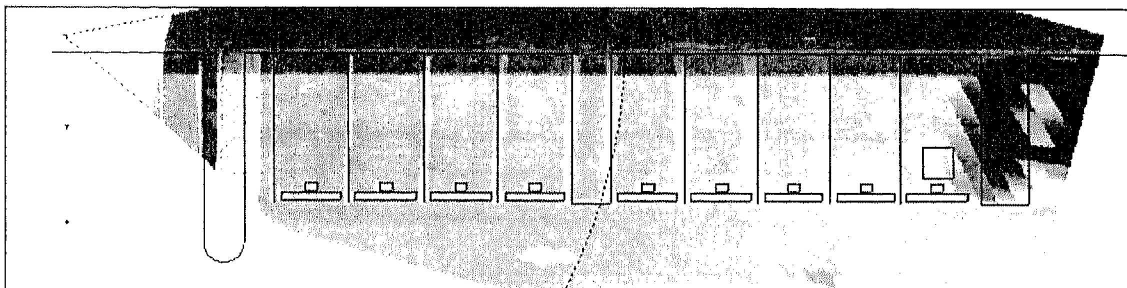


Image 7a


Image 7b

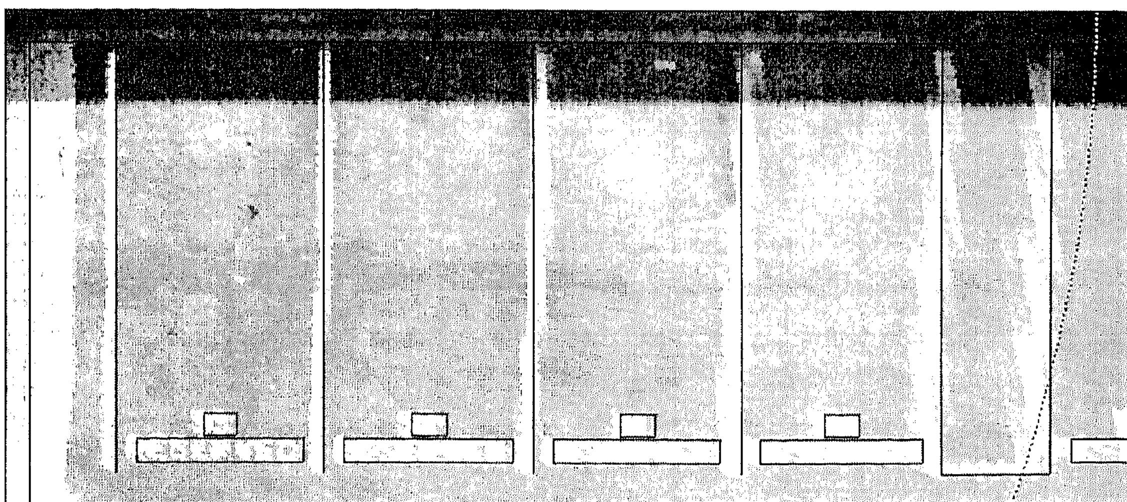
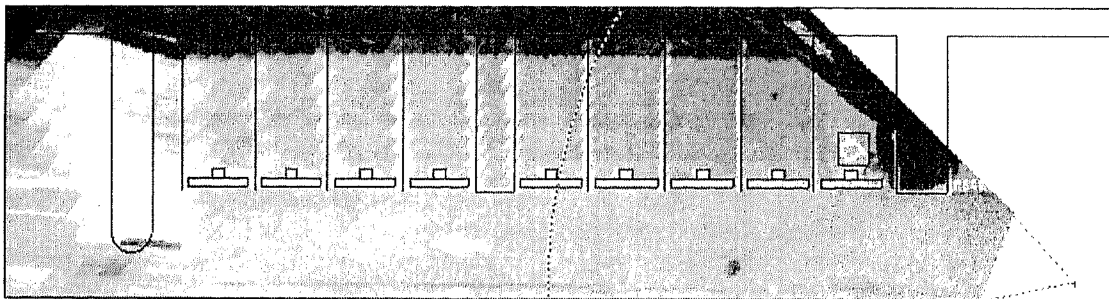
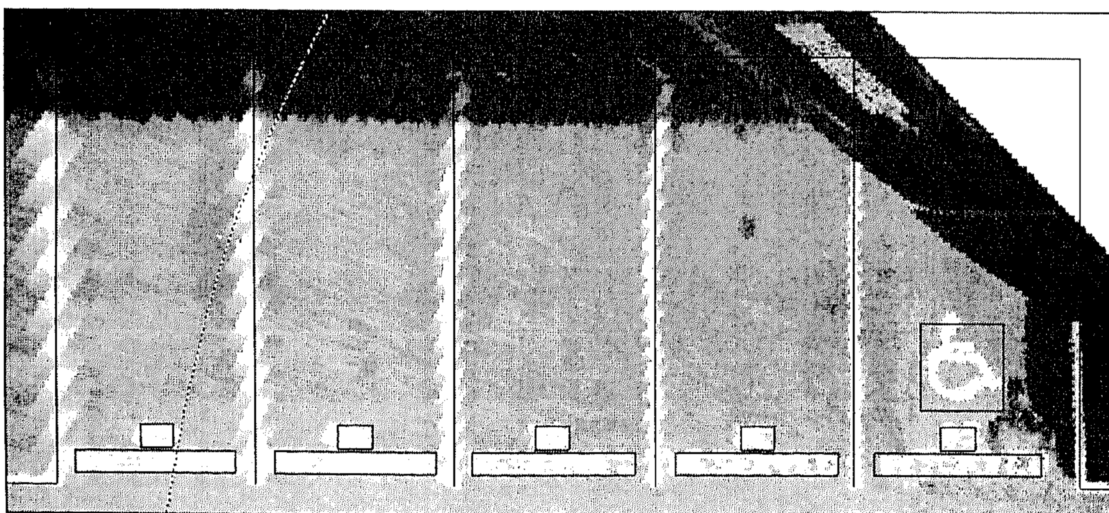

Image 7c



Image 8a


Image 8b


Image 8c

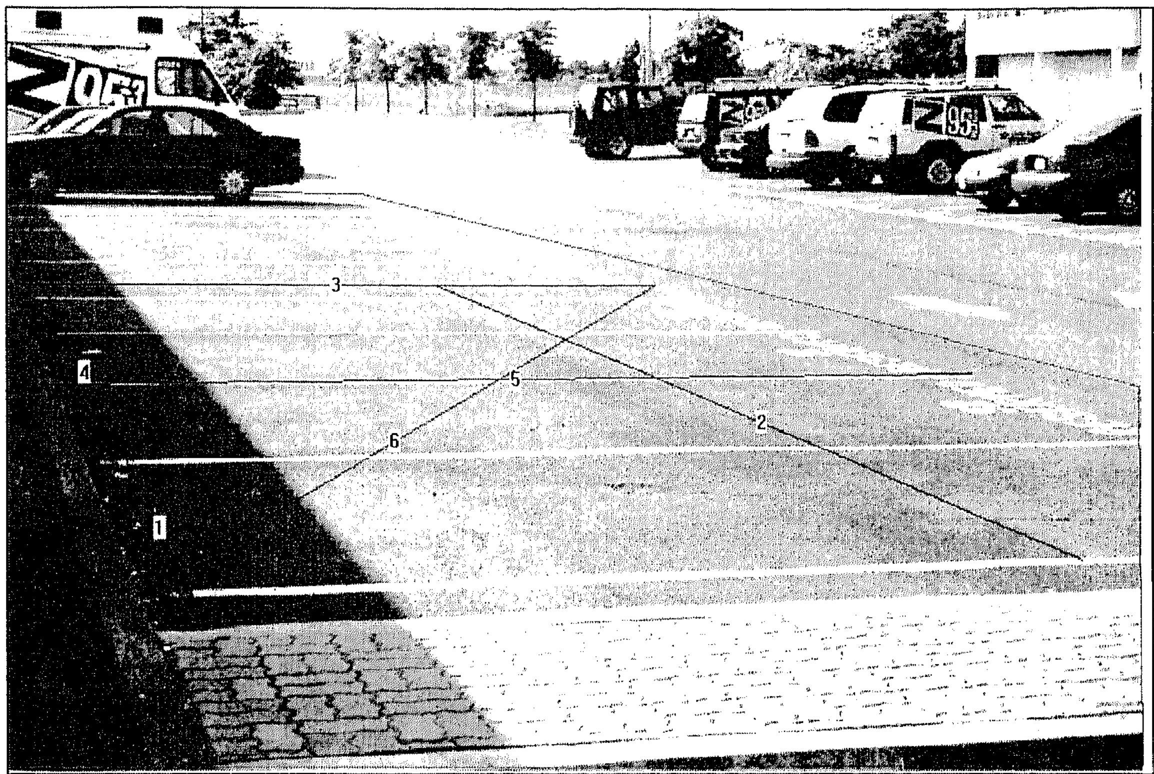
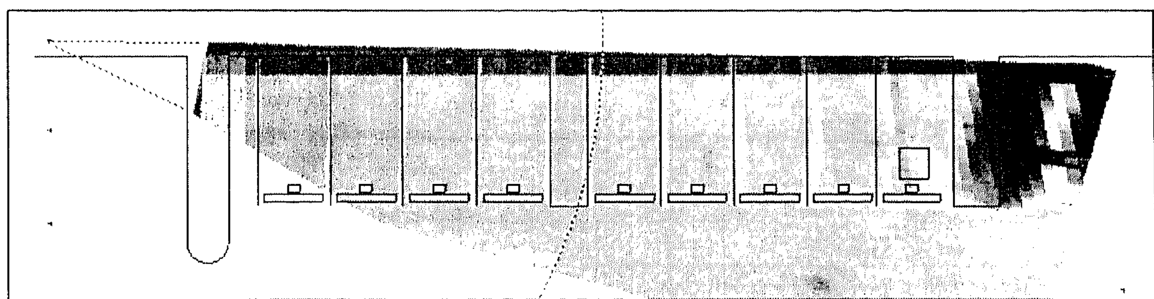
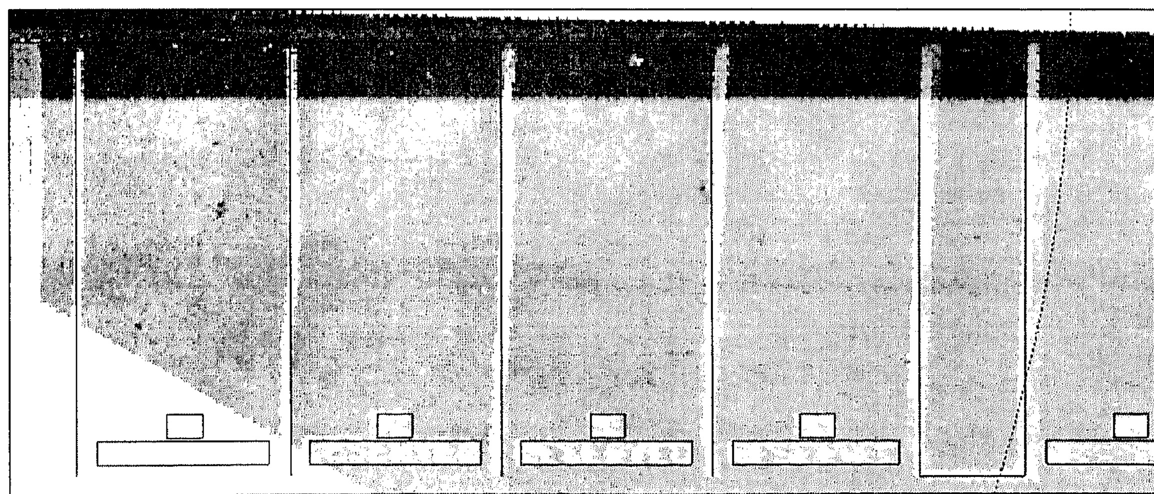


Image 9a


Image 9b


Image 9c

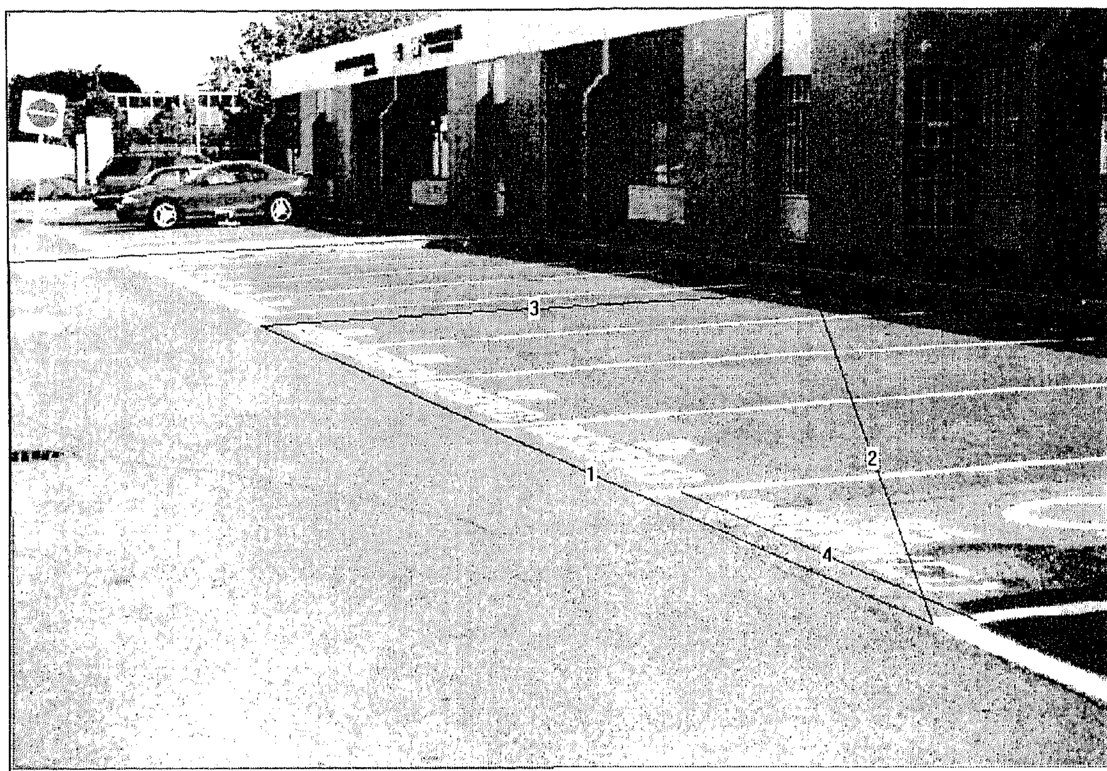
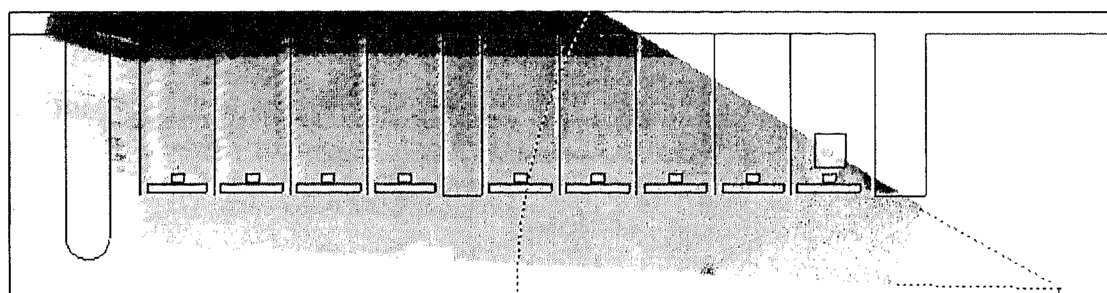
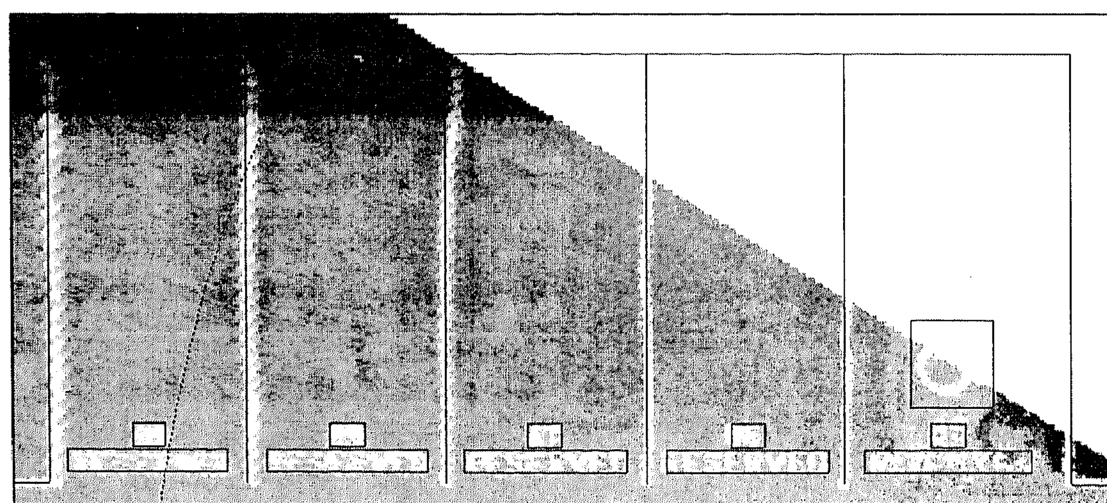


Image 10a


Image 10b


Image 10c

APPENDIX B

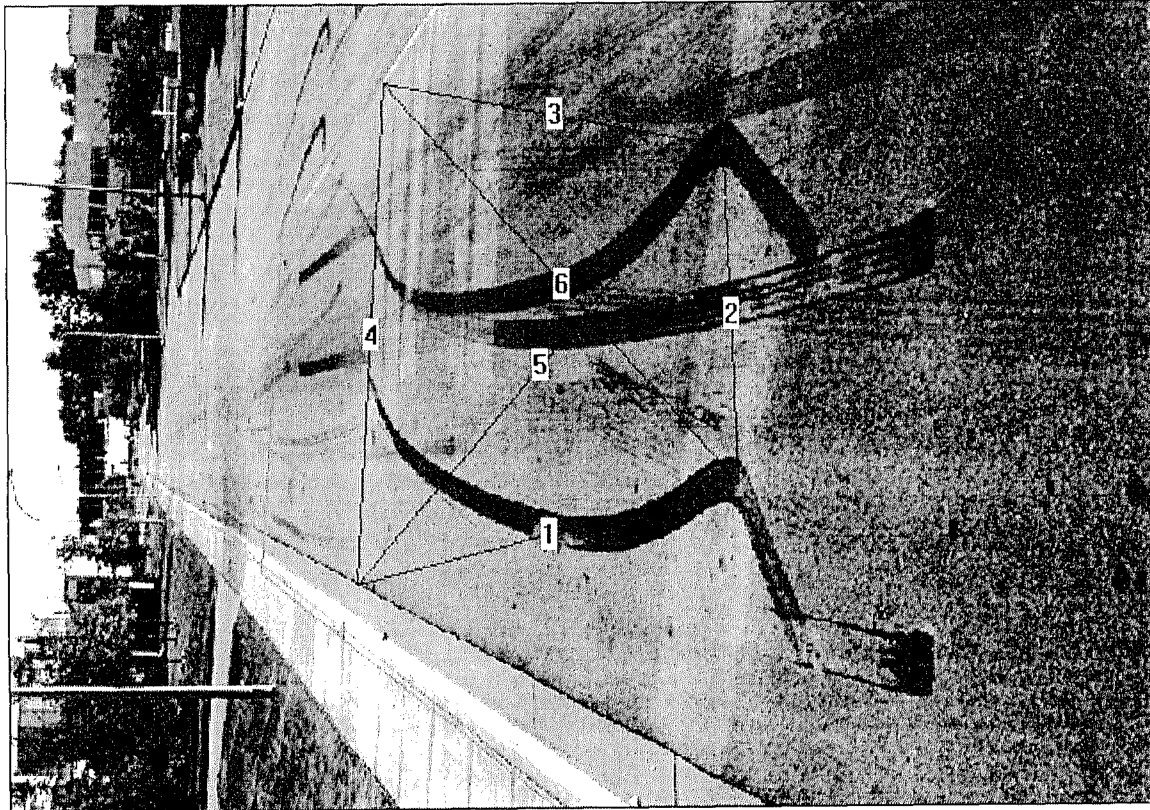


Image 11a

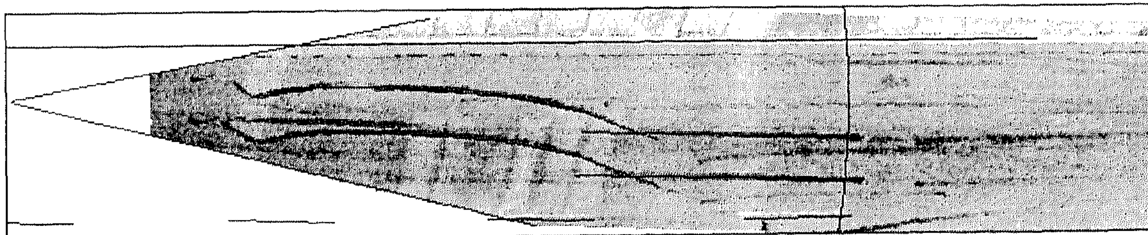


Image 11b

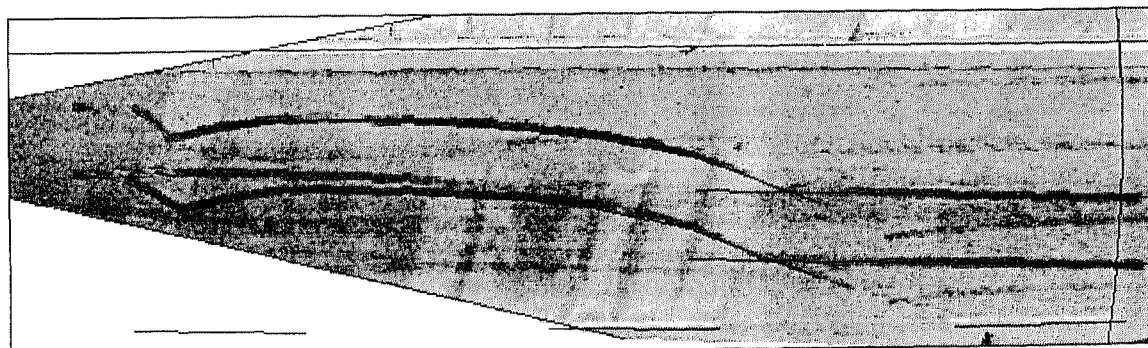


Image 11c

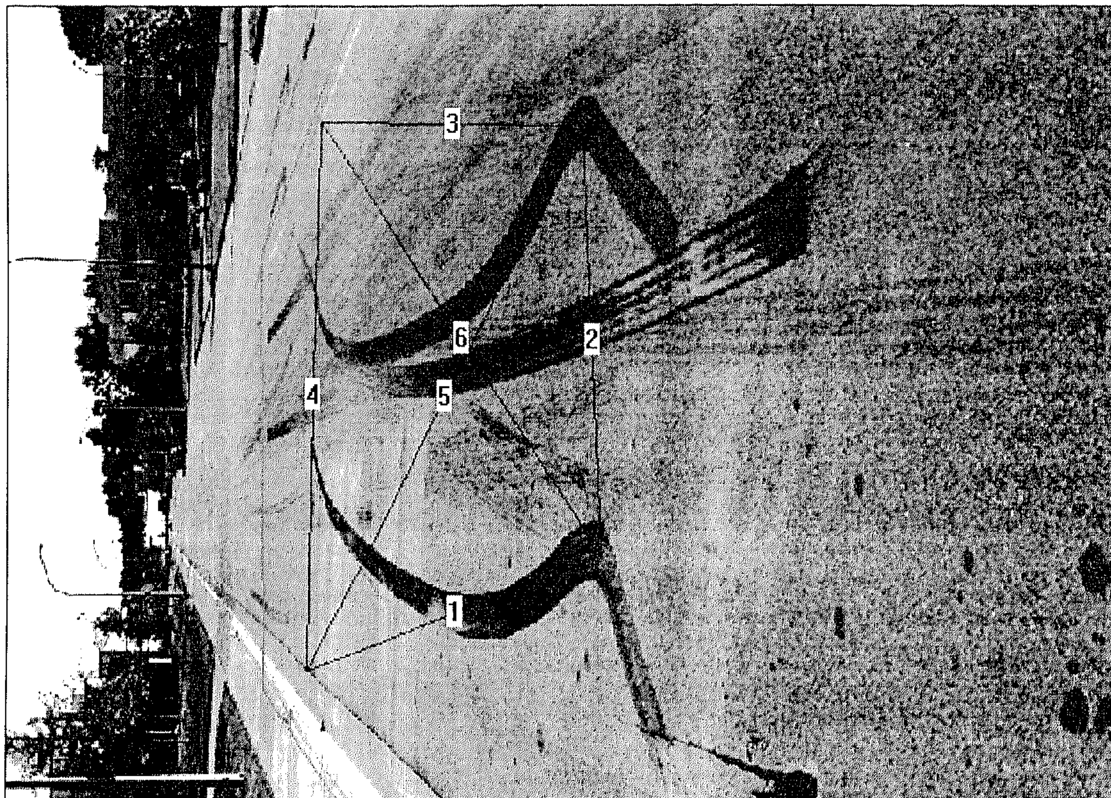
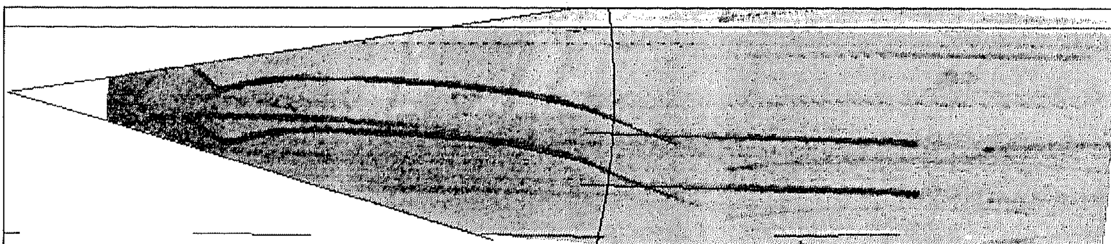
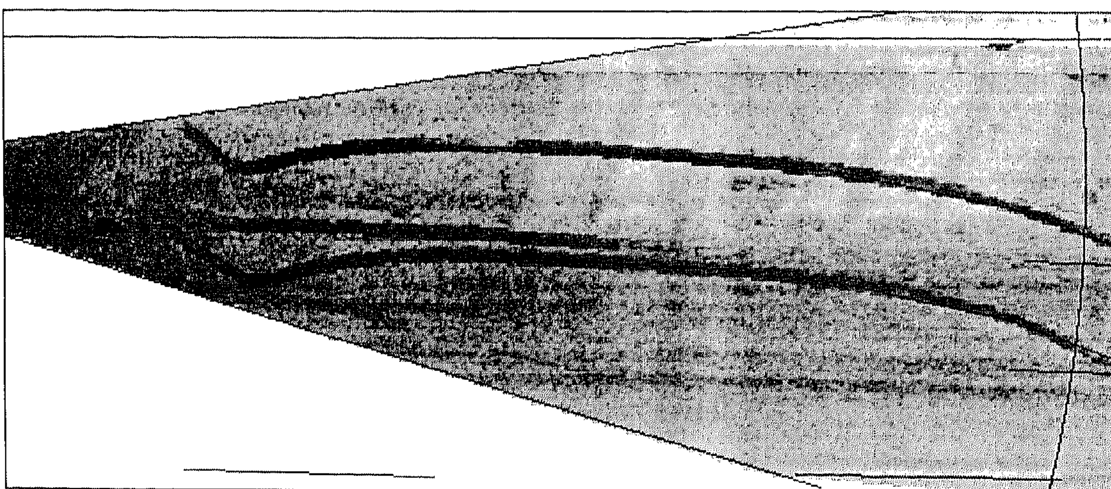


Image 12a


Image 12b


Image 12d

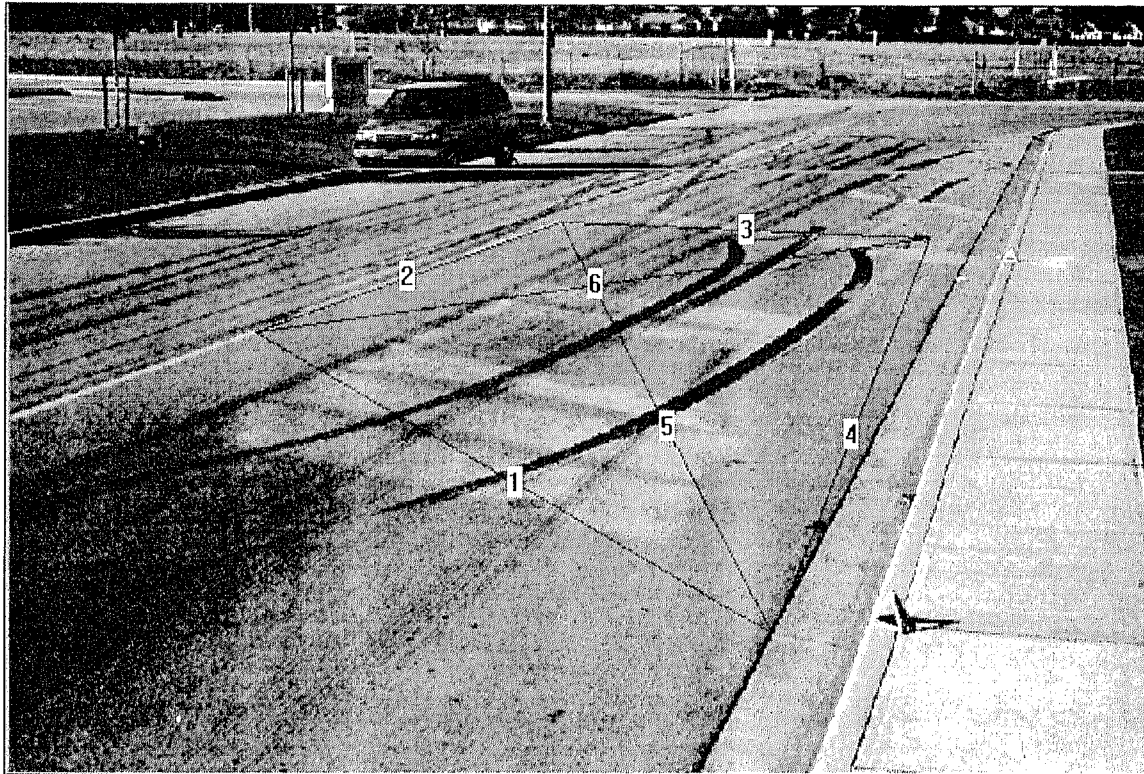
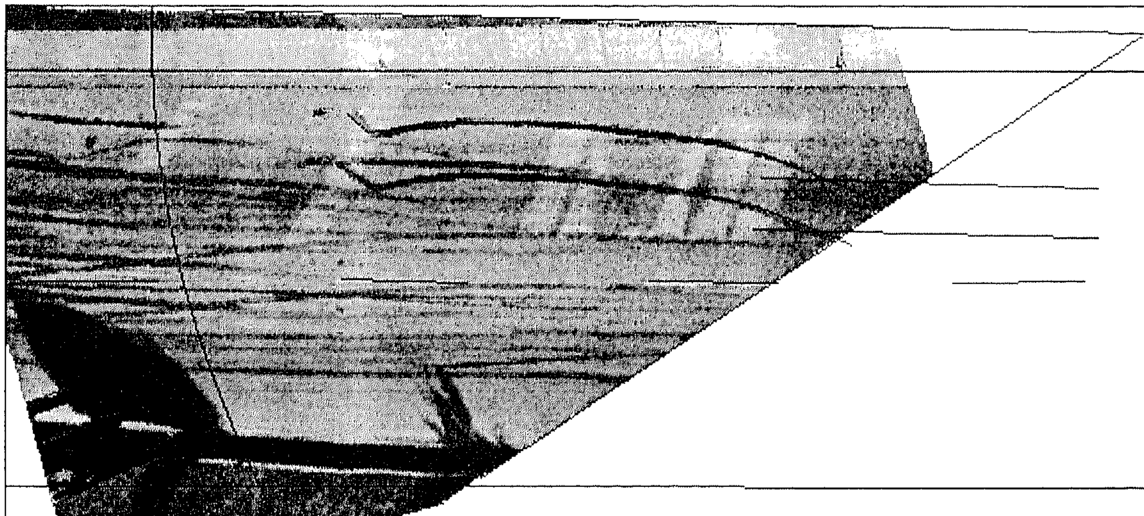


Image 13a


Image 13b

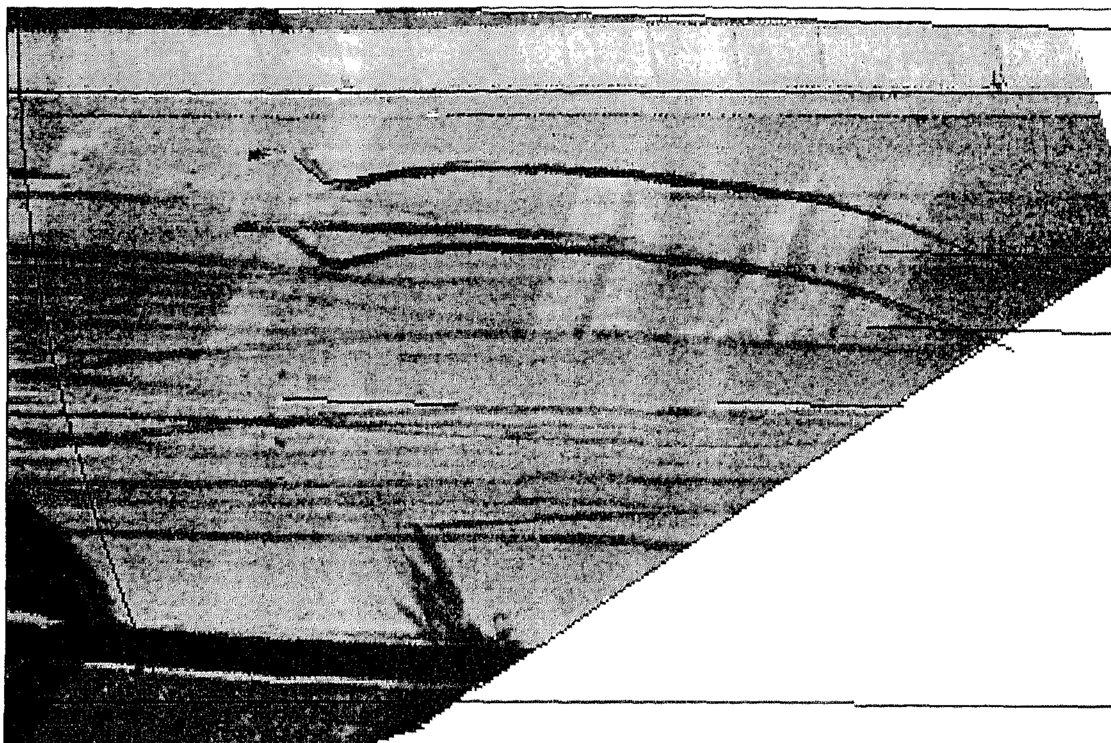


Image 13c

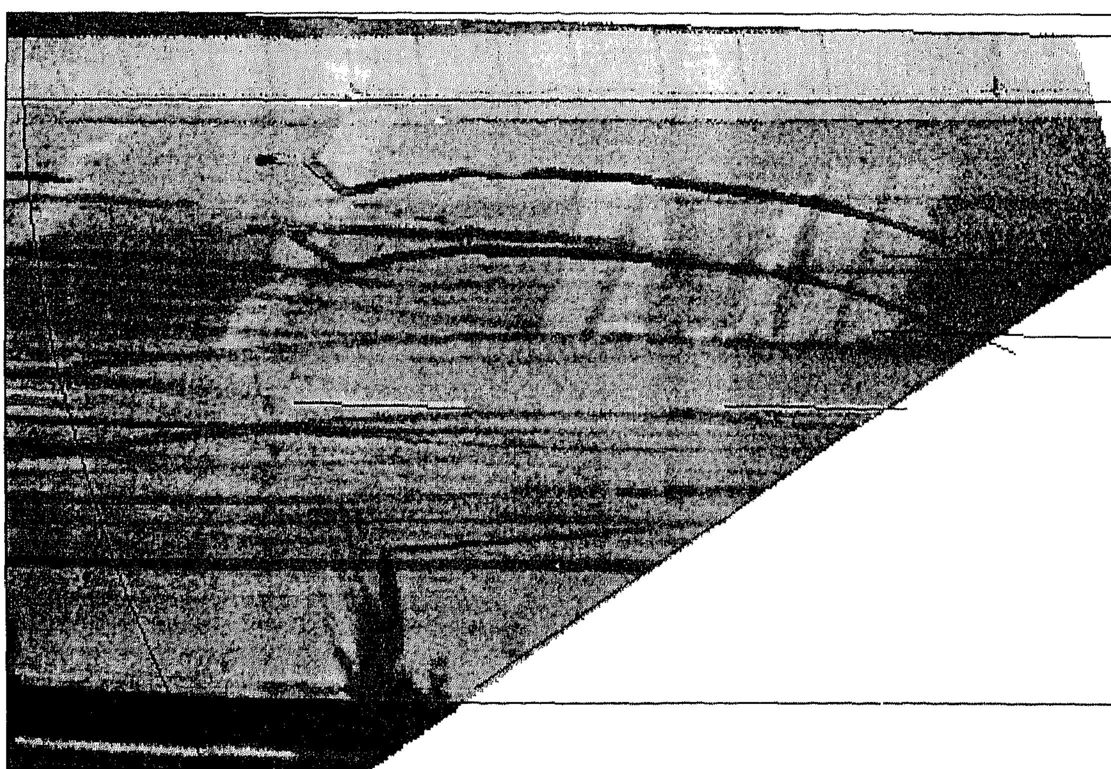


Image 13d

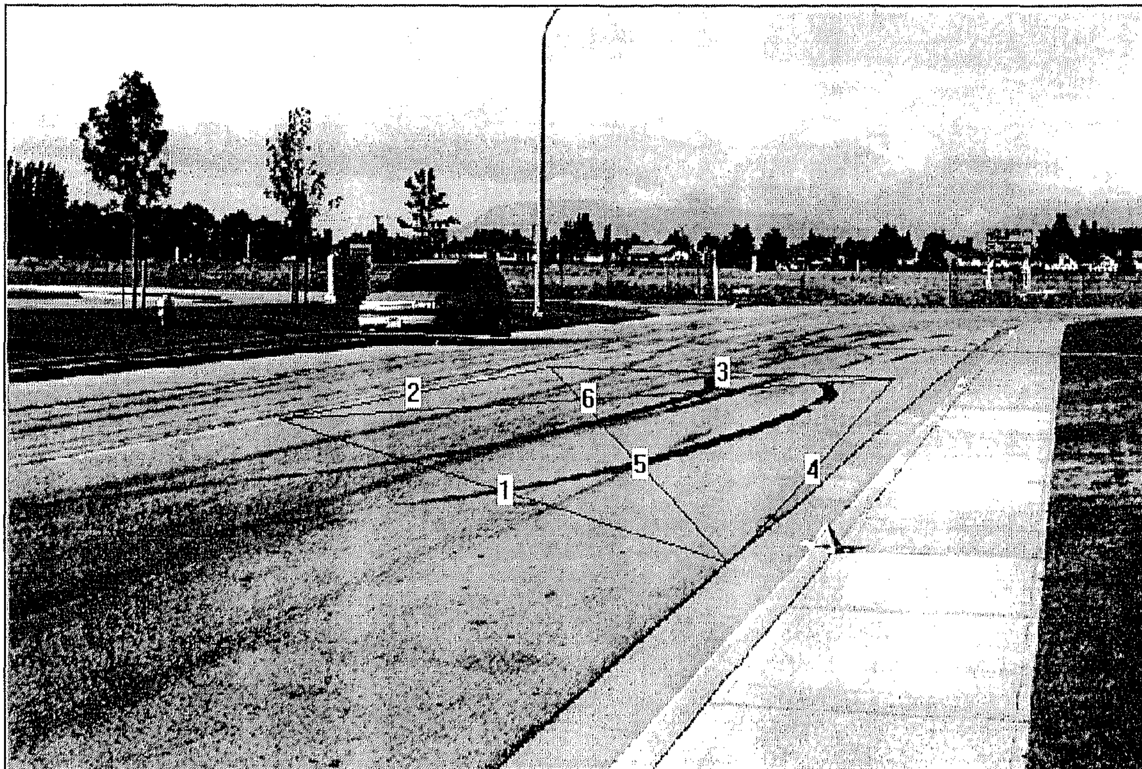


Image 14a

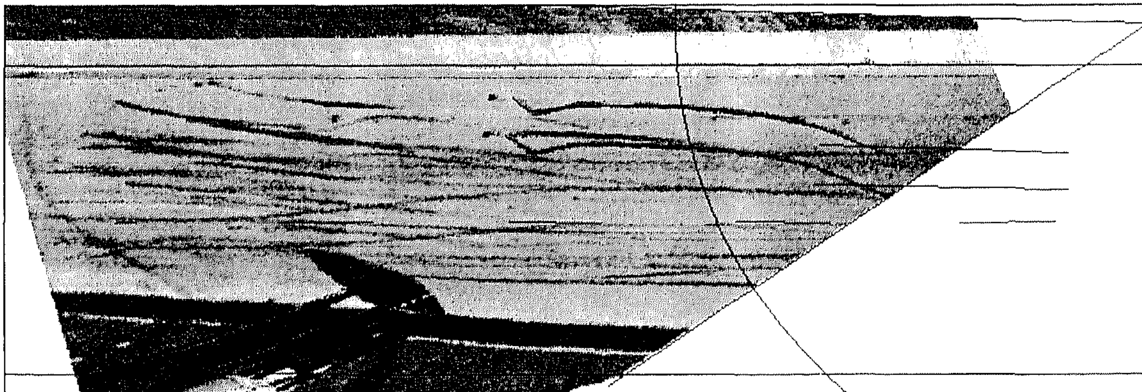

Image 14b



Image 14c

APPENDIX C - 3D and 2D Theory

The generalized 3D case is developed first, then simplified for the 2D "co-planarity" condition.

EXTERIOR ORIENTATION

The first task in any analytical photogrammetric project is determination of the properties of the camera, such as the geometric distortion, the precise focal length, and the coordinates of the principal point (the point where the optical axis of the camera passes through the film plane). This is termed interior orientation, and a camera with these properties known is termed a "metric" camera. For the mathematics that follow, a perfect camera is assumed: distortion-free with known focal length and with the optical axis aligned with the center of the image, or photograph.

The second task is to locate the camera exposure center , *O*, (also called the focal point) in space. This is termed exterior orientation.

Exterior orientation consists of locating the camera exposure center on the global coordinate system, T_x, T_X, and T_Y, determination of the alignment of the camera optical axis, ω, ϕ, and k,and determination of the scale factor, s, between the image and the object space (the ground plane in our case).

NOMENCLATURE

The following nomenclature will be used throughout this paper:

- X_J, Y_J, Z_J = coordinates of object point J on global frame X, Y, Z
- x_j, y_j, z_j = coordinates of photo point j on camera frame x, y, z, origin at exposure center, 0.
- x'_j, y'_j, z'_j = coordinates of photo point j on frame x', y', z', parallel to X, Y, Z, origin at exposure center, 0.
- T_X, T_Y, T_Z = coordinates of exposure center, 0, on global frame
- ω, ϕ, k, = direction of optical axis from exposure center, 0,
- s = scale between the image and the object space.
- h = camera height.

These parameters are shown in Figures C1 and C2.

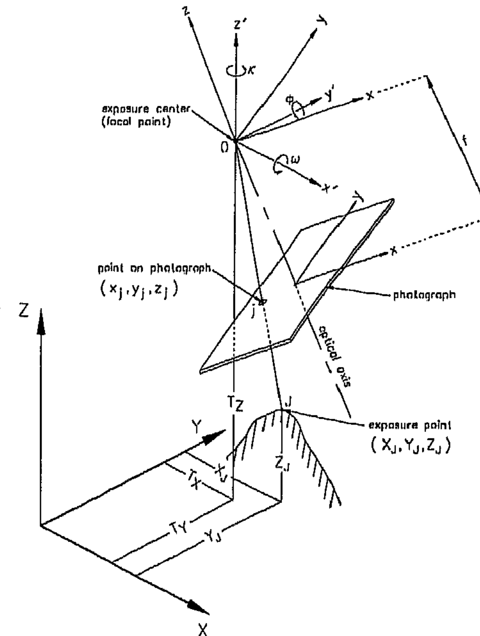


Figure C1 - The image space and the object space

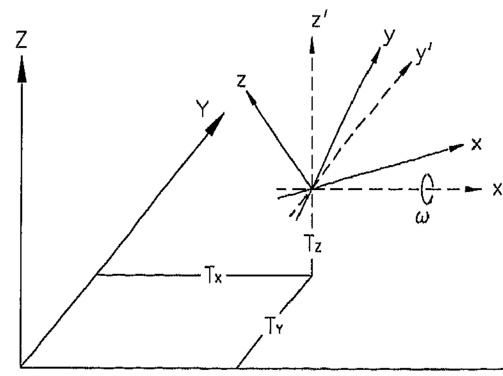


Figure C2 - The three coordinate frames of reference

ROTATIONS

The convention is to first convert (by rotations) from coordinates in the parallel frame x', y', z' to the coordinates of the photo point in the camera frame to x_j, y_j, z_j, and later, by matrix transformation (because it is x', y', z' that is usually given) convert back to x_j, y_j, z_j by $X = MX'$ (see later). These rotations take place about the three axes of the parallel frame in the order shown in Figure C3.

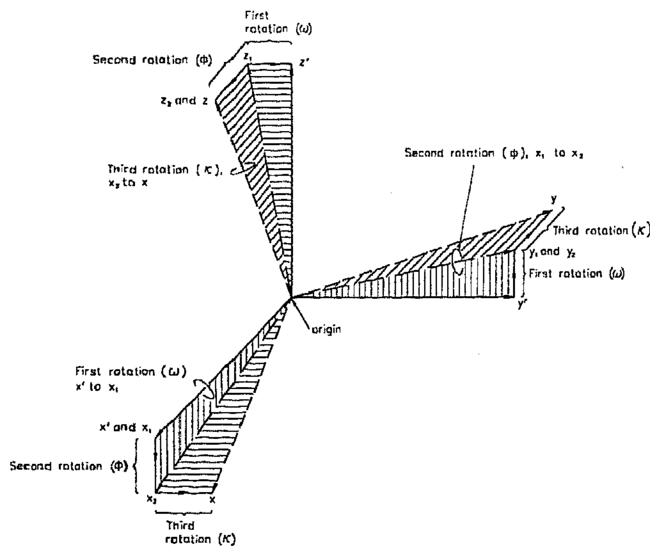


Figure C3 - Order of rotations

First, rotate through ω about x' axis, get x_1, y_1, z_1, from x', y', z', as shown in Figure C4:

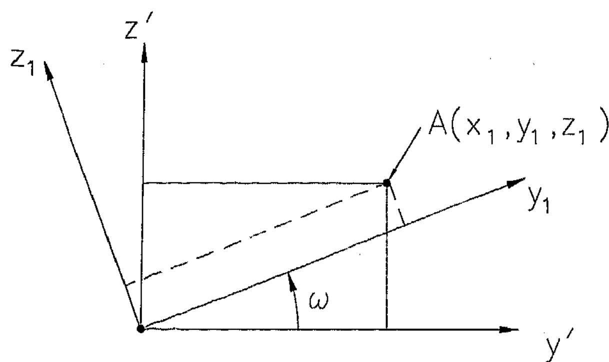


Figure C4 - Rotation about x' axis

$$\begin{aligned} x_1 &= x' \\ y_1 &= y' \cos\omega + z' \sin\omega \\ z_1 &= -y' \sin\omega + z' \cos\omega \end{aligned} \qquad \text{(C1)}$$

Second, Rotate through ϕ about y_1 axis, get x_2, y_2, z_2, as shown in Figure C5:

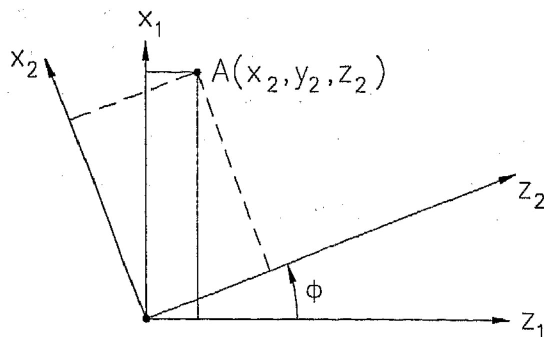


Figure C5 - Rotation about y_1 axis

$$\begin{aligned} x_2 &= x_1 \cos\phi - z_1 \sin\phi \\ y_2 &= y_1 \\ z_2 &= x_1 \sin\phi + z_1 \cos\phi \end{aligned} \qquad \text{(C2)}$$

and by substitution:

$$\begin{aligned} x_2 &= x' \cos\phi - \left(-y' \sin\omega + z' \cos\omega\right) \sin\phi \\ y_2 &= y' \cos\omega + z' \sin\omega \\ z_2 &= x' \sin\phi + \left(-y' \sin\omega + z' \cos\omega\right) \cos\phi \end{aligned} \qquad \text{(C3)}$$

Third, rotate through k about z_2 to x, y, z, as shown in Figure C6:

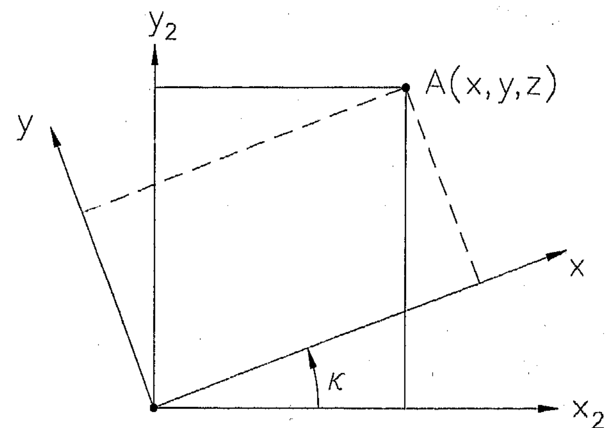


Figure C6 - Rotation about z_2 axis

$$\begin{aligned} x &= x_2 \cos k + y_2 \sin k \\ y &= -x_2 \sin k + y_2 \cos k \\ z &= z_2 \end{aligned} \qquad \text{(C4)}$$

and substitute again:

$$\begin{aligned} x &= \left(x' \cos\phi - \left(-y' \sin\omega + z' \cos\omega\right) \sin\phi\right) \cos k \\ &+\left(y' \cos\omega + z' \sin\omega\right) \sin k \\ y &= -\left(x' \cos\phi - \left(-y' \sin\omega + z' \cos\omega\right) \sin\phi\right) \sin k \\ &+\left(y' \cos\omega + z' \sin\omega\right) \cos k \\ z &= x' \sin\phi + \left(-y' \sin\omega + z' \cos\omega\right) \cos\phi \end{aligned}$$

(C5)

Expanding and factoring:

$$\begin{aligned} x &= x'\left(\cos\phi \cos k\right) - \sin\phi \cos k\left(-y' \sin\omega + z' \cos\omega\right) \\ &\quad + y'(\cos\omega \sin k) + z'(\sin\omega \sin k) \\ &= x'\left(\cos\phi \cos k\right) - y'\left(-\sin\omega \sin\phi \cos k\right) \\ &\quad - z'\left(\cos\omega \sin\phi \cos k\right) \\ &\quad + y'(\cos\omega \sin k) z'(\sin\omega \sin k) \end{aligned}$$

$$= (\cos\phi\cos k)x' + (\cos\omega\sin k + \sin\omega\sin\phi\cos k)y' + (\sin\omega\sin k - \cos\omega\sin\phi\cos k)z' \quad \text{(C6)}$$

By similar means, the relationships for y and z can be found.

These rotations can be expressed in matrix form by substituting m_{ij} for the coefficients of x', y', z'

$$\begin{aligned} x &= m_{11}x' + m_{12}y' + m_{13}z' \\ y &= m_{21}x' + m_{22}y' + m_{23}z' \\ z &= m_{31}x' + m_{32}y' + m_{33}z' \end{aligned} \quad \text{(C7)}$$

where,

$$\begin{aligned} m_{11} &= \cos\phi\cos k \\ m_{12} &= \cos\omega\sin k + \sin\omega\sin\phi\cos k \\ m_{13} &= \sin\omega\sin k - \cos\omega\sin\phi\cos k \end{aligned} \quad \text{(C7)}$$

and, proven similarly,

$$\begin{aligned} m_{21} &= -\cos\phi\sin k \\ m_{22} &= -\sin\omega\sin\phi\sin k + \cos\omega\cos k \\ m_{23} &= \cos\omega\sin\phi\sin k + \sin\omega\cos k \\ m_{31} &= \sin\phi \\ m_{32} &= -\sin\omega\cos\phi \\ m_{33} &= \cos\omega\cos\phi \end{aligned} \quad \text{(C8)}$$

In matrix form

$$X = MX' \quad \text{(C9)}$$

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix},\ M = \begin{bmatrix} m_{11}m_{12}m_{13} \\ m_{21}m_{22}m_{23} \\ m_{31}m_{32}m_{33} \end{bmatrix},\ X' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$$

Recall the convention is to go from photo points on parallel axes (x', y', z') to photo points on rotated axes (x, y, z). Because M is an orthogonal Matrix,

$$M^{-1} = M^T \quad \text{(C10)}$$

$$\therefore\ X' = M^T X \quad \text{(C11)}$$

Yielding the relationship for going from general parallel axes to photo points (i.e. transposed $m_{ij} = m_{ji}$)

$$\begin{aligned} x' &= m_{11}x + m_{21}y + m_{31}z \\ y' &= m_{12}x + m_{22}y + m_{32}z \\ x' &= m_{13}x + m_{23}y + m_{33}z \end{aligned} \quad \text{(C12)}$$

This is the general relationship to go back from the rotated axes (x, y, z) to the parallel axes, (x', y', z').

SCALING AND TRANSLATION

To get the coordinates of the photo points in X, Y, Z, the global coordinate system, we must deal with the position of the exposure center and the scale factor.

$$\begin{aligned} X &= sx' + T_X = s(m_{11}x + m_{21}y + m_{31}z) + T_X \\ Y &= sy' + T_Y = s(m_{12}x + m_{22}y + m_{32}z) + T_Y \\ Z &= sz' + T_Z = s(m_{13}x + m_{23}y + m_{33}z) + T_Z \end{aligned} \quad \text{(C13)}$$

Or, in matrix form:

$$\overline{X} = SM^T X + T \quad \text{(C14)}$$

Where

$$\overline{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix},\ T \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

This can read as the object points equals the scaled rotated photo points plus the translations,

SOLVING FOR SEVEN UNKNOWNS

Recall that we have seven unknowns, the three translations T_X, T_Y, and T_Z, the three rotations ω, ϕ, and k, and the scale factor s.

Each point on the photograph whose global coordinates are known will yield three equations. Thus, since we need at least seven equations, we need at least three points whose coordinates are known in both systems. We will for convenience label these points p, q, r in the photograph, and P, Q, R in the global system.

$$\begin{aligned} X_P &= s(m_{11}x_p + m_{21}y_p + m_{31}z_p) + T_X \\ Y_P &= s(m_{12}x_p + m_{22}y_p + m_{32}z_p) + T_Y \\ Z_P &= s(m_{13}x_p + m_{23}y_p + m_{33}z_p) + T_Z \\ X_Q &= s(m_{11}x_q + m_{21}y_q + m_{31}z_q) + T_X \\ Y_Q &= s(m_{12}x_q + m_{22}y_q + m_{32}z_q) + T_Y \\ Z_Q &= s(m_{13}x_q + m_{23}y_q + m_{33}z_q) + T_Z \\ X_R &= s(m_{11}x_r + m_{21}y_r + m_{31}z_r) + T_X \\ Y_R &= s(m_{12}x_r + m_{22}y_r + m_{32}z_r) + T_Y \\ Z_R &= s(m_{13}x_r + m_{23}y_r + m_{33}z_r) + T_Z \end{aligned} \quad \text{(C15)}$$

These equations are non-linear (degree>1)

LINEARIZATION

Using Taylor's theorem, with the angles given in radians:

$$X_P = X_{P_0} + \left(\frac{\partial X_P}{\partial s}\right)_0 ds + \left(\frac{\partial X_P}{\partial \omega}\right)_0 d\omega + \left(\frac{\partial X_P}{\partial \phi}\right)_0 d\phi$$

$$+ \left(\frac{\partial X_P}{\partial k}\right)_0 d\kappa + \left(\frac{\partial X_P}{\partial T_X}\right)_0 dT_X + \left(\frac{\partial X_P}{\partial T_y}\right)_0 dT_Y$$

$$+ \left(\frac{\partial X_P}{\partial T_Z}\right)_0 dT_Z \quad \text{(C16)}$$

A total of nine equations can be written in this form for x, y, and z for each point p, q, and r.

The first few coefficients will be:

$$a_{11}\left(\frac{\partial X_P}{\partial s}\right)$$

$$X_P = s\left(m_{11}x_p + m_{21}y_p + m_{31}z_p\right) + T_X$$

$$\therefore a_{11} = m_{11}x_p + m_{21}y_p + m_{31}z_p \quad \text{(C17)}$$

$$a_{12} = \frac{\partial X_P}{\partial \omega}$$

$$X_P = s\left((\cos\phi\cos k)x_p + (-\cos\phi\sin k)y_p + (\sin\phi)z_p\right) + T_x$$

now, because $\frac{\partial X_P}{\partial \omega} = 0$

$$a_{12} = 0 \quad \text{(C18)}$$

$$a_{13} = \frac{\partial X_P}{\partial \phi}$$

$$\therefore a_{13} = s\left[(-\sin\phi\cos k)x_p + (\sin\phi\sin k)y_p + (\cos\phi)z_p\right] \quad \text{(C19)}$$

through to

$$a_{97} = \frac{\partial Z_R}{\partial T_Z} \quad \text{(C20)}$$

The partial derivative coefficients can be expressed as a_{ij}:

$$a_{11}ds + a_{12}d\omega + a_{13}d\phi + a_{14}dk + a_{15}dTx + a_{16}dTy + A_{17}dT_z = \left[X_P - X_{P_0}\right] + V_{X_P} \quad \text{(C21)}$$

through to:

$$a_{91}ds + a_{92}d\omega + a_{93}d\phi + a_{94}d\kappa + a_{95}dT_X + a_{96}dT_Y + a_{97} = \left[Z_R - Z_{R_0}\right] + V_{Z_R} \quad \text{(C21)}$$

or

$$nAX = L + V \quad \text{(C22)}$$

where,

$A = \begin{bmatrix} a_{11} \ldots\ldots\ldots\ldots a_{17} \\ \\ a_{91} \ldots\ldots\ldots\ldots a_{97} \end{bmatrix}$, the coefficient matrix,

$X = \begin{bmatrix} ds \\ d\omega \\ d\partial \\ dk \\ dT_X \\ dT_Y \\ dT_Z \end{bmatrix}$, the unknowns,

$L = \begin{bmatrix} X_P - X_{P_0} \\ \\ Z_R - Z_{R_0} \end{bmatrix}$, the constants (should tend to zero)

$V = \begin{bmatrix} V_{X_P} \\ \\ V_{Z_R} \end{bmatrix}$, the residual values

Using the least squares solution

$${}_mA_{nm}X_1 = {}_mL_1 + {}_mV_1 \quad \text{(C23)}$$

Which can also be expressed as

$$A^TAX = A^TL \quad \text{(C24)}$$

Which is the matrix of the "normal" equation coefficients of the unknowns.

Multiplying by $\left(A^TA\right)^{-1}$

$$\left(A^TA\right)^{-1}A^TAX = \left(A^TA\right)^{-1}A^TL$$

$$IX = \left(A^TA\right)^{-1}A^TL \quad \text{(C25)}$$

$$X = \left(A^TA\right)^{-1}A^TL$$

(equally weighted)

Standard Deviation

$$S_0 = \sqrt{\frac{V^T V}{r}} \qquad \text{(C26)}$$

r = degrees of freedom
= # equations - # unknowns
= 9-7
= 2

check:

$$L: \begin{bmatrix} X_P - X_{P_0} \\ \\ Z_R - Z_{R_0} \end{bmatrix}$$

These terms can be tested to satisfy the required precision.

SIMPLIFICATION FOR 2D RECTIFICATION

In motor vehicle collision analysis, the events most often occur on a two-dimensional plane, the road surface. 2D rectification requires only one photograph and deals with the projection of the photograph plane (image plane) onto the road surface plane (object plane). These conditions lead to a number of simplifying assumptions that reduce the mathematical complexity.

As stated earlier the fundamental assumptions of 2D rectification are that the focal point of the camera, the point on the negative (or positive), and the corresponding surface point are collinear, and that the surface is planar. The geometry of 2D rectification is illustrated in Figures C7 and C8.

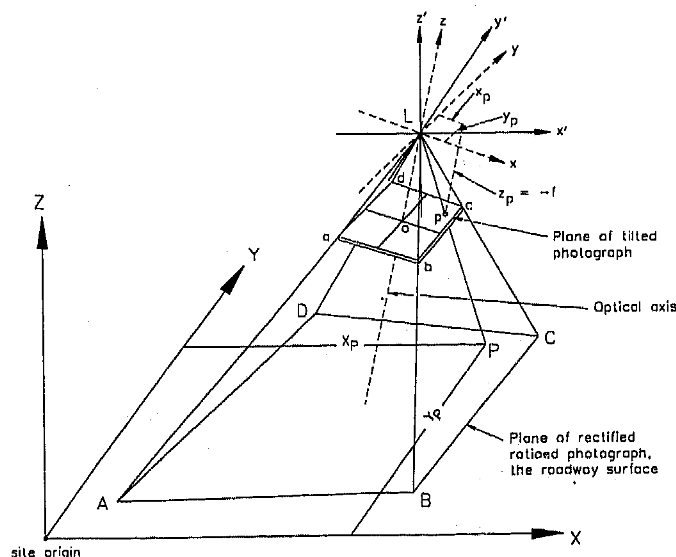


Figure C7 - The image space and the object space

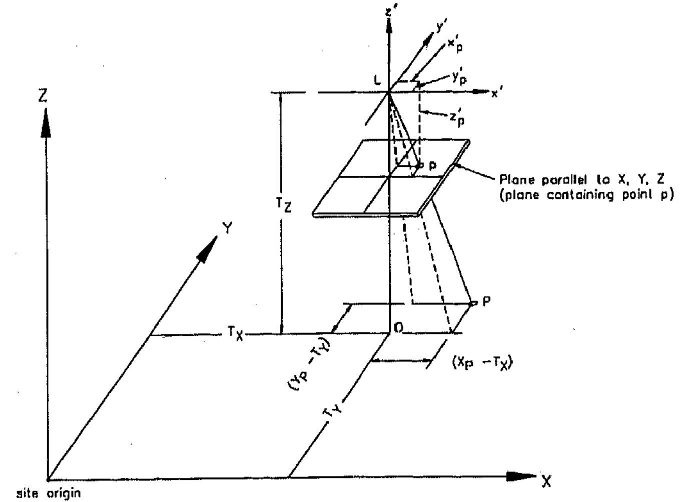


Figure C8 - The image space rotated parallel to the object space

Starting from the equations for rotation of the image plane:

$$\begin{aligned} x_p &= m_{11}x'_p + m_{12}y'_p + m_{13}z'_p \\ y_p &= m_{21}x'_p + m_{22}y'_p + m_{23}z'_p \\ z_p &= m_{31}x'_p + m_{32}y'_p + m_{33}z'_p \end{aligned} \qquad \text{(C27)}$$

From similar triangles in Figure 8:

$$x'_p = \frac{z'_p(X_P - T_X)}{T_Z}, \qquad \text{(C28)}$$

$$y'_p = \frac{z'_p(Y_P - T_Y)}{T_Z}, \text{ and} \qquad \text{(C29)}$$

$$z'_p = \frac{z'_p(T_Z)}{T_Z} \qquad \text{(C30)}$$

Substituting and rearranging:

$$\begin{aligned} x_p &= m_{11}z'_p\frac{(X_P - T_X)}{T_Z} + m_{12}z'_p\frac{(Y_P - T_Y)}{T_Z} + m_{13}z'_p\frac{(T_Z)}{T_Z} \\ y_p &= m_{21}z'_p\frac{(X_P - T_X)}{T_Z} + m_{22}z'_p\frac{(Y_P - T_Y)}{T_Z} + m_{23}z'_p\frac{(T_Z)}{T_Z} \\ z_p &= m_{31}z'_p\frac{(X_P - T_X)}{T_Z} + m_{32}z'_p\frac{(Y_P - T_Y)}{T_Z} + m_{33}z'_p\frac{(T_Z)}{T_Z} \end{aligned}$$

(C31)

Expanding and collecting terms:

$$x_p = \frac{z'_p}{T_Z}[m_{11}(X_P - T_X) + m_{12}(Y_P - T_Y) + m_{13}T_Z]$$

$$y_p = \frac{z'_p}{T_Z}[m_{21}(X_P - T_X) + m_{22}(Y_P - T_Y) + m_{23}T_Z]$$

$$z_p = \frac{z'_p}{T_Z}[m_{31}(X_P - T_X) + m_{32}(Y_P - T_Y) + m_{33}T_Z] \qquad \text{(C32)}$$

Next, dividing the equations for x and y by the equation for z:

$$x_p = z_p \frac{[m_{11}(X_P - T_X) + m_{12}(Y_P - T_Y) + m_{13}T_Z]}{[m_{31}(X_P - T_X) + m_{32}(Y_P - T_Y) + m_{33}T_Z]}$$

$$y_p = z_p \frac{[m_{21}(X_P - T_X) + m_{22}(Y_P - T_Y) + m_{23}T_Z]}{[m_{31}(X_P - T_X) + m_{32}(Y_P - T_Y) + m_{33}T_Z]} \qquad \text{(C33)}$$

Next, by dividing each term in the numerator and denominator by $m_{33}T_z$:

$$x_p = \frac{\frac{z_p m_{11}}{m_{33}T_Z}(X_P - T_X) + \frac{z_p m_{12}}{m_{33}T_Z}(Y_P - T_Y) + \frac{z_p m_{13}}{m_{33}T_Z}T_Z}{\frac{m_{31}}{m_{33}T_Z}(X_P - T_X) + \frac{m_{32}}{m_{33}T_Z}(Y_P - T_Y) + \frac{m_{33}T_Z}{m_{33}T_Z}}$$

$$y_p = \frac{\frac{z_p m_{21}}{m_{33}T_Z}(X_P - T_X) + \frac{z_p m_{22}}{m_{33}T_Z}(Y_P - T_Y) + \frac{z_p m_{23}}{m_{33}T_Z}T_Z}{\frac{m_{31}}{m_{33}T_Z}(X_P - T_X) + \frac{m_{32}}{m_{33}T_Z}(Y_P - T_Y) + \frac{m_{33}T_Z}{m_{33}T_Z}} \qquad \text{(C34)}$$

Since the camera height (T_Z), focal length ($-z_p$), rotation angles (φ, ω, κ), and translations (T_X, T_Y) are all constant:

$$x_p = \frac{c_1 X_P + c_2 Y_P + c_3}{c_4 X_P + c_5 Y_P + 1}$$

$$y_p = \frac{c_6 X_P + c_7 Y_P + c_8}{c_4 X_P + c_5 Y_P + 1} \qquad \text{(C35)}$$

where

$$c_1 = \frac{z_p m_{11}}{m_{33}T_Z} \qquad c_2 = \frac{z_p m_{12}}{m_{33}T_Z}$$

$$c_3 = \frac{z_p m_{13}}{m_{33}} - \frac{z_p m_{11}T_X}{m_{33}T_Z} - \frac{z_p m_{12}T_Y}{m_{33}T_Z}$$

$$c_4 = \frac{m_{31}}{m_{33}T_Z} \qquad c_5 = \frac{m_{32}}{m_{33}T_Z} \qquad \text{(C36)}$$

$$c_6 = \frac{z_p m_{21}}{m_{33}T_Z} \qquad c_7 = \frac{z_p m_{22}}{m_{33}T_Z}$$

$$c_8 = \frac{z_p m_{23}T_Z}{m_{33}T_Z} - \frac{z_p m_{21}T_X}{m_{33}T_Z} - \frac{z_p m_{22}T_Y}{m_{33}T_Z}$$

The above equations are general and represent the projectivity between two planes. Thus, the reverse transformation, from photo points to road surface points, can also be expressed as:

$$X_P = \frac{c'_1 x_p + c'_2 y_p + c'_3}{c'_4 x_p + c'_5 y_p + 1}$$

$$Y_P = \frac{c'_6 x_p + c'_7 y_p + c'_8}{c'_4 x_p + c'_5 y_p + 1} \qquad \text{(C37)}$$

where

$$c'_1 = \frac{T_Z m_{11}}{m_{33}z_p} \qquad c'_2 = \frac{T_Z m_{21}}{m_{33}z_p}$$

$$c'_3 = \frac{T_Z m_{31}z_p + T_X}{m_{33}z_p}$$

$$c'_4 = \frac{m_{13}}{m_{33}z_p} \qquad c'_5 = \frac{m_{23}}{m_{33}z_p} \qquad \text{(C38)}$$

$$c'_6 = \frac{T_Z m_{12}}{m_{33}z_p} \qquad c'_7 = \frac{T_Z m_{22}}{m_{33}z_p}$$

$$c'_8 = \frac{T_Z m_{32}z_p + T_Y}{m_{33}z_p}$$

For 2D rectification there are eight constants and two equations per control point, therefore at least four control points are required for a solution. The projectivity equations can be written in matrix form and the eight constants can be solved directly. If more than four control points are used a least squares solution of the eight constants can be performed. The eight constants themselves are made up of seven independent unknowns.

- Three translations (T_X, T_Y, T_Z)
- Three rotations (ϕ, ω, κ)
- Camera focal length $(-z_p)$

PC-Rect makes some assumptions about the orientation of the camera which simplifies the solution

process. The optical axis is defined to be in the Y-Z plane, and the focal point is defined to be at a height T_z above the object plane origin such that:

T_X = 0

T_Y = 0

T_Z = Camera height

κ = Rotation about Z-axis = 0

Thus, in PC-Rect there are only four independent unknowns:

- Camera height (T_Z)
- Two rotations (ϕ, ω), defined as the camera inclination and rotation respectively
- Camera focal length ($-z_p$)

In PC-Rect the user inputs known distances between points on the road surface rather than the xy coordinates of points. Each known distance provides one equation.

$$L_i = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \qquad \text{(C39)}$$

where

L_i = known distance on object space

$X_2, Y_2, ..$ = end point coordinates of known distance on object space

Substituting for ($X_2, Y_2, ..$) with Equation C37, results in one equation for each known distance. Since there are four unknowns, PC-Rect requires four known distances for a solution. The equations are solved by varying the unknowns in an iterative process until the weighted total error is minimized. The weighted total error provides the user with an indication of the precision expected from the results.

The individual and weighted total errors are calculated in PC-Rect 2.4 as follows:

$$\text{Individual Weighted Error} = w_i \delta L_i \qquad \text{(C40)}$$

$$\text{Weighted Total Error} = \frac{\sqrt{\sum_i^n (w_i \delta L_i)^2}}{\sum_i^n w_i} \qquad \text{(C41)}$$

where

w_i = weighting of known distance, *i*

δL_i = error in known distance *i*

APPENDIX D

Images 15 to 18 were rectified at different resolutions from a 600 dpi bitmap of a 4" x 6" photograph print of Image 2a (original file size 8336 kB). The resolution and file size of each image are as specified below.

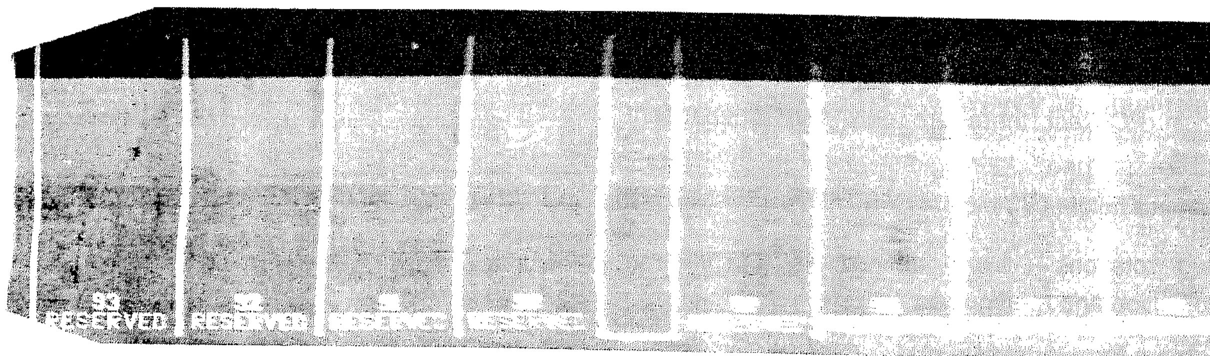


Image 15 - 600 dpi Original - 100 pixel/meter Rectification - Rectified file size 1442 kB

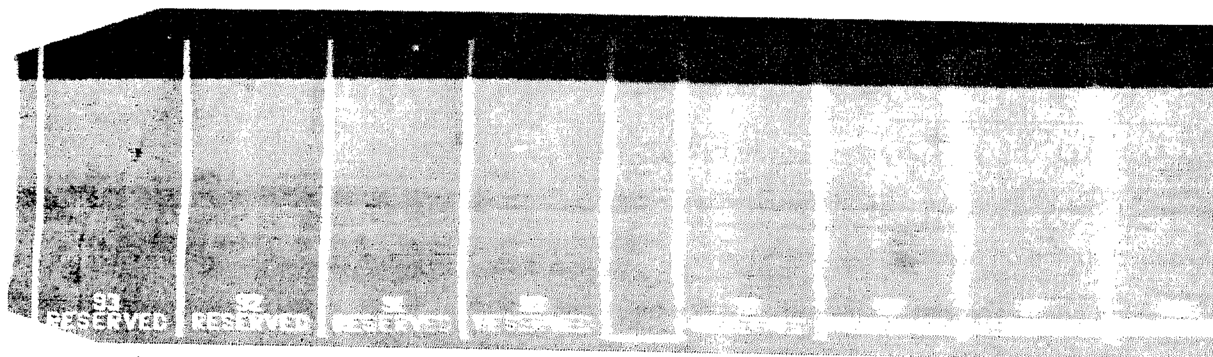


Image 16 - 600 dpi Original - 50 pixel/meter Rectification - Rectified file size 361 kB

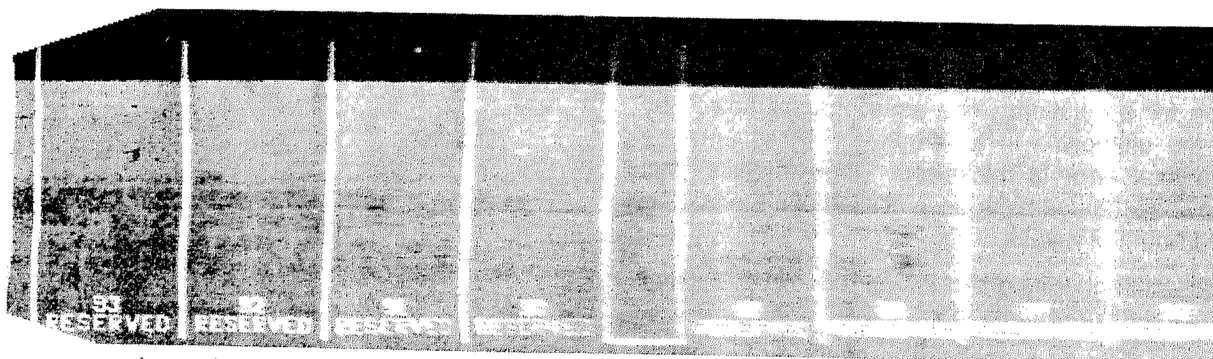


Image 17 - 600 dpi Original - 30 pixel/meter Rectification - Rectified file size 130 kB

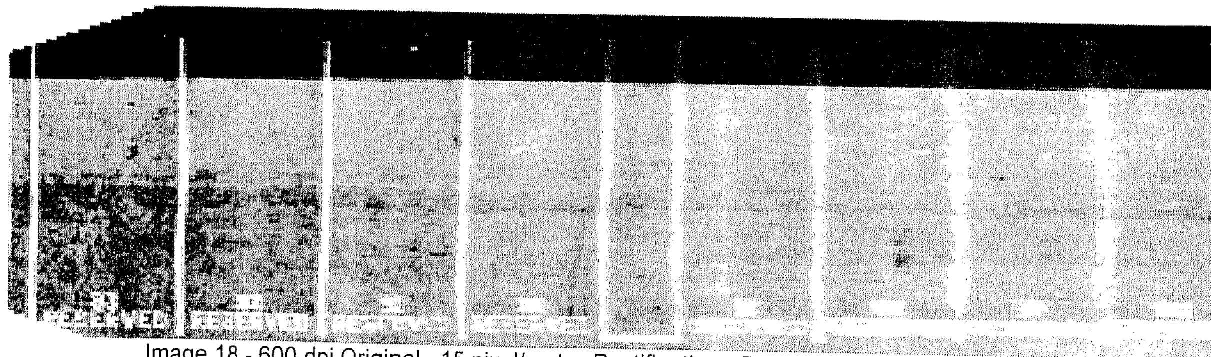


Image 18 - 600 dpi Original - 15 pixel/meter Rectification - Rectified file size 32 kB

Rectified images 19 to 22 were also based on a 4" x 6" photograph print of Image 2a. For these, however, the scanning resolution of the photograph print was varied and the rectification resolution was kept at 30 pixels/meter. This resulted in a file size of 130 kB for each rectified image. The original bitmap file size is given below each image.

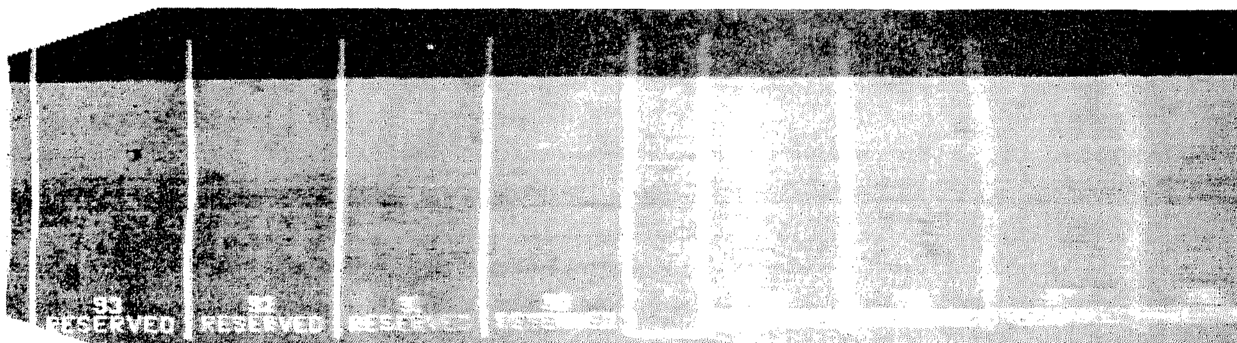


Image 19 - 600 dpi Original - 30 pixel/meter Rectification - Original file size 8336 kB

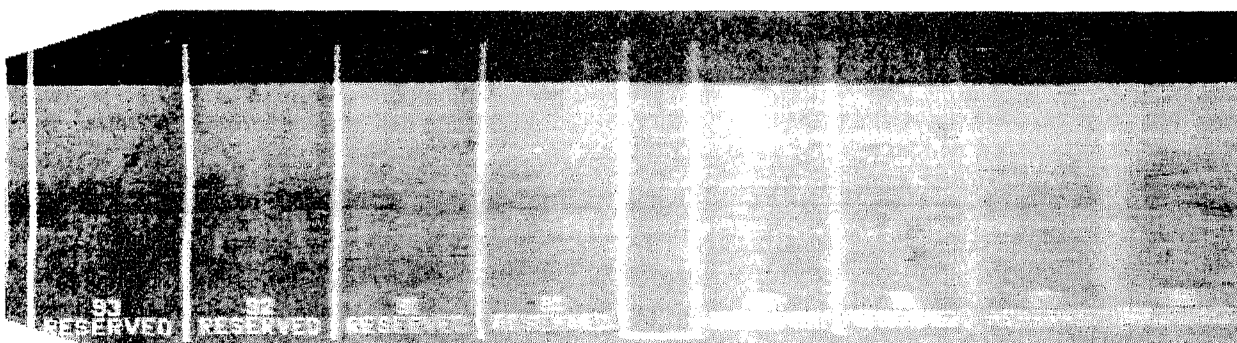


Image 20 - 300 dpi Original - 30 pixel/meter Rectification - Original file size 2088 kB



Image 21 - 150 dpi Original - 30 pixel/meter Rectification - Original file size 523 kB

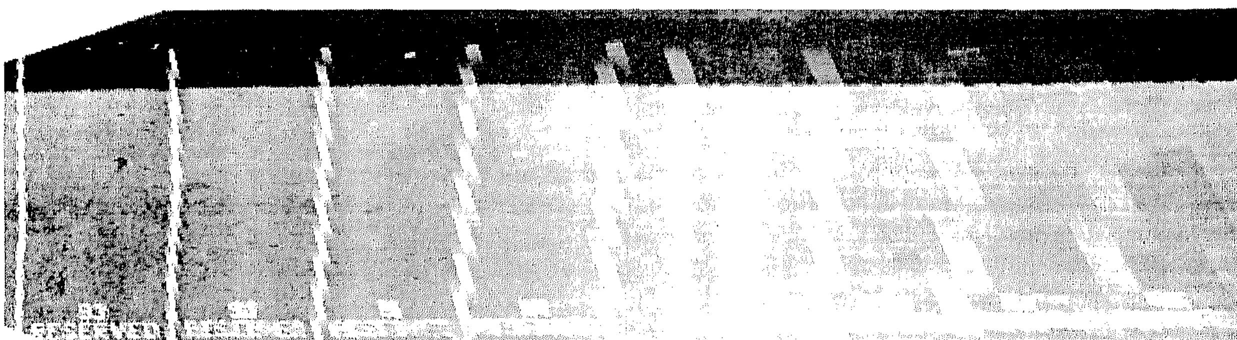


Image 22 - 75 dpi Original - 30 pixel/meter Rectification - Original file size 132 kB

970954

Drag and Steering Effects of Under Inflated and Deflated Tires

Ric Robinette, Darrell Deering, and Richard J. Fay
Fay Engineering Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Under inflated or deflated tires are known to cause increased forward drag and lateral steering effects on vehicles. These effects are commonly suggested to be the cause of driver loss of control and subsequent vehicular accidents. The increased drag and steering effects of under inflated or deflated tires are frequently an issue in an accident reconstruction. This paper documents the results of a series of tests conducted to determine the magnitude and effects of under inflated or deflated tires on cars and light trucks. The test also establishes a method of testing that can be used to determine steering effects for other vehicles and speed conditions. Six vehicles ranging from a compact passenger car to a 3/4 ton pickup truck were tested. The test methodology was simple and produced repeatable test results up to the 45 mph speed defined as a limit for the tests.

The first phase of the tests involved the measurement of forward deceleration rates during a series of coast to stop tests with all of the vehicle tires initially inflated to the rated pressure. This was followed by tests at various under inflation pressures. The deceleration rates versus inflation pressures are documented and discussed. As anticipated all vehicles experienced an increase in the coast to stop deceleration rates at under inflated tire pressure conditions. However, the increase in the deceleration rates were not dramatic in magnitude. When all four tires on the test vehicles were deflated to 10 psi, the coast to stop deceleration rates still represented relatively small deceleration profiles.

The second phase of the tests was performed with a totally deflated tire on either the front or rear axle. The forward deceleration and lateral acceleration rates, and the corresponding dimensional lateral motion profiles were measured during the coast to stop tests with the deflated tires. The forward and lateral acceleration rates and the lateral offset travel distances are documented and discussed. During the test, the test driver was always able to easily maintain control of the test vehicles and steer them in either a straight path or curved path with either a deflated front or rear tire. The driver was always able to easily regain control and steer the test vehicles after initially releasing the steering wheel and allowing the vehicle to initiate lateral motion as a result of the deflated tires. This suggested that a blow-out single tire event can be a controllable condition.

These tests provided consistent results within the test speed ranges up to 45 mph. The results may not be similar at higher speeds and the results should not be randomly extrapolated to higher vehicle speeds without a valid basis. If higher speed affects are desired, the same test methology is considered to be a reasonable approach to gather data.

INTRODUCTION

The effects of under inflated or deflated tires are known to increase rolling resistance and to cause a steering effect wherein the vehicle tends to steer to the side of the vehicle with the deflated tire. This lateral steering effect after a tire blow-out has been alleged to be the cause of many accidents. A review of the literature provided limited objective data defining these lateral steer effects. Limpert (4) summarizes some lateral travel versus time data after a blow-out condition on the steering axle of a semi tractor-trailer combination. A series of tests were planned to produce subjective and objective results quantifying the forward and lateral acceleration rates and lateral motion with under inflated tires and deflated tires on either a front or rear axle. This paper documents the results of the first phases of the proposed test program. The tests were designed to be simple to perform with the minimum of test equipment so that the test could be easily repeated at a future date for any vehicle type that may be associated with a specific accident investigation or reconstruction issue. The tests were designed to permit the accumulation of a data base that can be utilized in conjunction with vehicle dynamic steering response simulator programs such as the Single Vehicle Simulator (SVS), HVOSM (5) or the Engineering Dynamics Corporation's EDSVS (6).

A range of vehicles was selected for the first phase of the testing program to include various sizes and types of passenger cars and light trucks. The vehicles were selected to include

TABLE 1

VEHICLE AND TIRE SPECIFICATION DATA

TEST NO.	1	2	3	4	5	6
MFG.	MAZDA	NISSAN	NISSAN	OLDS	FORD	JEEP
MODEL	RX-7	SENTRA	ALTIMA	CIERA	F250 4WD	CJ-5 4WD
YEAR	1979	1996	1996	1996	1984	1974
TEST WEIGHT	2710	2590	3130	3410	4680	2890
% FRONT AXLE	51	63	62	63	61	50
% REAR AXLE	49	37	38	37	39	50
TRANSMISSION	4 MAN	AUTO	AUTO	AUTO	4 MAN	4 MAN
WHEELBASE	95	100	103	105	133	84
TRACK	55	57	58	58	66	51
WB/TRACK	1.73	1.75	1.78	1.81	2.02	1.65
TIRE MFG.	PIRELLI	GENERAL	GENERAL	UNIROYAL	FIRESTONE	GOODRICH
TIRE MODEL	T-300	AMERI*	XP2000	TIGER PAW	STEELTEX	T/A RADIAL
TIRE SIZE	185/70-13	175/70-13	205/60-15	195/75-14	245/75-16	235/75-15
RATED PRESS.	32 PSI	32/29 PSI	32 PSI	30 PSI	50 PSI	51 PSI

front and rear wheel drive configurations. The test vehicles, tires and selected specification characteristics are listed in Table 1. All the vehicles were standard production vehicles. All of the tires used in the test were standard production tires with tread depths in excess of 6/32 inch.

TEST METHODOLOGY

The test were conducted on a dry asphalt air strip runway at the decommissioned Stapleton Airport in Denver, Colorado. The runway extended approximately 5000 feet in the north and south direction. The runway was 100 feet wide. The runway surface elevation profile was surveyed and was considered relatively level. To minimize any effects of grade and wind, all vehicles were subjected to repeated tests in the north and south directions and the results were arithmetically averaged. All of the test data presented in this paper is the arithmetic average of the repeated tests for each individual vehicle.

STRAIGHT LINE COAST TO STOP TESTS – The initial phase of the testing involved coast to stop tests that was designed to identify vehicle/tire deceleration characteristics at rated pressure and at lower tire pressure values. All of the coast to stop tests were conducted with the vehicles entering the deceleration test zone at a constant speed of 30 mph. Automatic transmission vehicles were driven in "D" drive gear and the four wheel drive vehicles were in two (rear) wheel drive mode. The initial test was conducted with all tires inflated at the rated tire pressure. Subsequent tests were then performed with all tire pressures reduced to 25 psi and 10 psi. The vehicles were driven at a constant 30 mph speed into the deceleration test zone and shifted into neutral and allowed to coast to a stop. The vehicle was then subjected to another similar coast down test while traveling in the opposite direction. It was decided to place all vehicles in neutral during the coast down test to minimize the potential differential effects of vehicle drive train braking and inertia that may be encountered with the different combination of manual and automatic transmission configurations that were being tested. Although there is still some drag on the vehicles from bearings and viscous friction, positioning the vehicles in neutral should minimize the mechanical friction and maximize the relative effects of variable tire inflation pressures. During the coast down tests in neutral, the driver continued to hold the steering wheel and guide the vehicle on a straight path.

During the coast to stop testing, the travel distance and time were recorded and utilized to independently calculate overall average deceleration rates. The deceleration rates from the coast to stop tests in opposite directions were compared for anomalies and individual tests were repeated if an anomaly was observed. The set of deceleration data were arithmetically averaged. Only the average values are presented in the paper and are used for comparison purposes.

DECELERATION AND LATERAL STEER TESTS WITH DEFLATED TIRES – The second phase of the testing involved coast to stop tests and lateral motion tests with totally deflated tires in order to simulate a tire blow-out condition. These tests were designed to define the forward and lateral deceleration rates and lateral travel distances with either a front axle or a rear axle deflated tire. Most tests were conducted with the vehicles entering the deceleration test zone at a constant speed of 30 mph with either a totally deflated front or rear tire. The same transmission and two wheel drive settings as the coast to stop tests were used. The test driver entered the test zone at a constant 30 mph with the required steering input to maintain travel along a straight line parallel to the edge of the asphalt runway. At the beginning of the deceleration test zone, the driver removed his hands from the steering wheel, removed his foot from the accelerator pedal, and simultaneously shifted into neutral (or depressed the clutch then shifted into neutral). The vehicle was thus allowed to coast and to freely respond to the deflated front or rear axle tire. It was decided to perform the test with the driver releasing the steering wheel to simulate a

scenario of an on road blow-out condition followed by no driver response or attempt to control the vehicle. The release of the steering wheel and allowing the vehicle to freely respond should produce the maximum lateral acceleration and lateral motions. Any reasonable attempts to control the vehicle by the driver would result in lesser lateral acceleration and lateral motion. The test driver was always able to drive all vehicles in a straight line and to steer the vehicles in either direction with either a deflated front or rear tire.

During the deflated tire tests the forward and lateral deceleration rates were recorded via a Valentine Research G-Analyst bi-axial accelerometer test instrument. The trajectory path of the vehicle was documented by means of a spray paint can mounted along passenger door area and operated by front seat passenger during the coast down. A less reliable method of defining the trajectory path was the deflated tire mark on the runway surface. This mark was not always clearly visible for all the vehicles. The forward and lateral motions were measured at predefined distances along the paint trail. The paint trail data was also useful in defining the changing radius of curvature of the trajectory path.

The numerical test results were obtained by applying the previously defined data averaging process for multiple tests in opposite directions. The intent for all the vehicle tests was to allow the vehicle to coast to a stop. However, in several tests the vehicle lateral travel exceeded the 100 feet width of the runway before the vehicle came to a stop. In those instances, the driver braked the vehicle to a stop at the opposite edge of the runway.

In order to have meaningful subjective test results, the same vehicle test driver was used for all of the deflated tire tests.

COAST TO STOP TESTS WITH RATED AND UNDER INFLATED TIRE PRESSURES

The results of the rolling resistance deceleration tests with rated and under inflated tire pressures are shown in Table 2 and graphically illustrated in Figure 2. At rated tire pressure conditions, the Mazda RX-7 exhibited the lowest overall deceleration and the Jeep had the highest deceleration rate. The aerodynamic drag resistance of the test vehicles may contribute somewhat to this trend as the Mazda is expected to have the lowest aerodynamic drag and the Jeep probably has the highest drag. In one test the Mazda RX-7 coasted over 3700 feet. At these deceleration levels, a small grade effect in the test surface can have a noticeable effect on the coast to stop distances.

TABLE 2

VEHICLE DECELERATION RATES VERSUS TIRE INFLATION PRESSURES

TEST CONDUCTED BY APPROACHING TEST AREA AT CONSTANT 30 MPH THEN SHIFTING THE VEHICLE INTO NEUTRAL AND COASTING TO A STOP. THE DATA VALUES SHOWN ARE AVERAGE RESULTS FROM MULTIPLE TEST RUNS IN OPPOSITE DIRECTIONS.

TEST NO.	1	2	3	4	5	6
MFG.	MAZDA	NISSAN	NISSAN	OLDS	FORD	JEEP
MODEL	RX-7	SENTRA	ALTIMA	CIERA	F250 4WD	CJ-5 4WD
YEAR	1979	1996	1996	1996	1984	1974
RATED PRESSURE	32 PSI	32/29 PSI	32 PSI	30 PSI	50 PSI	51 PSI

AVERAGE VEHICLE DECELERATION DURING COAST TO STOP, g'S

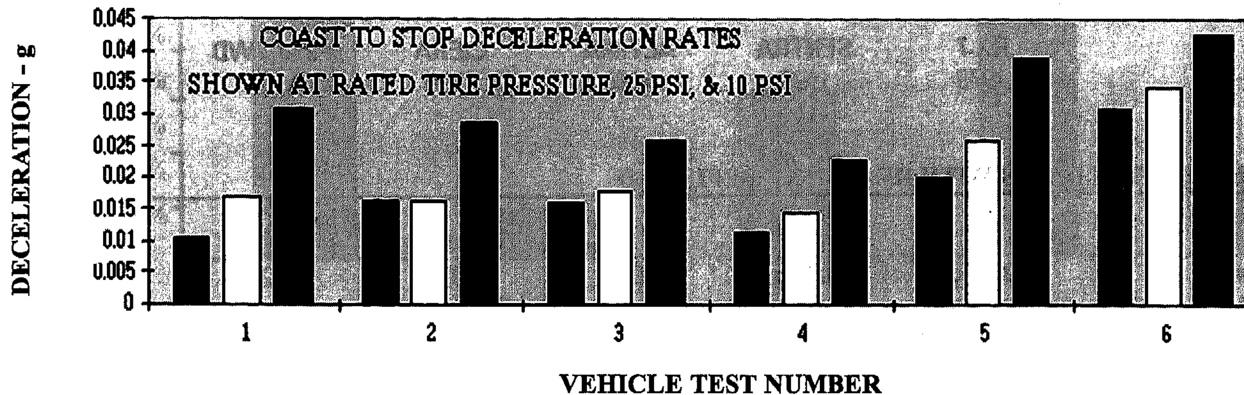
TEST NO.	1	2	3	4	5	6
RATED PRESSURE	0.0106	0.0161	0.0158	0.0113	0.0201	0.0309
25 PSI TIRE PRESSURE	0.0168	0.0163	0.0181	0.0143	0.0258	0.0343
10 PSI TIRE PRESSURE	0.0308	0.0288	0.0258	0.0227	0.0389	0.0427

AVERAGE VEHICLE DECELERATION AS % CHANGE OF DECELERATION AT RATED PRESSURE

TEST NO.	1	2	3	4	5	6
RATED PRESSURE	100%	100%	100%	100%	100%	100%
25 PSI TIRE PRESSURE	158%	101%	115%	127%	128%	111%
10 PSI TIRE PRESSURE	291%	179%	163%	201%	194%	138%

FIGURE 2

DECELERATION RATES VS TIRE PRESSURE



Overall average deceleration rate calculations based upon the coast to stop distances were consistently about 5% higher then the overall average deceleration rates calculated from the elapsed times associated with the coast to stop tests. This slight difference was very repeatable and consistent throughout the tests.

Each individual vehicle exhibited an expected trend of increased deceleration rates with reducing tire inflation pressure. However, as a whole, the test group did not exhibit large increases in the deceleration rates. With the exception of the Mazda RX-7, the rates did not increase more than 200% between rated pressure and 10 psi. The absolute deceleration rates recorded for the Mazda at reduced tire pressures are similar to the other test vehicles. However, the percentage increases for the Mazda are larger because the baseline value at the rated tire pressure was lower.

The Ford F-250 pickup and the Jeep were equipped with tires with a higher load capacity rating relative to the actual tire loads associated with the vehicle weights. These higher load capacity tires probably have a stiffer side wall design then the tires on the test cars. The Ford pickup and Jeep tires exhibited less tire deflection than the car tires at the under inflated conditions. Inspite of this reduced side wall deflection characteristic, lower tire inflation pressures resulted in similar deceleration rates for the cars, the pickup truck and Jeep.

A subjective observation that was supported by the numerical test results was that the reduced tire pressure conditions were not always very perceptible to the driver in terms of vehicle handling. In contrast, the reduced pressure conditions were observable in the tire side wall contour at the lower pressures. Thus, a case could be made that handling and driveability characteristics may not alert a driver to an under inflated tire, and hence vehicle owners should regularly monitor their tire pressures to insure the tires are being maintained at the proper tire pressure.

SINGLE FRONT AND REAR DEFLATED TIRE TEST

Each vehicle was subjected to a series of tests with a deflated rear axle tire, and then the tests were repeated with a deflated front axle tire. Forward and lateral deceleration data was continuously recorded during these tests with a Valentine Research G-Analyst. Other dimensional data related to forward travel, lateral offset motions from the original straight line of travel, and radius of travel were recorded at 50 feet and 100 feet increments from the point of the initiation of the test.

The Jeep responses to the initial deflated tire tests were in conflict with the anticipated trends. At times Old Yeller (as the Jeep has affectionately become known) appeared to have a mind of its own and proceed in a repeatable but unanticipated direction. Subsequent testing determined that an improper steering alignment was resulting in a continuous lateral steer in one direction. Since the deflated tire lateral motion tests were skewed by the improper steering alignment, this test data was declared invalid and is not presented. However, the responses of the Jeep did provide some valuable subjective information that was consistent with other observation on the other vehicles. These will be discussed in a later section of the paper.

The forward deceleration rates with a totally deflated rear axle tire and with a totally deflated front axle tire are presented in Table 3 for travel distances of 50 feet and 100 feet after the initiation of the test. Section C in Table 3 illustrates the absolute and percentage increase of the forward deceleration rates relative to the tests with all tires at rated pressure. Figures 3-A and 3-B present these percentage increases graphically.

The following general observations were made. In general, all the vehicles exhibited the anticipated increase in forward deceleration with a deflated tire relative to the test with all tires at the rated tire pressure. Surprisingly, the Ford pickup did not exhibit any appreciable reduction with a deflated rear

TABLE 3

DEFLATED TIRE TEST

ALL DEFLATED TIRE TESTS WERE CONDUCTED BY ENTERING THE DECELERATION ZONE AT A CONSTANT 30 MPH. AT THE START OF THE TEST THE DRIVER REMOVED HIS HANDS FROM THE STEERING WHEEL AND

MODEL	RX-7	SENTRA	ALTIMA	CIERA	F250 4WD	CJ-5 4WD
YEAR	1979	1996	1996	1996	1984	1974

FOWARD DECELERATION RATES WITH REAR TIRE DEFLATED

TEST NO.	1	2	3	4	5	6
A. At forward travel distance of 50 ft	0.04 g	0.03 g	0.03 g	0.03 g	0.02 g	n/a
B. At forward travel distance of 100 ft	0.04 g	0.03 g	0.03 g	0.03 g	0.02 g	n/a
C. Increase in forward deceleration rate for deflated tire versus all four tires at rated pressure	0.029 g 277%	0.014 g 86%	0.014 g 90%	0.019 g 165%	0.00 g n/a	n/a n/a

FORWARD DECELERATION RATES WITH FRONT TIRE DEFLATED

TEST NO.	1	2	3	4	5	6
A. At forward travel distance of 50 ft	0.06 g	0.03 g	0.07 g	0.07 g	0.04 g	n/a
B. At forward travel distance of 100 ft	0.06 g	0.03 g	0.06 g	0.07 g	0.04 g	n/a
C. Increase in forward deceleration rate for deflated tire versus all four tires at rated pressure	0.049 g 466%	0.014 g 86%	0.049 g 311%	0.059 g 519%	0.02 g 99%	n/a n/a

FIGURE 3-A

INCREASE IN FORWARD DECELERATION WITH REAR TIRE DEFLATED VS ALL TIRES AT RATED PRESSURE

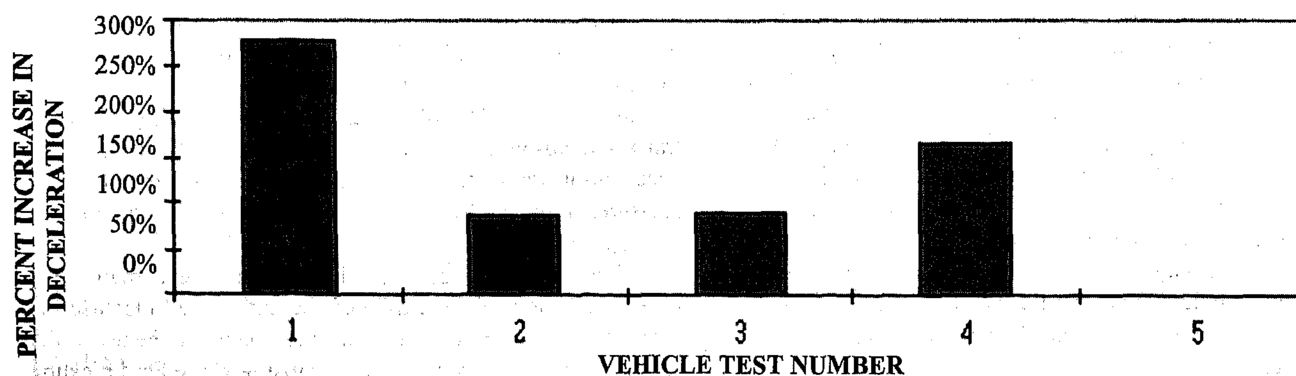


FIGURE 3-B

INCREASE IN FORWARD DECELERATION WITH FRONT TIRE DEFLATED VS ALL TIRES AT RATED PRESSURE

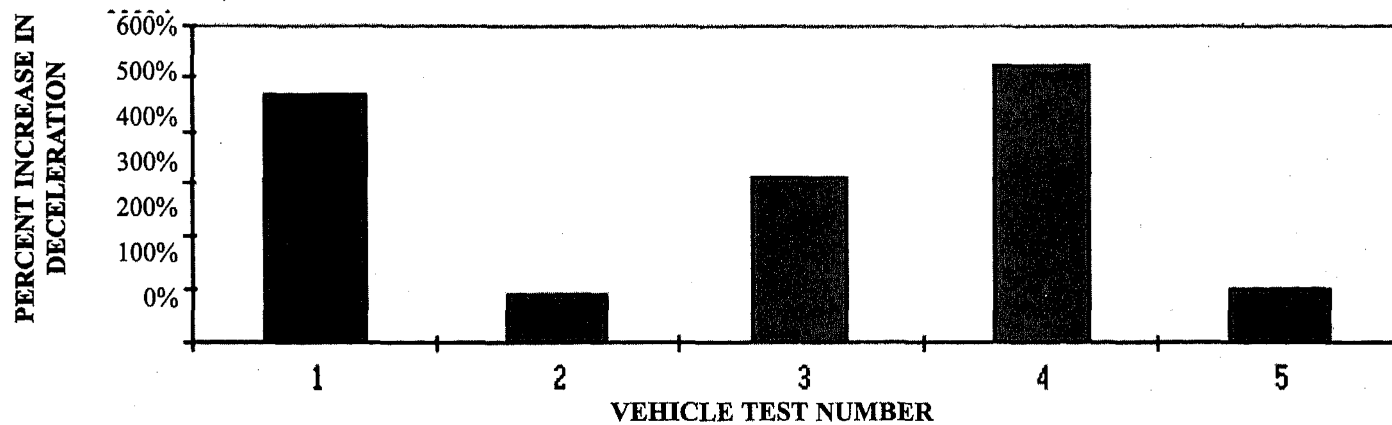


TABLE 4

DEFLATED TIRE TEST

ALL DEFLATED TIRE TESTS WERE CONDUCTED BY ENTERING THE DECELERATION ZONE AT A CONSTANT 30 MPH. AT THE START OF THE TEST THE DRIVER REMOVED HIS HANDS FROM THE STEERING WHEEL AND PLACED THE TRANSMISSION INTO NEUTRAL

TEST NO.	1	2	3	4	5	6
MFG.	MAZDA	NISSAN	NISSAN	OLDS	FORD	JEEP
MODEL	RX-7	SENTRA	ALTIMA	CIERA	F250 4WD	CJ-5 4WD
YEAR	1979	1996	1996	1996	1984	1974

LATERAL OFFSET DISTANCES IN FEET WITH REAR TIRE DEFLATED

TEST NO.	1	2	3	4	5	6
A. At forward travel distance of 50 ft	2.5	0.1	1.1	n/a	0.4	n/a
B. At forward travel distance of 100 ft	7	1.1	1.3	1.4	2.6	n/a
C. At forward travel distance of 200 ft	19.9	3.2	4.1	4	5.6	n/a

LATERAL OFFSET DISTANCES IN FEET WITH FRONT TIRE DEFLATED

TEST NO.	1	2	3	4	5	6
A. At forward travel distance of 50 ft	2.7	1.9	2.6	5	2.2	n/a
B. At forward travel distance of 100 ft	11.3	4.8	6.5	14.6	12	n/a
C. At forward travel distance of 200 ft	51.2	14.3	29.4	60.6	63.2	n/a

FIGURE 4-A

LATERAL OFFSET TRAVEL DISTANCES WITH A REAR TIRE DEFLATED

GRAPHS SHOWING CUMULATIVE OFFSET TRAVEL DISTANCES DURING TESTS:

OFFSET TRAVEL AFTER 50' LOWEST - PORTION OF BAR
CUMULATIVE OFFSET TRAVEL AFTER 100' - MIDDLE PORTION OF BAR
CUMULATIVE OFFSET TRAVEL AFTER 200' - TOP PORTION OF BAR

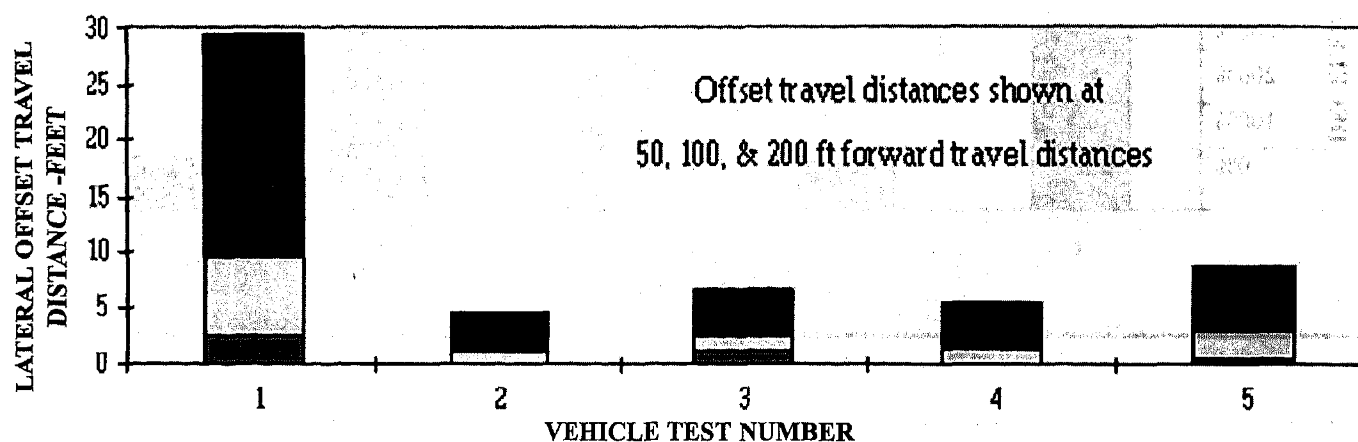
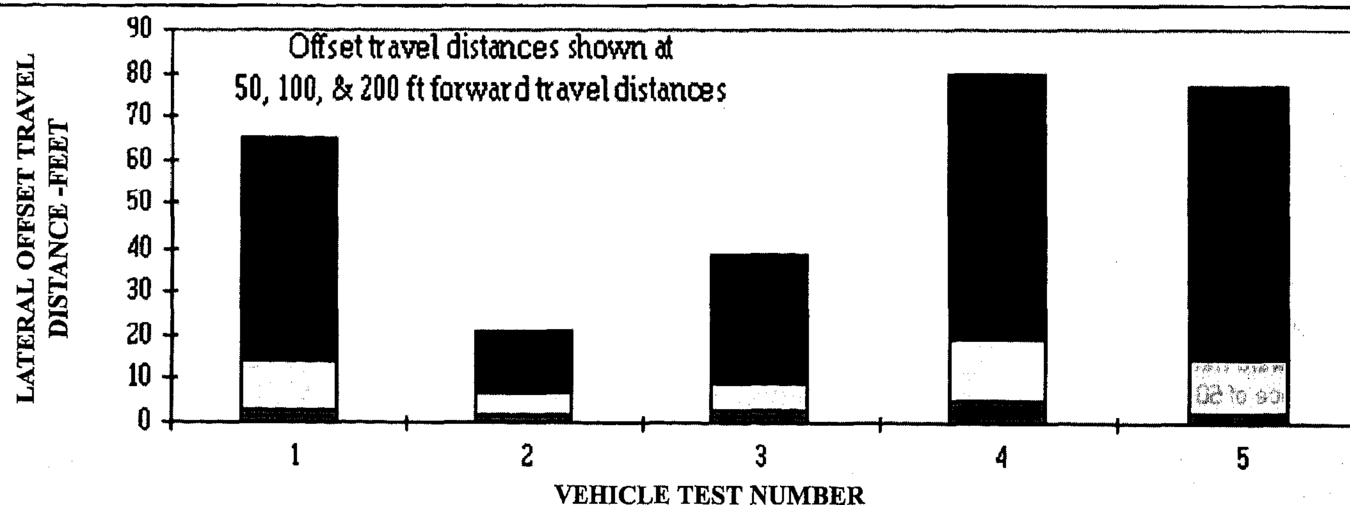


FIGURE 4-B

LATERAL OFFSET TRAVEL DISTANCES WITH A FRONT TIRE DEFLATED



axle tire. A front axle deflated tire did create an increase in the deceleration rate on the Ford. The Mazda RX-7 exhibited much higher percentage increases than the other test vehicles. Since the absolute deceleration values for the Mazda are similar to the other vehicles at the reduced tire pressure conditions, the differential percentage increase is once again a result of the relatively low Mazda deceleration rate with rated tire pressure. Another observation was that the vehicles that exhibited the smaller deceleration increases with a deflated rear tire also exhibited the smaller deceleration increases with a deflated front tire.

The lateral travel distances and lateral acceleration rates with deflated front and rear tires are shown in Tables 4 and 5. The lateral offset measurements relative to the initial straight line path are shown in Table 4 for forward travel distances of 50, 100, and 200 feet after the driver released the steering wheel and the vehicle was allowed to freely respond to the lateral accelerations associated with the deflated tires. These lateral travel distances would represent the maximum anticipated lateral travel distances a vehicle would experience with an on road blow-out if the driver provided no steering input or control at 30 mph. If the driver responded with any corrective steering input, the lateral travel would be reduced or potentially eliminated. For all tests performed, the test driver was always able to provide sufficient corrective steer to maintain the vehicle in a straight path or to steer the vehicle in either a right or left direction. These corrective steering maneuvers were accomplished with minor steering effort on the part of the test driver.

The data in Tables 3, 4 and 5 defining the deflated tire and lateral motion tests are presented in terms of forward travel distances, the data could have been developed in terms of time after start of the event. The following approximate times are provided for information purposes. The test vehicles

TABLE 5

DEFLATED TIRE TEST

ALL DEFLATED TIRE TESTS WERE CONDUCTED BY ENTERING THE DECELERATION ZONE AT A CONSTANT 30 MPH. AT THE START OF THE TEST THE DRIVER REMOVED HIS HANDS FROM THE STEERING WHEEL AND PLACED THE TRANSMISSION INTO NEUTRAL

LATERAL ACCELERATION RATES IN g's WITH REAR TIRE DEFLATED

TEST NO.	1	2	3	4	5	6
A. At forward travel distance of 50 ft						
1.) Measured test results	0.05	0.02	0.01	0.02	0.04	n/a
2.) Calculated from lateral offset motions	0.121	0.048	0.028	0.016	0.065	n/a
B. At forward travel distance of 100 ft						
1.) Measured test results	0.05	0.01	0.01	0.01	0.03	n/a
2.) Calculated from lateral offset motions	0.084	0.014	0.024	0.012	0.032	n/a

LATERAL ACCELERATION RATES IN g's WITH FRONT TIRE DELATED

TEST NO.	1	2	3	4	5	6
A. At forward travel distance of 50 ft						
1.) Measured test results	0.12	0.04	0.06	0.08	0.11	n/a
2.) Calculated from lateral offset motions	0.144	0.092	0.125	0.241	0.108	n/a
B. At forward travel distance of 100 ft						
1.) Measured test results	0.11	0.04	0.06	0.07	0.14	n/a
2.) Calculated from lateral offset motions	0.136	0.058	0.076	0.176	0.145	n/a

FIGURE 5-A

LATERAL ACCLERATION RATES WITH REAR TIRE DEFLATED

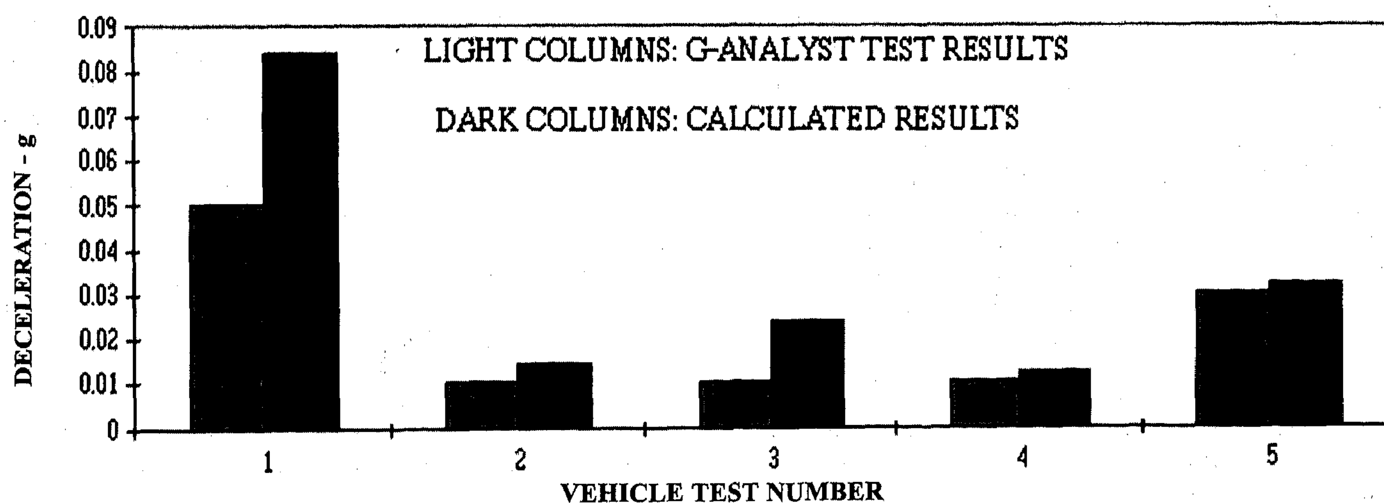
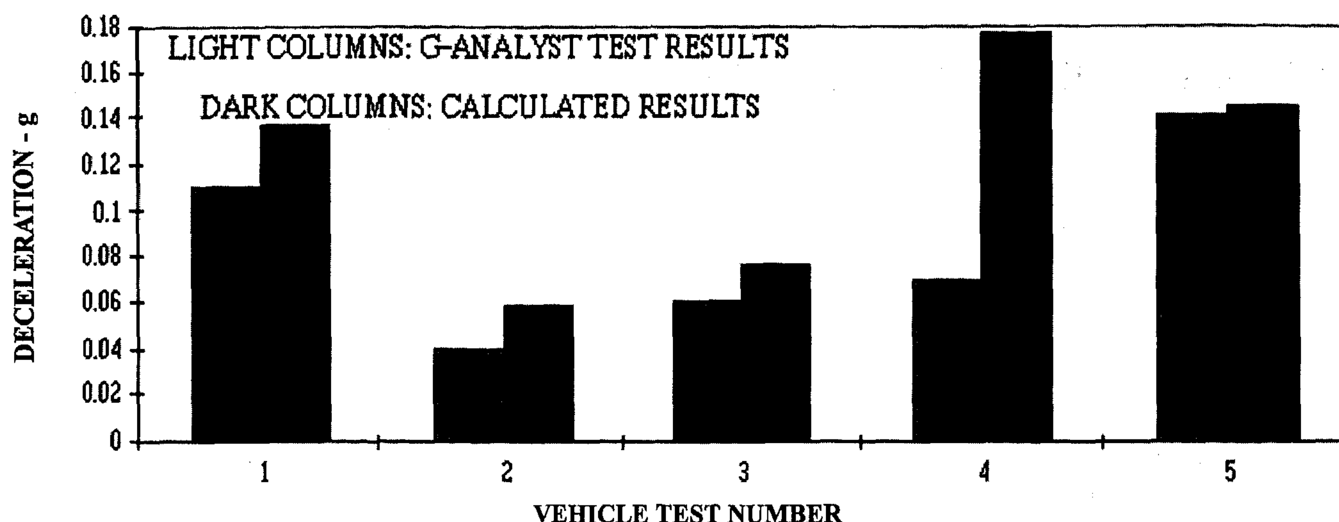


FIGURE 5-B

LATERAL ACCLERATION RATES WITH FRONT TIRE DEFLATED



passed the 50 feet station between 1.1 and 1.2 seconds after the start of the test, and passed the 100 feet station between 2.25 and 2.45 seconds.

The primary observed trend of the data on Table 4 was that the magnitude of lateral travel was much greater with a front deflated tire then with a rear deflated tire. This was expected. The lateral offset of the Mazda with a deflated rear tire was much greater than any of the other vehicles. This trend with the Mazda was repeatable with addition tests runs, however, the reason for this trend was not identified.

Other general trends and observations were that vehicles with front steer/drive axles had less lateral offset motion then vehicles without a front drive axle during the deflated rear tire tests. This trend was not observed during deflated front tire tests. Another general trend was that shorter wheelbase vehicles had less lateral offset then longer wheelbase vehicles during both front and rear deflated tire tests. The large lateral offset distances recorded with the Mazda was the obvious exception to this trend. This trend was not expected because it was anticipated that the longer wheelbase vehicles with higher moments of inertia would not be as susceptible to higher lateral accelerations and travel distances.

The lateral acceleration rates with deflated tires are presented in Table 5. Figure 5-A and 5-B graphically present the lateral acceleration data at the time that the vehicles are 100 feet beyond the start of the tests. Table 5 also includes the results of attempts to calculate a lateral acceleration rate based upon the measured lateral offset dimensions shown in Table 4. The calculated results were developed by applying the lateral offset equations while traveling on an assumed constant radius. The assumption of a constant radius was not expected, and the results demonstrated that the vehicles did not follow a constant radius. However, the assumption of a constant radius for short portions of the first portion of the travel path provided reasonable agreement between the calculated and measured lateral acceleration values.

Basic motion equations of centrifugal acceleration were used to calculate the lateral acceleration from the offset distance data. Several sources provide a derivation of this type of analysis. One such derivation is included in Daily (7). He offers the following relationship for a lateral offset:

$$D = 0.366\ S\ \sqrt{(M/F)}$$

Where D = forward travel distance along the initial direction of travel, feet

S = speed, mph

M = lateral offset distance from initial direction of travel, feet

F = lateral acceleration rate, g's

As expected, the trends in Table 5 follow the data trends in Table 4. The lateral acceleration rates follow the pattern of the lateral offset travel distances. There were higher lateral acceleration rates with a deflated front tire compared to a rear tire.

Figures 5-A and 5-B illustrate the relationship between the calculated lateral accelerations based upon the lateral offset dimensions and the lateral acceleration rates measured with the G-Analyst. The trend of the calculated values followed the trend of the measured values, but the calculated values were consistently higher than the measured values.

The localized radius of curvature of the lateral travel path were analyzed at 100 feet travel distances for the front and rear deflated tire tests. The radius of curvature was clearly changing throughout the trajectory path. However, the trend of the change was somewhat unexpected. With a deflated rear tire, all of the vehicles experienced a continuously increasing radius after the first 50 to 100 feet of travel. With a deflated front tire, the majority of the vehicle tests exhibited a continuously decreasing radius, with the other vehicle tests having an increasing radius for the first 100-200 feet of travel, followed by the decreasing radius pattern.

PHOTOGRAPHIC ILLUSTRATIONS

Photographs 1 through 7 illustrate the deflated tire mark and paint trail for one of the deflated left rear tire tests with the Mazda RX-7. The deflated tire mark and the paint trail are highlighted with arrows.

GENERAL OBSERVATIONS

A few general observations are offered from the combined test results. The Mazda

RX-7 exhibited the lowest rolling resistance drag during the rated tire pressure coast to stop tests. The Jeep exhibited the highest rolling resistance during the same test. During the deflated tire tests, the Mazda RX-7 generally exhibited much higher lateral travel distances and acceleration rates than the other vehicles. Frequently, the Mazda lateral motion responses conflicted with the observed trends in the other vehicles. One proposed explanation is that there was a greater differential drag between the relatively low rolling resistance of the Mazda inflated tires versus the deflated tire. This greater differential apparently resulted in more lateral acceleration and displacement. In contrast, the Jeep exhibited much lower lateral responses to a deflated tire than any other vehicle. The higher rolling resistance of the inflated tires resulted in a smaller differential drag between the inflated tires and the deflated tire. This would correspond to the much lower lateral accelerations and displacement that was observed during the tests with the Jeep.

The detailed test data presented in this paper are all based upon a common initial speed of 30 mph. The Nissan Altima and the Oldsmobile Ciera were subjected to limited testing at 45 mph with deflated tires. The following subjective observations are made relative to these limited test. During all of the vehicle tests, the driver felt the harshest vibrations in the steering wheel at very low speeds. As the speeds were increased, the vibration levels tended to drop to a reduced level and then remain fairly constant as speeds increased. The driver perceived steering input torque and steer wheel rotation to maintain straight travel at speeds of 10, 30 and 45 mph were similar. This was supported by the G-Analyst data that illustrated a fairly constant lateral and forward deceleration rate for each individual vehicle during the coast to stop tests with a deflated tire. When the Nissan Altima was subjected to a deflated rear tire test at 45 mph, the magnitude of the lateral travel profile was not notably different than the characteristics at 30 mph.

There were some driver perceived differences in the required steering wheel input torque to maintain a vehicle on a straight path for deflated front versus rear tire conditions, particularly on the Ford pickup and the Oldsmobile. The Oldsmobile exhibited much greater vehicle and steering wheel vibration with either a deflated front or rear deflated tire than any other vehicle. It was unclear if this was a function of the tire construction or of the vehicle. Regardless of the cause, the driver had no difficulty maintaining the Oldsmobile on a straight path or achieving controlled lateral steering in either direction with a deflated tire.

The documented lateral acceleration rates during the deflated tire tests were all below 0.15 g's when the vehicles were allowed to respond freely without any restraint at the steering wheel. These lateral acceleration levels were all within easily attainable lateral steer capabilities of roadway vehicles and could be readily controlled by a fairly minor response from the test driver.

The tests results provided consistent results within the test speed ranges up to 45 mph. The results may not be similar at higher speeds and the results should not randomly extrapolated to higher vehicle speeds without a valid basis. If higher speed affects are desired, the same test methology is considered to be a reasonable approach to gather data.

ACKNOWLEDGMENTS

The authors wish to express our gratitude to Mr. Philip Hight and Mr. Donald Rudney for their review and input to the paper and to Mr. Mark Robinette, Mr. Kevin Robinette, and Mr. Bruce Birza who assisted with the testing activities. Special appreciation is extended to Mr. Mark Robinette and Mr. Kevin Robinette for contributing to the production of this paper.

REFERENCES

1. "1996 Year Book, The Tire and Rim Association Inc." The Tire and Rim Association, Inc., 1996
2 ."N.A.D.A. Official Used Tire Guide", Mountain States Edition, N.A.D.A, 1977-1996
3. "Expert Autostats", 4N6XPERT SYSTEMS, La Mesa, California
4. Limpert, Rudolf, "Motor Vehicle Accident Reconstruction and Cause Analysis", Third Edition, 1986
5. "Highway Vehicle Object Simulation Model - 1976 Users Manual", Report FHWA-RD-76-162 Calspan Corporation, under USDOT Contract DOT-FH-11-8265
6. "Vehicle Analysis Package, EDSVS Program Manual", Engineering Dynamics Corporation, Beaverton, Oregon
7. Daily, John, "Fundamentals of Traffic Accident Reconstruction", Institute of Police Technology and Management, Jacksonville, Florida, 1988

PHOTO NO. 1

STANDARD TEST PROCEDURE PHOTOS SHOWING 1979 MADZA WITH TRAVEL PATH DEFINED BY DEFLATED TIRE AND PAINTED WHITE LINE. (photos 1-7)

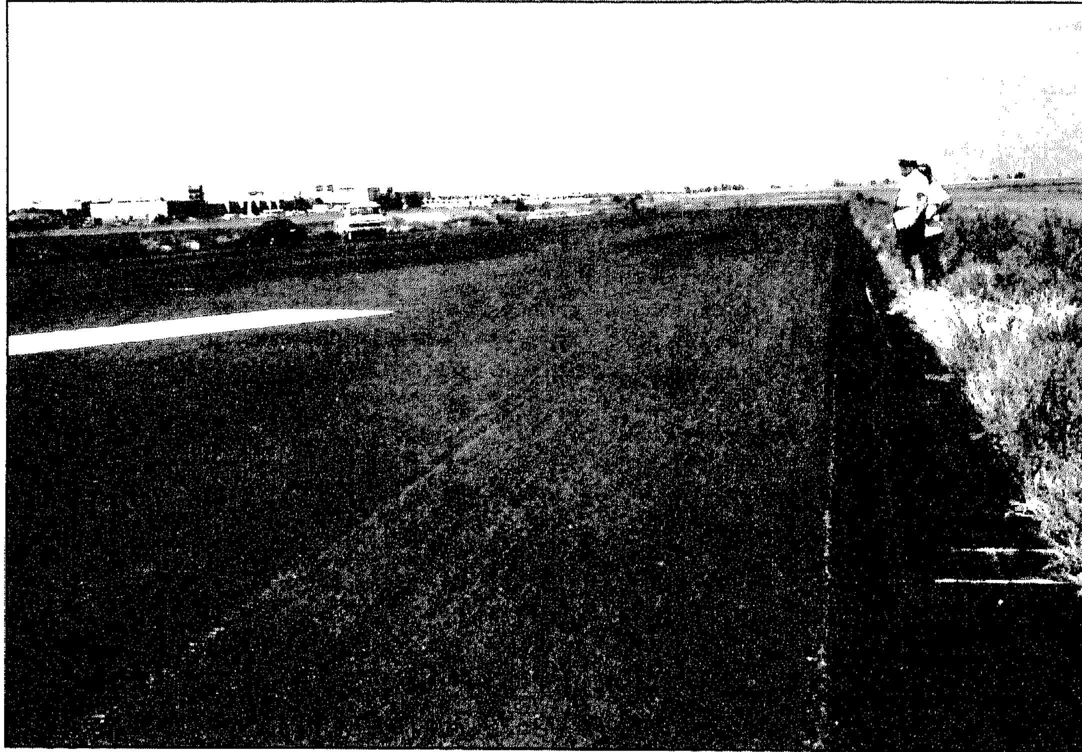


PHOTO NO. 2

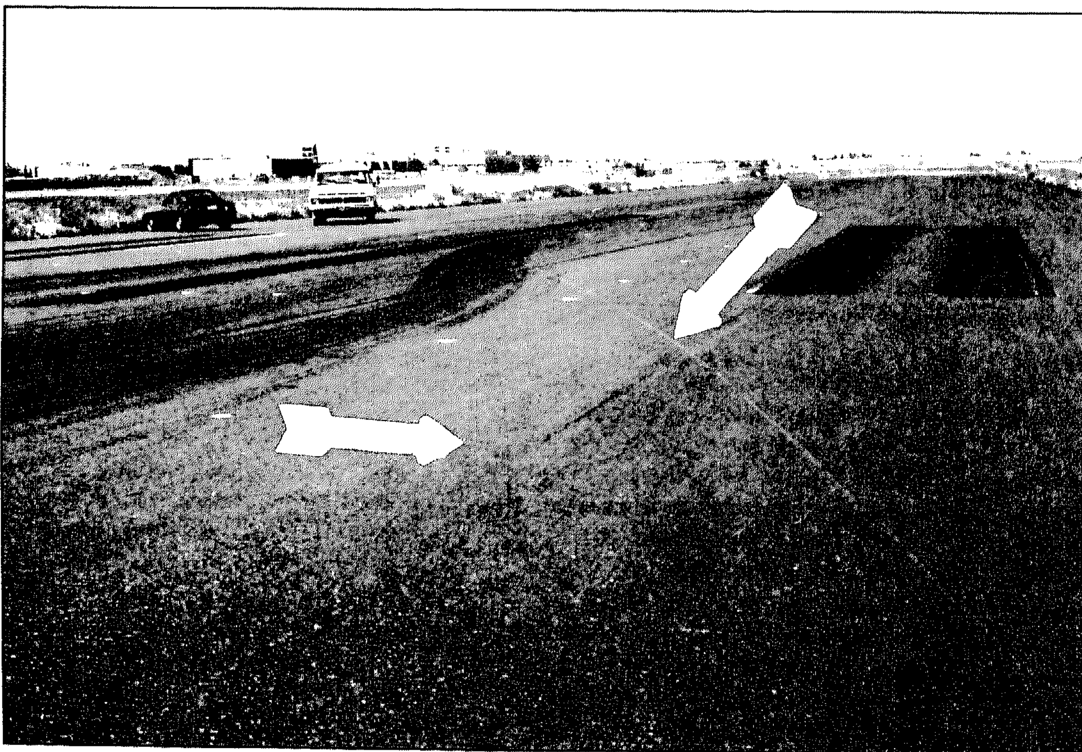


PHOTO NO. 3

STANDARD TEST PROCEDURE PHOTOS SHOWING 1979 MADZA WITH TRAVEL PATH DEFINED BY DEFLATED TIRE AND PAINTED WHITE LINE. (photos 1-7)

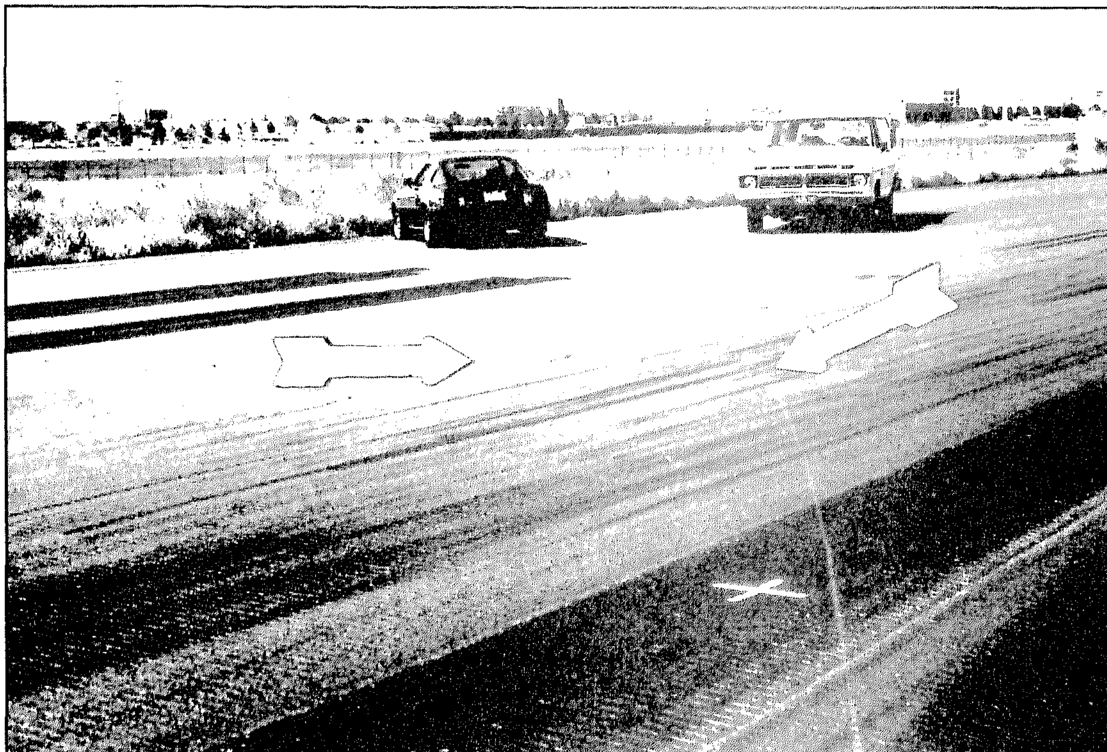


PHOTO NO. 4



PHOTO NO. 5

STANDARD TEST PROCEDURE PHOTOS SHOWING 1979 MADZA WITH TRAVEL PATH DEFINED BY DEFLATED TIRE AND PAINTED WHITE LINE. (photos 1-7)

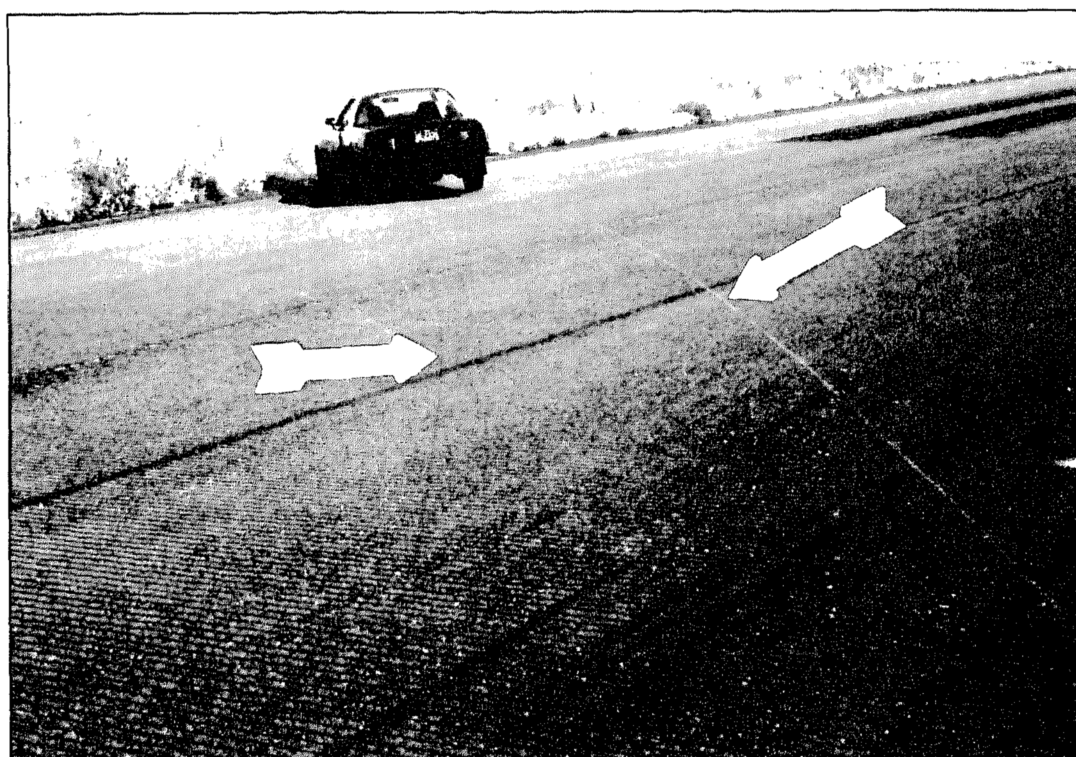


PHOTO NO. 6



PHOTO NO. 7

STANDARD TEST PROCEDURE PHOTOS SHOWING 1979 MADZA WITH TRAVEL PATH DEFINED BY DEFLATED TIRE AND PAINTED WHITE LINE. (photos 1-7)



DEFLATED LEFT REAR TIRE

SAE #970954
Reviewer's Discussion
by Philip V. Hight, P.E., Accident Research Santa Barbara
Drag and Steering Effects of Under Inflated and Deflated Tires
Richard J. Fay, Ric D. Robinette, Darrell Derring, Authors

This paper reports the results of automobile deceleration tests and lateral displacement experienced by vehicles with low tire pressure and also deflated tires. It describes clearly the methodology that was used. The findings generally were that the vehicles with deflated tires behaved in a mundane manner when it came to vehicles handling and control.

In the real world punctures occur more frequently to the rear tires than the front tires. Currently travel speeds may be well in excess of 88 km/h (55 mph) when tire deflation occurs. Although the authors have stated the tests speeds in their paper it is important for the researchers not to blindly extrapolate the findings to high speeds. Research has shown that when deflation occurs on curves it also increases the probability of vehicle handling difficulties for the driver. Even on a straight road with a minor uphill upgrade, a van driver, with cruise control engaged, can get into oversteering troubles when a rear tire explodes. At higher speeds the vehicles may still maintain static stability characteristics but with increased speed there is a change to the steady state path curvature stiffness. With increased speed the tire slip angle, force and moment assumptions may not remain linear. Other changes that occur at speed is that the circumference of the tires become slightly enlarged, more circular at the road contact location with less foot print length and change of contact patch geometry. When a rear puncture occurs the cornering stiffness at the rear of the vehicle is much reduced. Controlling an over steering vehicle at high speed is much more difficult than at lower speed.

When the front tire punctures then the trim changes due to aligning torque effects. The vehicle becomes more under steering but becomes poor for avoidance. Just because the vehicle is dynamically stable it does not mean that after the puncture its heading is correct. It would certainly be most helpful to have a follow up paper at much higher speeds and using a fifth wheel to obtain deceleration data. One would then not have to rely on the vehicle rolling at low deceleration values to a complete stop.

Reference:

R.J. Grogan, TR Watson, "Tyre Puncture - How, Why and Where," J. Forensic Sci Soc 1974, 14, 165.

F. Dell'Amico, "Steady - State Stability of the Automobile," Calspan Technical Report YC-3004-K-1, 1971.

WF Mulliken and DL Milliken, **Race Car Vehicle Dynamics**, TL 243.M55 1995.

SAE #970954
Reviewer's Discussion
by Donald F. Rudny, Rudny and Sallmann Engineering. Ltd.
Drag and Steering Effects of Under Inflated and Deflated Tires
Richard J. Fay, Ric D. Robinette, Darrell Derring, Authors

Although this paper provides some useful data for use in accident reconstruction analysis, it best serves as a guide to establishing a test methodology. The variability encountered in the results suggests that if critical conclusions are relying on such data, independent testing should be conducted. The authors have provided good basic test procedures and baseline data for others to conduct their own individual tests with some confidence.

In the future it might be interesting to include testing of locked wheels totally deflated and with variable inflation pressure.

970955

Project Y.A.M. (Yaw Analysis Methodology) Vehicle Testing and Findings - Victoria Police, Accident Investigation Section

Peter Bellion
Victoria Police

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A vehicle oversteered or cornering at excessive speed results in the tyres of the vehicle loosing traction with the road surface. As a result tyre yaw marks may be left on the road surface. Yaw marks are common at fatal collision sites.

Various methods are reported to estimate the speed of the vehicle that leaves yaw marks on a road surface. The difference in each method is how to determine the radius and whether the peak or average friction is used.

Tests were conducted with four different vehicles. Variations in tyre pressures, and driver inputs of acceleration, braking and steering over-correction were investigated. Yawing followed by emergency braking with and without ABS was further studied.

The radar speed of the vehicle for each test was compared with speed estimates from the yaw marks.

INTRODUCTION

Sliding of a tyre on a motor car on a road surface occurs not only when skidding during an emergency braking situation but also sometimes when steering. When a tyre on a car slides on the road in a steering situation the tyre mark that is left on the road surface is known as a tyre scuff mark, yaw mark or critical speed scuff mark. The tyre yaw marks are curved and will normally have striations angled across their width. Following Newton's Laws, a motor car once set in motion, tends to keep going in a straight line. When a driver turns the front wheels of the car at an angle, they would have to slide sideways if the car was to go straight ahead. Friction acting between the tyres and the road prevents such sliding and moves the front of the car in the direction it is steered, causing the car to follow a curved path. The faster the car travels, the more frictional force is needed to redirect the car into a curved path. If the speed of the vehicle is high enough, the force tending to keep the car going straight ahead (inertia) exceeds the friction acting between the tyres of the car and the road surface and as a result the tyre begins to slide sideways whilst still rotating. If the radius of the curved path of the car is known when the car started to slide out to the side and the friction (traction) force resisting slippage is also known it is possible to estimate the speed of the vehicle at which it commenced to sideslip. From the principles of circular motion the metric equation derived to calculate the speed at which side sliding occurs is shown in equation 1.

$$S = Sqrt[127R(f+e)] \dots (1)$$

where:

- S = Speed in kilometres per hour
- R = Radius of curved path of car in metres
- f = coefficient of friction
- e = grade or crossfall of road in direction of slippage of tyre in metre/metre.

*Note (e) is positive if grade/crossfall assists the car in passing around the curve and negative if it acts against the car.

In Accident Reconstruction the equation used to calculate radius from a chord (straight line intersecting two points on a curve) and middle ordinate (line which passes out at a right angle from centre of chord to the curve) is as follows in equation 2.

$$R = C^2/8m + m/2 \dots (2)$$

where:

- R = radius in metres
- C = chord measurement in metres
- m = middle ordinate measurement in metres

ASPECTS LOOKED AT DURING TESTING

Various measurement methods are reported in previous literature, [3,4,6,13,14,20], in respect to speed determination from yaw marks. To analyse various measurement methods to determine radius using equation (2) and whether the peak or average friction value should be used in conjunction with the determined radius to calculate speed, using equation (1), the Victoria Police, Accident Investigation Section conducted tests at Avalon Airport, Avalon, Victoria on the 13th, 14th, 15th and 28th of February, 1996. The tests incorporated 4 different test vehicles which were all examined by qualified police mechanical investigators attached to the Accident Investigation Section and found to be roadworthy prior to testing. The vehicles used were:-

- 1995 VR Holden Commodore sedan, 3.8 litre, 6 cylinder, automatic transmission, fitted with Goodyear Eagle GA tyres.
- 1995 ED Ford Falcon sedan, 4.0 litre, 6 cylinder, automatic transmission, fitted with Goodyear Eagle GA tyres.
- 1994 TR Mitsubishi Magna Executive sedan, V6 engine, automatic transmission, fitted with Goodyear Eagle GA tyres.
- 1995 Toyota Corolla Seca hatchback sedan, 4 cylinder, automatic transmission, fitted with Bridgestone RD-339 tyres.

Along with the various vehicles tested, tyre pressures were also varied from 26psi to 34 psi to see what effects this would have. Various inputs of acceleration, braking and steering over-correction were also looked at during testing. The effects of yawing and braking with and without anti-lock braking systems were also looked at during testing.

TEST NUMBERING, VEHICLE, MANOEUVRE, TYRE PRESSURE

Table (1) shows how the tests were numbered in respect to vehicle type, vehicle manoeuvre which led to the production of yaw marks on the road surface and the pressure in the tyres for that vehicle during the test. The test number is recorded in Column 1 of the result spreadsheets in Appendix 1.

Table 1 Test Numbering

Test No	Vehicle	Manoeuvre	Tyre Pressure
1-22	Commodore	straight->oversteer	28psi
23-36	Falcon	straight->oversteer	28psi
39-52	Magna	straight->oversteer	28psi
70-83	Corolla	straight->oversteer	28psi
53-57	Commodore	straight->oversteer, constant acceleration.	34psi
58-61	Commodore	straight->oversteer->50% brakes	34psi
62	Commodore	straight->oversteer->locked brakes	34psi
63-64	Commodore	straight->oversteer->accelerate	34psi
65-69	Commodore	straight->oversteer left then right	34psi
84-85	Falcon	straight->oversteer left then right	28psi
86-91	Falcon	straight->oversteer right	26psi
92-94	Falcon	straight->oversteer left then right	28psi
93ii-94ii	Commodore	corner at excessive speed	28psi
95	Commodore	corner at excessive speed, constant steer	28psi
96-98	Commodore	corner at excessive speed, accelerating	28psi
99	Falcon	corner at excessive speed, const. acc.	28psi
100-101	Falcon	corner at excessive speed, accelerating	28psi
102-103	Falcon	corner at excessive speed, const. acc.	28psi
104	Falcon	corner at excessive speed, accelerating	28psi
105	Falcon	corner ex. speed, braking with ABS	28psi
106	Falcon	corner ex. speed, brake to lock up.	28psi
107-110	Commodore	straight->oversteer, split friction surface	28psi

SPEED OF TEST VEHICLE COMING INTO STEERING MANOEUVRE

A Falcon hand held radar, speed measuring device was used to measure the speed of each test vehicle coming into the manoeuvre which led to the production of yaw marks. The radar was held at approximately 90 degrees to the front of the vehicle coming into the steering manoeuvre to minimise radar error. For most of the tests the speedometer reading of the vehicle was recorded for comparison with the radar reading. Columns 2,3 of test result spread sheets contained in Appendix 1 shows the radar speed and speedometer speed respectively. Tests were conducted from speeds of 40 kilometres per hour, incrementing by 10 kilometres per hour up to 100 kilometres per hour. Some of the cornering tests were conducted at speeds over 100 kilometres per hour.

CHORD & MIDDLE ORDINATE MEASUREMENTS AT TEST SITE

After each test 15, 20 and 30 metre chords were used to determine the radius of the front and rear, outside tyre yaw marks. Middle ordinate measurements taken from the mid point on each chord to the outer edge of both the front and rear outside tyres were then taken and recorded. Radius of curvature of the yaw marks were then calculated for each chord and middle ordinate measurement using Equation (2). " Tajima", 100 metre, "Symron" and "Engineer Super", yellow fibreglass tape and nylon-coated steel tapes, accurate between -40 degrees Celsius to 60 degrees Celsius were used for chord measuring, 3 police members ensuring tape is straight. Steel pocket tape measures were used for middle ordinate measurements. Photographs (1) and (2) show examples of the measuring sequence described above.



Photo 1, Chord & Middle Ordinate Measuring



Photo 2, Middle Ordinate Measuring

DETERMINATION OF FRICTION INCLUDING GRADE/CROSSFALL DURING TESTING

A VS model Holden Commodore sedan was utilised for skid testing in direction of slippage of test vehicle tyres on each day of the testing. This vehicle was fitted with "Goodyear Eagle GA Tyres". Although this vehicle was different to the test vehicles, previous research and testing [28], indicated that the difference in drag factor would lead to negligible differences in speed results under dry conditions. Dry road surface conditions were encountered for the first 3 days of testing, however on the last day of testing slightly wet conditions were encountered, which led to lower drag factors. Skid tests were conducted on two or more occasions on each day of the testing. A VC2000PC, Portable Road Test, Dynamometer and Braking Test Computer giving output of time, speed, distance and peak/average friction at 1/100 second intervals was set up inside the skid test vehicle to record measurements for the skid tests. A "drag sled" device was also used, on occasions, to record drag factor measurements in order to compare results obtained with the VC2000PC.

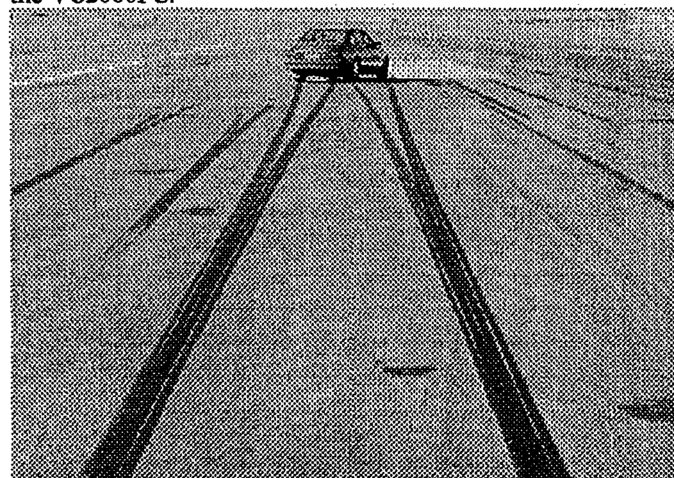


Photo 3, VS Commodore and Skid Marks.

The surfaces encountered during the testing were predominantly of an open graded, asphaltic concrete type, which had a smooth macro-texture and a rough micro-texture. This type of surface was consistent with a number of asphaltic concrete type surfaces found in the Melbourne, Metropolitan Area of Victoria, Australia.

The split friction tests were conducted on grass and a sealed taxiway surface. This sealed surface had a layer of liquid bitumen sprayed over its top which gave lower drag factors. The other type of sealed surface encountered was a Portland cement type concrete surface, which also had a smooth macro-texture with a rough micro-texture.



Photo 4, The VC2000.

Table 2, Drag Factors used in Speed Calculations

Day	Location	VC2000 Peak Factor	VC2000 Average Factor	c.f. Drag Sled
1	main runway north/south	0.95	0.78	0.78av.
1	taxiway split friction	0.67/0.61	0.46/0.45	
2	main runway north	0.93	0.77	0.84-0.93
2	main runway south	0.92	0.76	0.84-0.93
3	main runway north	0.93	0.78	
4	concrete taxiway (damp)	0.90	0.75	
4	cornering onto main runway	0.90	0.72	

RESULTS OF SPEED CALCULATIONS DETERMINED FROM CHORD AND MIDDLE ORDINATE MEASUREMENTS AT TEST SITE

The results of the speed calculations, determined from radius derived from chord and middle ordinate measurements, and the peak and average drag factors from the test site for each day are listed, in rows, starting with the test number on the spreadsheets in Appendix 1.

The second last column of data in the spreadsheets, Appendix 1, gives maximum percentage variation of calculated speed using peak drag factor from the radar speed. The maximum percentage variation relates to the chord measurement and tyre which gave the most deviation from the radar speed.

The last column shows for all chord lengths whether taken off the front or rear outside tyre, using the average drag factor, whether the calculated speed was less, equal or greater than the radar speed of the vehicle coming into the manoeuvre.

Line graphs of results have been prepared for the majority of the tests. Where prepared these are shown subsequent to the spreadsheet results also in Appendix 1.

PERCENTAGE ERRORS FROM SPEED RESULTS

The following tables show a summary of:-

- Sample number in each test group, for the given chord length used to determine radius,
- The percentage range for each test group, ie. the lowest to the highest percentage error from the radar speed for the speeds calculated from the given chord length measurements, eg. 15, 20 or 30 metre chord, off the front or rear tyre, and with either peak or average friction.
- The percentage of tests in the test group which gave a calculated speed, greater than radar speed, less than radar speed and equal to radar speed.

Table 3, Percentages, front tyre, 15 metre, peak.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	21	-8.6 to +11.4	85	10	5
23-36c/28	14	-3 to +32	71	14	15
39-52c/28	14	-1.4 to +9.7	93	7	0
53-57c/34	5	0 to +3.3	60	0	40
58-62b/34	5	-5.7 to +2.9	60	20	20
63-64a/34	2	0 to +5.3	50	0	50
65-69f/34	10	+3 to +38	100	0	0
70-83c/28	14	-8.1 to +6.5	78	14	8
84-85f/28	3	+10.1to +19.4	100	0	0
86-91c/26	6	-1.2 to +2.2	50	33	17
93-98cor/28	2	0 to +12.3	50	0	50
107-109s/28	3	-15 to +7.3	33	0	67

NB c = constant throttle, a = accelerating, b = braking, f = flick steering, cor = cornering, s = split friction and /28 refers to tyre pressure in pounds per square inch

Table 4, Percentages, front tyre, 15 metre, average.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	21	-11.3to +1.3	5	85	10
23-36c/28	14	-12.3to +20	7	93	0
39-52c/28	14	-10 to 0	0	93	7
53-57c/34	5	-9.8 to -5.5	0	100	0
58-62b/34	5	-15.7to -6.9	0	100	0
63-64a/34	2	-9.4 to -4	0	100	0
65-69f/34	10	-7.8 to +26.5	22	60	18
70-83c/28	14	-8.1 to +6.5	78	14	8
84-85f/28	3	0 to +7.4	67	0	33
86-91c/26	6	-12.8to -9.5	0	100	0
92-94f/28	6	-11 to +3.6	33	67	0
93-98cor/28	2	-9.1 to 0	0	50	50
107-109s/28	3	-29 to -9.8	0	100	0

Table 5, Percentages, front tyre, 20 metre, peak.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	22	-2 to +8.5	68	32	0
23-36c/28	13	-3.6 to +7.5	38	32	30
39-52c/28	14	-4.3 to +9.7	78	8	14
53-57c/34	4	-3.3 to 0	0	25	75
58-62b/34	5	-7.1 to +2.9	40	60	0
63-64a/34	2	0 to +1.3	50	0	50
65-69f/34	8	-3.9 to +14.3	87.5	12.5	0
70-83c/28	14	-5.8 to +4.7	35	35	28
84-85f/28	4	+7.2 to +13.4	100	0	0
86-91c/26	6	-6 to 0	67	33	0
92-94f/28	3	+2.7 to +6.8	100	0	0
93-98cor/28	4	-1 to +19.8	75	25	0
99-106cor/28	8	-10.3to +16.8	63	37	0
107-109s/28	3	-1.2 to -12.9	0	100	0

Table 6, Percentages, front tyre, 20 metre, average.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	22	-2.5 to -11.6	0	100	0
23-36c/28	13	-2.5 to +9.2	8	92	0
39-52c/28	14	-12.8to 0	0	92	8
53-57c/34	4	-13.1to -7.6	0	100	0
58-62b/34	5	-17.1to -8.3	0	100	0
63-64a/34	2	-10.8to -8	0	100	0
65-69f/34	8	-14.6to +4.4	25	75	0
70-83c/28	14	-13.4to -9.1	0	100	0
84-85f/28	4	-1.4 to +3	25	50	25
86-91c/26	6	-15.7to -9.7	0	100	0
92-94f/28	3	-8.2 to -5.1	0	100	0
93-98cor/28	4	-10 to +7.4	25	75	0
99-106cor/28a	3	-9.3 to +4.5	67	33	0
99-106cor/28c	3	-13.1to -2.3	0	100	0
99-106cor/28b	2	-19.6to -17.3	0	100	0
107-109s/28	3	-17 to -26	0	100	0

Table 7, Percentages, front tyre, 30 metre, peak.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
11-22c/28	12	-4 to +4.3	34	58	8
27-36c/28	9	-7.7 to +1.1	11	78	11
43-52c/28	10	-5.7 to +3.4	30	70	0
54-57c/34	4	-3.9 to -1	0	100	0
60b/34	1	-10	0	100	0
63-64a/34	2	-1.4 to -1.3	0	100	
65-69f/34	5	0 to +11	80	0	20
74-83c/28	10	-10 to +1	20	80	0
84-85f/28	2	+4.3 to +8.9	100	0	0
86-91c/26	6	-7.1 to 0	0	83	17
92-94f/28	3	0 to +9	67	0	33
93-98cor/28	4	-2.7 to +9.9	75	25	0
99-106cor/28a	3	0 to +7.9	67	0	33
99-106cor/28c	3	-7.9 to +1.1	33	67	0
99-106cor/28b	2	-11.3to -8	0	100	0
107-109s/28	1	-11	0	100	0

NB c = constant throttle, a = accelerating, b = braking, f = flick steering, cor = cornering, s = split friction and /28 refers to tyre pressure in pounds per square inch

Table 8, Percentages, front tyre, 30 metre, average.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
11-22c/28	12	-13 to -5.4	0	100	0
27-36c/28	9	-13.8to -3.1	0	100	0
43-52c/28	10	-14.3to -6.3	0	100	0
54-57c/34	4	-11.7to -9.9	0	100	0
60b/34	1	-10.8	0	100	0
63-64a/34	2	-11.9to -10.8	0	100	0
65-69f/34	5	-10.4to +1.1	20	80	0
74-83c/28	10	-20 to -11.3	0	100	0
84-85f/28	2	-5.8 to -1.5	0	100	0
86-91c/26	6	-18.6to -12.9	0	100	0
92-94f/28	3	-11 to -2.6	0	100	0
93-98cor/28	4	-11.8to -1.2	0	100	0
99-106cor/28a	3	-10.3to -3.4	0	100	0
99-106cor/28c	3	-17.1to -9.1	0	100	0
99-106cor/28b	2	-20.6to -17.3	0	100	0
109s/28	1	-26	0	100	0

Table 9, Percentages, rear tyre, 15 metre, peak.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	21	-1.5 to +8.9	86	4	10
23-36c/28	14	-4.6 to +15.3	71	22	7
39-52c/28	14	+1.6 to +15.7	100	0	0
53-57c/34	5	-3.3 to +1.3	40	40	20
58-62b/34	5	-5.7 to +4.3	40	20	40
63-64a/34	2	+1.3 to +2.7	100	0	0
65-69f/34	10	,0 to +6.8	70	0	30
70-83c/28	14	-20.6to +9.8	93	7	0
84-85f/28	3	+4.3 to +10.4	100	0	0
86-91c/26	6	-2.2 to +3.5	17	83	0
92-94f/28	6	-11.9to +9.6	83	17	0
93-98cor/28	3	-6.4 to +18.5	67	33	0
107-109s/28	3	-14 to +17	33	67	0

Table 10, Percentages, rear tyre, 15 metre, average.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	21	-10.1to -1.3	0	100	0
23-36c/28	14	-13.9to +4.1	7	93	0
39-52c/28	14	-6.8 to +4.3	7	79	14
53-57c/34	5	-7.7 to -6.5	0	100	0
58-62b/34	5	-15.3to -7.3	0	100	0
63-64a/34	2	-8.1 to -8	0	100	0
65-69f/34	10	-10.7to -2	0	100	0
70-83c/28	14	-29.9to -3.1	0	100	0
84-85f/28	3	-5.8 to 0	0	67	33
86-91c/26	6	-14.3to -10.2	0	100	0
92-94f/28	6	-21.4to -6.1	0	100	0
93-98cor/28	3	-14.6to +6.2	33	67	0
107-109s/28	3	-28 to -1.2	0	100	0

NB c = constant throttle, a = accelerating, b = braking, f = flick steering, cor = cornering, s = split friction and /28 refers to tyre pressure in pounds per square inch

Table 11, Percentages, rear tyre, 20 metre, peak.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	22	-4.3 to +6.7	59	23	18
23-36c/28	13	-6.2 to +7.5	54	31	15
39-52c/28	14	-5.7 to +9.8	64	29	7
53-57c/34	5	-9.9 to -1.3	0	100	0
58-62b/34	5	-7.1 to +1.4	20	60	20
63-64a/34	2	0 to +2.7	50	0	50
65-69f/34	8	-9.4 to +8.8	75	12.5	12.5
70-83c/28	14	-20.6to +8.7	57	29	14
84-85f/28	4	+2.9 to +10.5	100	0	0
86-91c/26	6	-4.3 to 0	0	83	17
92-94f/28	3	+1.4 to +6.9	100	0	0
93-98cor/28	6	-16 to +23.5	83	17	0
99-106cor/28a	3	+4.1 to +14.6	100	0	0
99-106cor/28b	2	-6 to -3	100	0	0
99-106cor/28c	3	-5.2 to +7.9	67	33	0
107-109s/28	3	-17.7to +2.4	33	67	0

Table 12, Percentages, rear tyre, 20 metre, average.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
1-22c/28	22	-13 to -3.4	0	100	0
23-36c/28	13	-15.4to -2.2	0	100	0
39-52c/28	14	-14.3to 0	0	93	7
53-57c/34	5	-16.9to -7.7	0	100	0
58-62b/34	5	-17.1to -8.3	0	100	0
63-64a/34	2	-9.3 to -8.1	0	100	0
65-69f/34	8	-8.8 to 0	0	87.5	12.5
70-83c/28	14	-30.9to -5.4	0	100	0
84-85f/28	4	-7.3 to 0	0	75	25
86-91c/26	6	-15.7to -12.8	0	100	0
92-94f/28	3	-9.6 to -5.1	0	100	0
93-98cor/28	6	-23.2to +9.9	16.5	67	16.5
99-106cor/28a	3	-7.2 to +2.3	50	50	0
99-106cor/28b	2	-15.5to -13.2	0	100	0
99-106cor/28c	3	-15.8to -3.4	0	100	0
107-109s/28	3	-30.6to -14.6	0	100	0

Table 13, Percentages, rear tyre, 30 metre, peak.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
11-22c/28	12	-3.8 to +1.1	8	67	25
27-36c/28	9	-10.8to +2	33	67	0
43-52c/28	10	-7.1 to +3.3	30	70	0
54-57c/34	4	-6.5 to -3	0	100	0
60b/34	1	-8.5	0	100	0
63-64a/34	2	-4 to -2.7	0	100	0
65-69f/34	5	0 to +8.8	40	0	60
74-83c/28	10	-10 to +4.3	20	60	20
84-85f/28	2	+4.4 to +9	100	0	0
86-91c/26	6	-10 to -3.3	0	100	0
92-94f/28	3	-4.1 to +7.7	67	33	0
93-98cor/28	6	-13.6to +13.6	50	33	17
99-106cor/28a	3	0 to +7.8	67	0	33
99-106cor/28c	3	-9.2 to 0	0	33	67
99-106cor/28b	2	-7.2 to -5.1	0	100	0
109s/28	1	-13	0	100	0

NB c = constant throttle, a = accelerating, b = braking, f = flick steering, cor = cornering, s = split friction and /28 refers to tyre pressure in pounds per square inch

Table 14, Percentages, rear tyre, 30 metre, average.

Test No.	Sample Number	% range	% > radar	% < radar	% = radar
11-22c/28	12	-13.9to -8.6	0	100	0
27-36c/28	9	-18.5to -7.1	0	100	0
43-52c/28	10	-15.7to -6.7	0	100	0
54-57c/34	4	-24 to -11.9	0	100	0
60b/34	1	-18.6	0	100	0
63-64a/34	2	-13.3to -12.2	0	100	0
65-69f/34	5	-10.4to -1.1	0	100	0
74-83c/28	10	-20 to -8.7	0	100	0
84-85f/28	2	-5.8 to -1.5	0	100	0
86-91c/26	6	-21.4to -14	0	100	0
92-94f/28	3	-13.7to -3.8	0	100	0
93-98cor/28	6	-21.6to +1.2	17	83	0
99-106cor/28a	3	-10.3to -5	0	100	0
99-106cor/28c	3	-18.4to -10.2	0	100	0
99-106cor/28b	2	-16.5to -15.3	0	100	0
109s/28	1	-27	0	100	0

COMMENTS ON CHORD & MIDDLE ORDINATE SPEED RESULTS

Front Tyre, 15 metre chord, peak friction

More often than not the calculated speed was greater than the radar speed, especially for the flick steering tests, where the vehicle was being oversteered from one direction to the other. There was no test group using this method which always gave conservative results. The maximum range for all tests with this method was -15% to +38%.

Front Tyre, 15 metre chord, average friction

More often than not the calculated speed was less than the radar speed, however problems were encountered once again in respect to both the Commodore and Falcon

flick steer tests. In five (5) out of the (13) test groups where this method was able to be applied, 100% of the time the calculated speed was less than the radar speed. These were:

- Falcon, 26 psi, constant throttle
- Commodore, 34 psi, constant throttle
- Commodore, 34 psi, braking
- Commodore, 34 psi, accelerating
- Falcon, all cornering tests.

Maximum range for this method was -29% (Commodore, split friction) to +26.5% (Commodore, flick steering).

Front tyre, 20 metre chord, peak friction

More often than not the calculated speed was greater than the radar speed, exceptions were:

- Commodore, 34 psi, constant throttle
- Seca, 28 psi, constant throttle
- Falcon, 26 psi, constant throttle
- Commodore, split friction tests.

In one (1) split friction group, out of the (14) test groups, 100% of time, calculated speed was less than radar speed. The maximum range encountered for this method was -10.3% (Falcon cornering) to +19.8% (Commodore cornering).

Front tyre, 20 metre chord, average friction

More often than not the calculated speed was less than the radar speed. In ten (10) out of the (16) test groups this occurred 100% of the time. General exceptions to the rule were the Falcon cornering at speed whilst under acceleration and all flick steer tests, most likely due to the fact that the front tyre starts to follow a straighter path earlier than the rear tyre in these circumstances.

The maximum range encountered for this method was -26% (Commodore, split friction) to +9.2% (Falcon, 28 psi, constant throttle, oversteer).

Front tyre, 30 metre chord, peak friction

The 30 metre chord was restricted in it use as the yaw marks left from speeds of 60 to 70 kilometres per hour and below did not enable a 30 metre chord to be used because of the overall length of the yaw marks, however when used for speeds of approximately 60 kilometres per hour up to approximately 100 kilometres per hour the method gave good results. Five (5) out of the (16) test groups had calculated speeds less than the radar speed 100% of the time. Problems tests found with this method were:

- Commodore flicks
- Falcon flicks
- Commodore cornering at speed
- Falcon cornering at speed & accelerating.

Maximum ranges encountered for this method were -11.3% (Falcon cornering & braking) to +11% (Commodore flicks).

Front tyre, 30 metre chord, average friction

Most of the time, fifteen (15) out of sixteen (16) test groups, this method gave a calculated speed less than the radar speed, 100% of the time. The exception was the Commodore flick tests, however 80% of the time for these tests the calculated speed was less than the radar speed.

Maximum ranges encountered for these tests were -26% (Commodore split friction) to +1.1% (Commodore flicks).

Rear tyre, 15 metre chord, peak friction

More often than not the calculated speed was greater than the radar speed, except for Falcon, 26 psi tests and Commodore split friction tests.

Maximum range encountered for this method was -20.6% (Seca last test) to +18.5% (Commodore cornering).

Rear tyre, 15 metre chord, average friction

More often than not the calculated speed was less than the radar speed. In nine (9) out of the (13) test groups this was the case 100% of the time. Far better results were obtained in respect to the flick tests on the rear tyre than for the front tyre with a 15 metre chord and peak friction. Main exception group was the Commodore cornering at speed tests where 33% of the time the calculated speed was greater than the radar speed. For speeds less than 70 kilometres per hour this method generally always gave conservative speed results.

Maximum range encountered for this method was -29.9% (Seca last test) to +6.2% (Commodore cornering)

Rear tyre, 20 metre chord, peak friction

This method gave a fairly average spread of calculated speed results above and below the radar speed. Main problem tests were:

- Commodore flicks
- Falcon flicks
- Commodore cornering at speed
- Falcon cornering at speed, constant throttle
- Falcon cornering at speed & accelerating.

Maximum range encountered for this method was -20.6% (Seca last test) to +23.5% (Commodore cornering @ 80 km/h).

Rear tyre, 20 metre chord, average friction

In fourteen (14) out of the (16) test groups where this method was able to be used, the calculated speed, 100% of the time, was less than the radar speed. The problem groups were:

- Commodore cornering at speed
- Falcon cornering under acceleration.

Maximum range encountered for this method was -30.6% (Commodore split friction) to +9.9% (Commodore cornering @ 80 km/h).

Rear tyre, 30 metre chord, peak friction

Once again this method restricted in its use for reasons already mentioned above. More often than not the calculated speed was greater than the radar speed, main problem groups were the Falcon, flicks and cornering under acceleration.

Maximum range was -13.6% to +13.6% (Commodore cornering at speed).

Rear tyre, 30 metre chord, average friction

Where used, 99% of the time the calculated speed was less than the radar speed. For the 1% overestimate, calculated speed did not exceed radar speed by more than 1.2%.

Maximum range was -27% (Commodore split friction) to +1.2% (Commodore cornering @ 80 km/h).

RECOMMENDATIONS FOR CHORD & MIDDLE ORDINATE MEASUREMENTS, FRICTION.

After analysing the test data the following is recommended in order to obtain conservative speed results, ie., to establish beyond reasonable doubt, that the speed of the vehicle calculated from the yaw marks is a minimum.

1. Generally 15 metre chords on the outside, rear, tyre mark, should be used where the subject vehicles speed is believed to be 70 kilometres per hour or less. The chord should begin at the start of the visible tyre yaw mark from the outside rear tyre.
2. Generally 30 metre chords should be used on the outside, rear, tyre mark, where the subject vehicles speed is believed to be greater than 70 kilometres per hour.
3. The average coefficient of friction taken in the direction of slippage of the tyres should be used. If this cannot be measured in the direction of slippage of the tyres from the subject vehicle, this value will have to be corrected for grade and or crossfall of the carriageway.
4. Average coefficient of friction should be determined by way of a skid test or tests at the collision site on the same surface as, and adjacent to the subject vehicles tyre yaw marks. A VC2000 braking test computer is an acceptable device for determining average friction. If skid tests are unable to be performed at the collision site an alternative friction measuring device should be used to establish friction. This could be either a drag sled or drag tyre device, however one should use their experience in examining and testing road surfaces to ensure that the friction readings obtained from the sled or tyre device are comparable with what one would expect the average friction to be for the surface under inspection.
5. A vehicle fitted with an Anti-lock Braking System (ABS) that leaves yaw marks on a road surface can have its speed estimated using equation (1). This is expected to provide a calculated speed which is less than the actual speed of the vehicle.
6. Close examination, photography and sketching of striation marks angled across a yaw mark should be done. The striation pattern will show what input there has been from the vehicle driver in terms of acceleration or braking. Acceleration striations are generally closely spaced and angle out rearwards (opposite to direction of travel of vehicle), whereas, braking striations are more spread out and angle forwards.
7. Examination of the striation pattern within the tyre marks will indicate where a yaw mark changes from a yaw mark into a skid mark. The striation pattern will go from angling across the mark to running parallel with the mark. If this occurs it is also likely that the centre of mass path of the vehicle will change from a curved path to a straight path. Accident Investigators and Reconstructionists should be fully aware of this as this means that the chord length will be restricted by the yaw section of the marks and not go into the skidding section of the marks.
8. Where a vehicle goes into a yawing condition, leaving tyre scuff marks which originated from the vehicle having travelled over split frictional surfaces, eg. left side grass or gravel, right side sealed surface, a resultant drag factor calculated using 2/3 of left side friction plus 1/3 right side friction input into equation (1) should give a conservative estimate.
9. Where a vehicle is oversteered from one direction to another, chord measurements should not be taken on the outside front tyre mark for the initial oversteer, yaw marks, as this will more often than not result in an overestimate of the vehicles speed.

SURVEY OF SCUFF MARKS TO PREPARE VEHICLE CENTRE OF MASS PLOTS

In addition, to chord and middle ordinate measurements taken at the test site, on the tyre yaw marks

produced from the test vehicles, some of the test yaw marks were surveyed using Geodimeter 400 and 600 model, Total Survey Stations, accurate to +/- 5mm. The reason for this was that the majority of collision sites which the Accident Investigation Section attend are surveyed using the Geodimeter. On completion of each survey the measurements are downloaded onto a computer loaded with 'Geocomp,' engineering/surveying/drafting software. Utilising this software package it is possible to determine radius or lineal measurements of surveyed marks at the original scale as which they were measured on the road, ie. 1:1 scale. This is performed using commands within the software package.

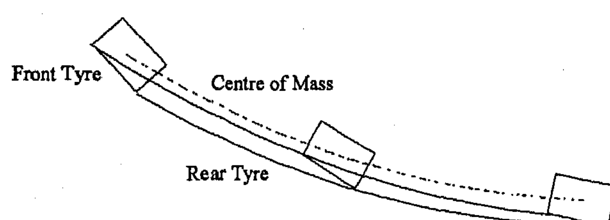


Diagram 1, An Example of Centre of Mass Plot, Falcon, Test No. 31.

RESULTS OF CENTRE OF MASS PLOTS

Wheelbase, track and centre of mass position diagrams of vehicles tested were plotted onto yaw marks surveyed, using the computer and 'Geocomp'. The centre of mass path plotted went from the commencement of the yaw marks where both the outside front and outside rear tyres coincided with the marks. Generally the centre of mass path did not exceed 30 metres in length. Utilising software commands the radius of the path followed by the centre of mass of the vehicle which left the marks was determined with the computer software tools. After the radius was determined, speeds were then calculated with both the peak and average friction using equation (1). Comparison of calculated speeds with peak/average friction was then made to radar speed. The results of this data analysis is contained in spreadsheet format in Appendix 2. Bar graphs have been prepared for illustration purposes, these are also contained in Appendix 2. Table 15 shows tests that were analysed with centre of mass plots.

Table 15, Tests Analysed by Centre of Mass

Vehicle	Manoeuvre	Tyre Pressure	Test Numbers
Commodore	straight then oversteer	28 psi	5,7,9,11,15,21
Falcon	straight then oversteer	28 psi	29,31,33,35
Magna	straight then oversteer	28 psi	48,49,51
Commodore	cornering at speed	28 psi	93,96,97,98
Falcon	cornering at speed	28 psi	99,101,102,103,104

COMMENTS ON CENTRE OF MASS RESULTS

Generally using this method gave more conservative speed results than for methods determined from chord and middle ordinate measurements taken off the front or rear outside tyre marks at the test site. This in fact should be the case as the centre of mass path should and did have a lower radius than that determined from the front or rear outside tyre mark.

Using the peak friction value the percentage error range for all 22 tests analysed using the centre of mass method was from -11% to +10%. 59% of the time the calculated speed was less than the radar speed, 14% of the time it was equal to the radar speed and 27% of the time it was greater than the radar speed.

Using the average friction value the percentage error range was from -20% to -1%, meaning 100% of the 22 tests analysed with this method had a calculated speed less than the radar speed.

Overall, although this method was not able to be applied to every test, it seems that conservative results would always be obtained if the average friction value is used.

If the peak friction value were used, in collisions cases, the actual speed of the vehicle would be expected to be within +/- 10% of the calculated speed.

SPEED COMPARED WITH RADIUS

The following is provided as a general rule or check for the accident investigator where a passenger car has left a set of yaw marks on a dry, asphaltic concrete or concrete surface.

Table 16, Speed km/h compared to Radius

Speed km/h	Approximate radius: yaw mark or centre of mass
60 km/h	25 metres
70 km/h	40 metres
80 km/h	50 to 60 metres
90 km/h	70 to 80 metres
100 km/h	80 to 90 metres
110 km/h	95 to 105 metres

MAXIMUM LATERAL ACCELERATION FACTOR

Maximum lateral acceleration factor was measured for the majority of the vehicle tests by VC2000PC which was mounted vertically on the centre interior of the vehicle windscreen. At the completion of each test the VC2000 was read and the maximum lateral acceleration factor was recorded. The average, maximum lateral acceleration factor was found to be slightly in excess of 1.0g.

CONCLUSION

In summary the testing carried out provided an extensive amount of experience for members of the Victoria Police, Accident Investigation Section. Estimating speed of passenger cars from yaw marks is a very useful tool for the Accident Investigator and Reconstructionist. Providing certain rules are followed the investigator can be reasonably confident that the estimate he is providing of the speed of a vehicle which has left a set of yaw marks at a collision site is a minimum. More often than not it is required for police to provide a minimum speed to the Court, especially in a criminal case where the standard of proof is beyond reasonable doubt.

ACKNOWLEDGEMENTS

All test work was carried out with the assistance of members of the Victoria Police, Accident Investigation Section all of which were dedicated to the task. The assistance of Inspector Neil O'Sullivan and Senior Sergeant Noel OSBORNE is appreciated in respect to organisation of the testing, coordinating legal indemnities and arranging for the troops to be fed during the four days. It is also correct to acknowledge former Acting Sergeant of Police, William KERAMIDAS who together with myself were the technical coordinators of the tests. Credit is further due to Mr. Rohan HARMAN of 'Drive Safe Australia' who was our driver for all tests. I would also like to thank, Acting Sergeant Cathy SMITH, of the Victoria Police, Policy and Education Section for her assistance in preparing the spread sheets contained in the appendices of this paper. I would also like to thank our sponsors, 'South Pacific Tyres' who provided money to pay for test location and facilities, and for providing 'Dunlop Monza 200 tyres' for ongoing testing as well as the Victoria Police, Transport Branch for arranging the test vehicles.

REFERENCES

1. Physics, revealing our world, Book one, Christina HART, Jacaranda Press 1991
2. Physics 2, Heinemann, Physics in Context
3. Lynn B. FRICKE. Traffic Accident Reconstruction, Vol. 2, Northwestern University Traffic Institute.
4. West Midlands Police, Accident Investigation Training Unit, Standard and Advanced Course Lesson Notes, City & Guilds Institute, Certificate in Road Accident Investigation.
5. Institute of Police Technology and Management, University of Florida, Course Notes.
6. AXUP. Chisholm Institute of Technology, Department of Civil Engineering, Police Studies Diploma, Road Traffic Accidents and Counter Measures.
7. Monash University, Caulfield Campus, 1993, Department of Police Studies, Accident Investigation Techniques 1. Course Notes.
8. Chisholm Institute of Technology,Nov 1988,Accident Investigation Techniques 1.
9. Accident Reconstruction, Speed Formula Derivations, TEEX, Law Enforcement and Security Training Division, Texas Engineering Extension Service, The Texas A&M University System.
10. M.G. LAY. Source Book for Australian Roads, Chapter 19, Road Geometry, Australian Road Research Board.
11. Road Construction Authority, Road Design Manual 1983, Page 2-08, Superevelevation and horizontal curve design.
12. Chisholm Institute of Technology, Bachelor of Civil Engineering Degree, Subject Code CIV317, Highway and Traffic Engineering, Course Notes.
13. Northwestern University Traffic Institute, Traffic Accident Reconstruction, Publication Number 806.
14. RF LAMBOURN. The calculation of motor car speeds from curved tyre marks. J Forensic Science Society 29 371-386, 1989.

15. BELLION. Speed Estimates from yaw marks, comparison between Lambourn's and Frickes's method, Conference Proceedings, Inaugural International Conference on Accident Investigation, Reconstruction Interpretation and the Law, Gold Coast Queensland, October. 1995.

16. Dr. Akhtar Qizilbash. To Skid or not to Skid-Tell Tale Tyre Marks, Central Queensland University, Rockhampton, Australia, presented at AIRIL '95 conference, Gold Coast, Queensland, 1995.

17. Mary REVELEY, Douglas BROWN & Dennis GUENTHER, Ohio State University, Columbus Ohio, A comparison study of skid and yaw marks, published in Motor Vehicle Accident Reconstruction, Review and Update, SP-777, International Congress and Exposition, Detroit Michigan, 1989, Society of Automotive Engineers.

18. John BROWN, Kenneth OBENSKI. Forensic Engineering Reconstruction of Accidents, 1990 Published by Charles C THOMAS, ISBN 0-398-05624-2.

19. Sergeant Mark GEORGE. Vehicle Curve Critical Speed Analysis, New South Wales Police, Published in Australian Police Journal, Jan-Mar 1991.

20. Albert T. BAXTER. An Examination of the Critical Speed Problem, Accident Reconstructionist, Published in Accident Reconstruction Journal, Vol 5, No. 6, Nov/Dec, 1993.

21. Luis MARTINEZ, Estimating Speed from Yaw Marks - An Empirical Study Published in Accident Reconstruction Journal, Vol 5, No. 3, May/June, 1993.

22. Letters to the Editor, Published in Accident Reconstruction Journal, Vol 6. No. 6, Nov/Dec, 1994.

23. Robert BROSHEARS, Richard SMITH. Drag Sled Handbook.

24. DICKERSON, ARNDT & MOWRY. Evaluation of Vehicle Velocity Predictions Using the Critical Speed Formula, Arndt & Associates, Ltd.

25. Richard F. LAMBOURN. Braking and Cornering Effects with and without Anti-Lock Brakes. SAE Technical Paper 940723.

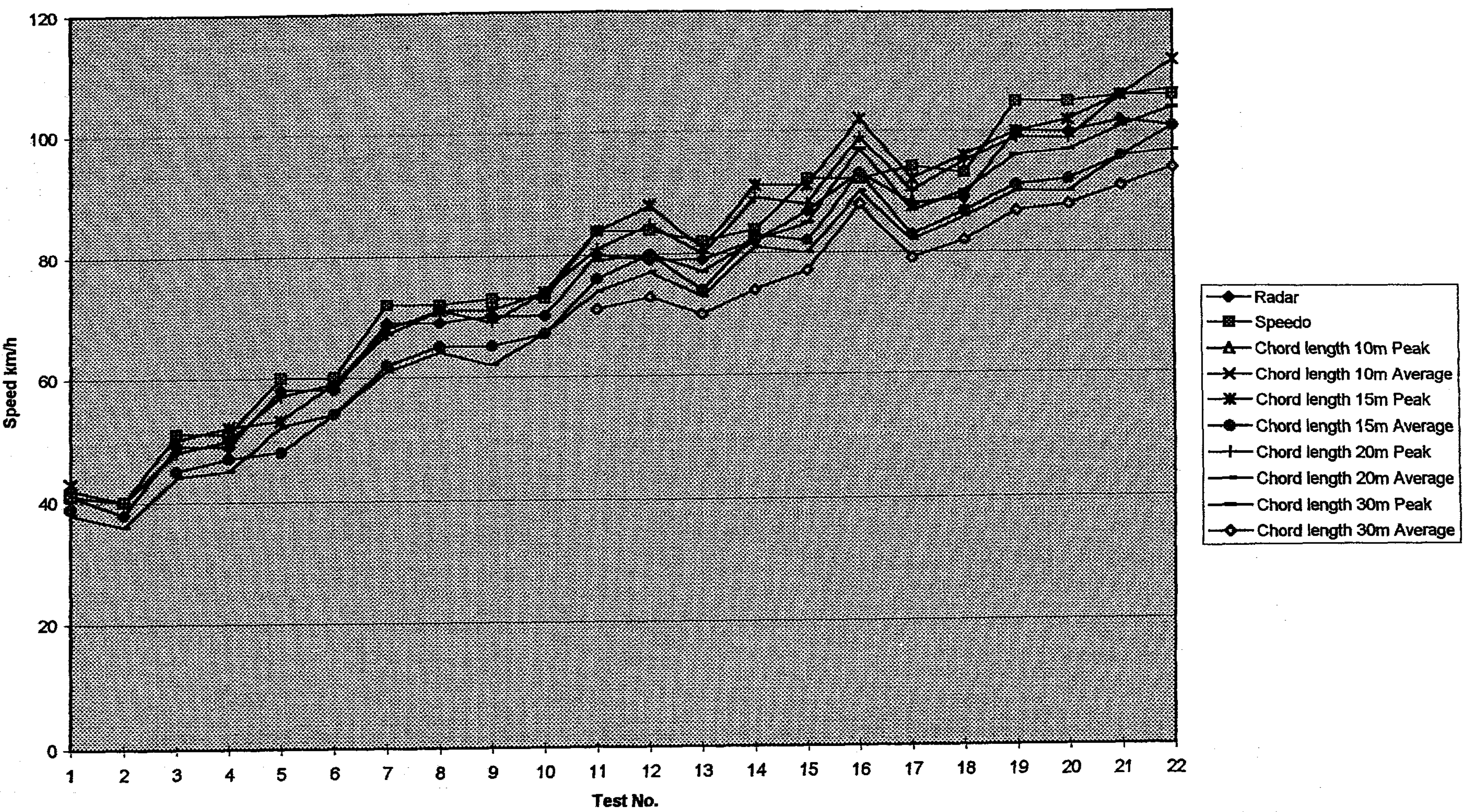
26. Dr Frank BULLEN and Mr John RULLER. Investigating Braking Influenced Crashes Using Data Extrapolated from In Situ Vehicle Tests. Road Accident Investigation Services Pty Ltd. Proceedings of the Inaugural, International

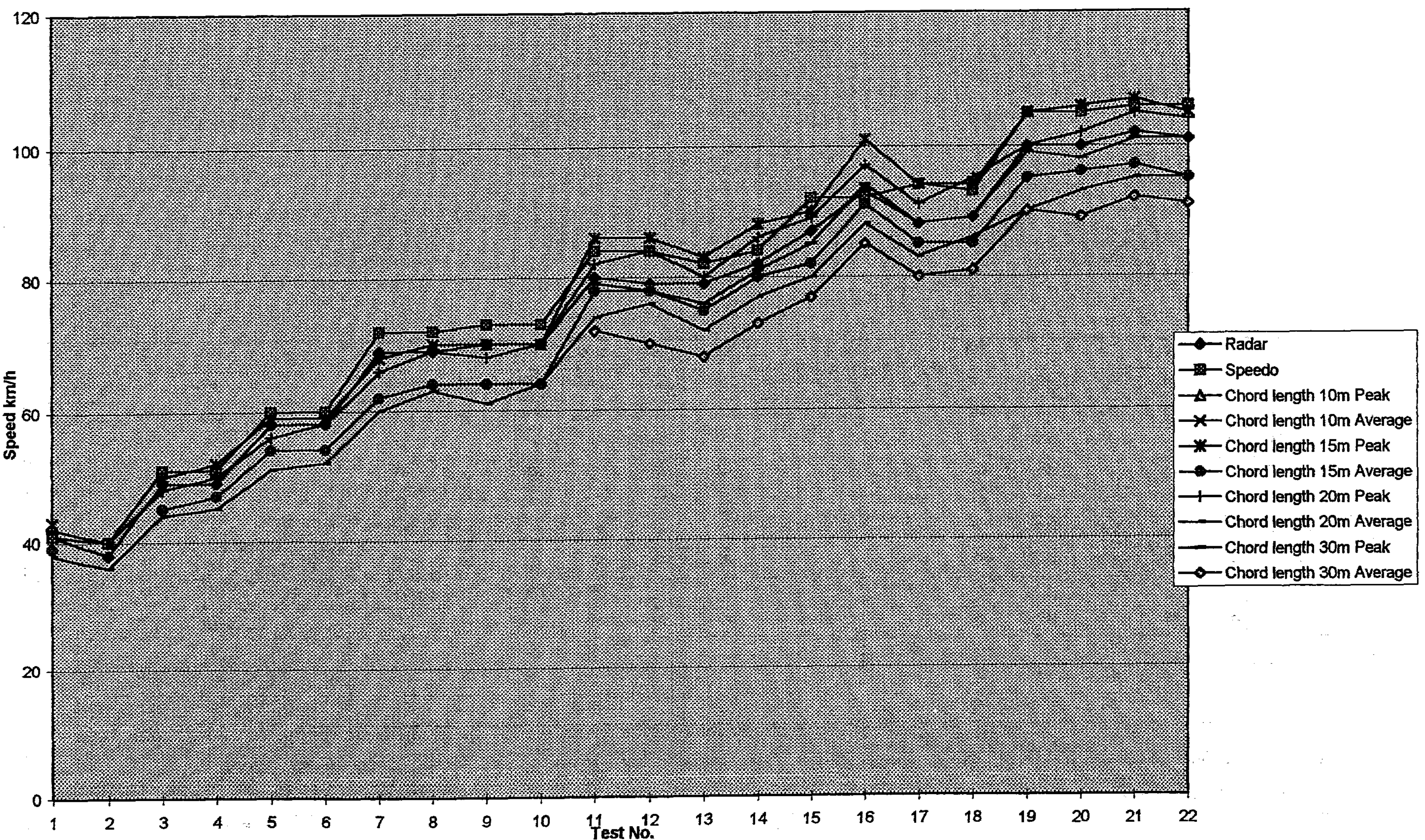
27. Conference on Accident Investigation, Reconstruction and the Law, October, 1995, Gold Coast, Australia.

APPENDIX 1
CHORD & MIDDLE ORDINATE METHOD SPEED RESULTS AND GRAPHS

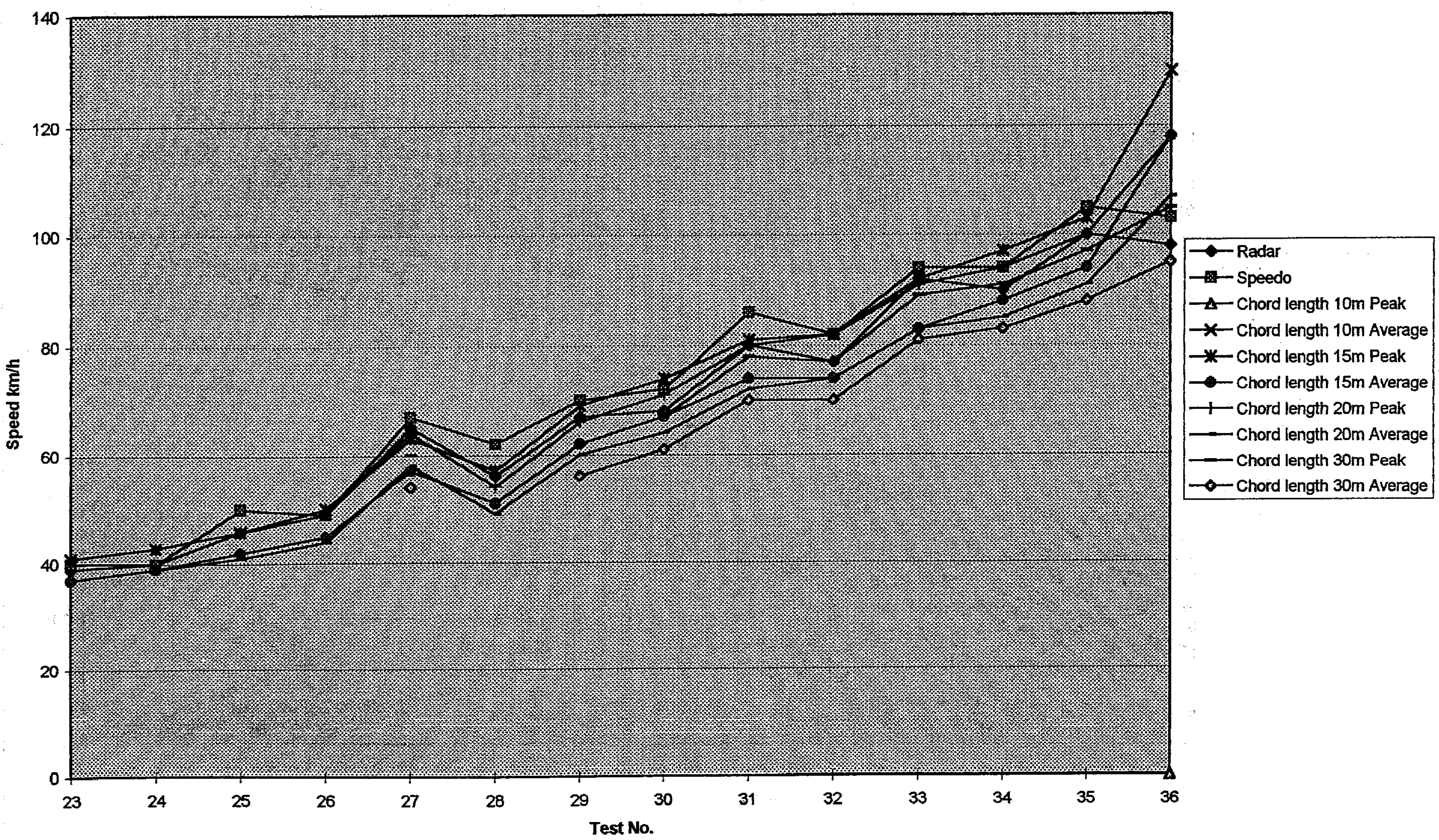
VR COMMODORE TESTS. TYRE PRESSURE AT 28PSI. 13 February 1996.

			Front Tyre								Rear Tyre									
Test No	Radar	Speedo	Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Comments	
			Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak %	Average
1	41	41			43	39	42	38					43	39	42	38			4.80%	<
2	38	40					40	36							40	36			5.20%	<
3	49	51			50	45	48	44					50	45	48	44			2%	<
4	49	51			52	47	50	45					52	47	50	45			6.10%	<
5	58	60			53	48	57	52					59	54	56	51			1.70%	<
6	58	60			59	54	59	54					59	54	58	52			1.70%	<
7	69	72			68	62	67	61					68	62	66	60			-1.40%	<
8	69	72			71	65	71	64					70	64	69	63			2.90%	<
9	70	73			71	65	69	62					70	64	68	61			1.40%	<
10	70	73			74	67	74	67					70	64	70	64			5.70%	<
11	80	84			84	76	81	74	79	71			86	78	82	74	79	72	7.50%	<
12	79	84			88	80	85	77	80	73			86	78	84	76	78	70	11.40%	<
13	79	82			81	74	80	73	77	70			83	75	80	72	76	68	5%	<
14	82	84			91	83	89	81	82	74			88	80	86	77	81	73	10.90%	<
15	87	92			91	82	88	80	85	77			90	82	89	80	85	77	4.60%	<
16	93	92			102	93	99	90	97	88			101	91	97	88	94	85	6.80%	<
17	88	94			92	83	90	82	87	79			94	85	91	83	88	80	7.80%	<
18	89	93			96	87	95	86	90	82			94	85	95	86	89	81	-2.25%	<
19	100	105			100	91	99	90	96	87			105	95	100	90	99	90	-9%	<
20	100	105			102	92	99	90	97	88			106	96	102	93	98	89	-8%	<
21	102	106			106	96	106	96	101	91			107	97	105	95	101	92	4.90%	<
22	101	106			112	101	107	97	104	94			105	95	104	95	101	91	10.90%	<=

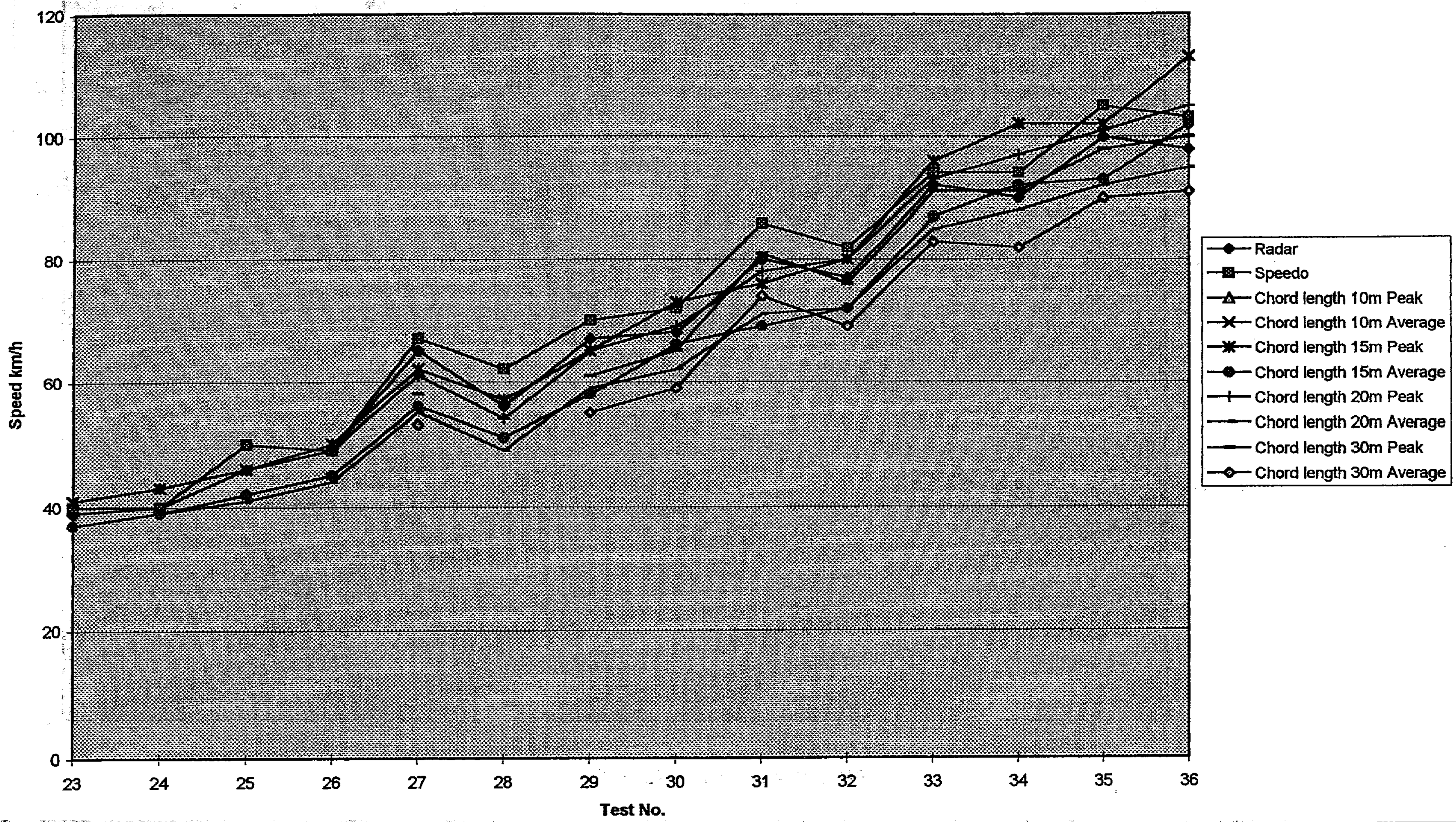

Commodore, Tests 1-22, Front Tyre
Speed km/h
0
20
40
60
80
100
120
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Test No.
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average


Commodore Tests 1-22, Rear Tyre
Speed km/h
120
100
80
60
40
20
0
Test No.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average

ED Falcon Tests. Tyre Pressure 28 PSI. Oversteer right. 13/14 February 1996																				
			Front Tyre								Rear Tyre									
Test No	Radar	Speedo	Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Comments	
			Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak %	Average
23	39	40			41	37							41	37					5%	<
24	40	40			43	39	43	39					43	39	43	39			7.50%	<
25	46	50			46	42	46	41					46	42	46	41			=	<
26	49	49			50	45	49	44					50	45	49	44			2%	<
27	65	67			63	57	64	58	60	54			62	56	61	55	58	53	-3%	<
28	56	62			57	51	54	49					57	51	54	49			2.90%	<
29	67	70			69	62	66	60	62	56			65	58	65	59	61	55	3	<
30	68	72			74	67	71	64	67	61			73	66	69	62	65	59	8.80%	<
31	80	86			81	74	80	72	78	70			76	69	78	71	81	74	1.25%	<
32	77	82			82	74	82	74	77	70			80	72	80	72	76	69	6.50%	<
33	92	94			92	83	91	83	89	81			96	87	93	85	91	83	4.30%	<
34	90	94			97	88	94	85	91	83			102	92	97	88	91	82	13.30%	<
35	100	105			103	94	100	91	97	88			102	93	101	92	98	90	3%	<
36	98	103			130	118	118	107	105	95			113	102	105	95	100	91	32%	8/6>


Ford, Tests 23-36,Front Tyre
Speed km/h
0
20
40
60
80
100
120
140
Test No.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average

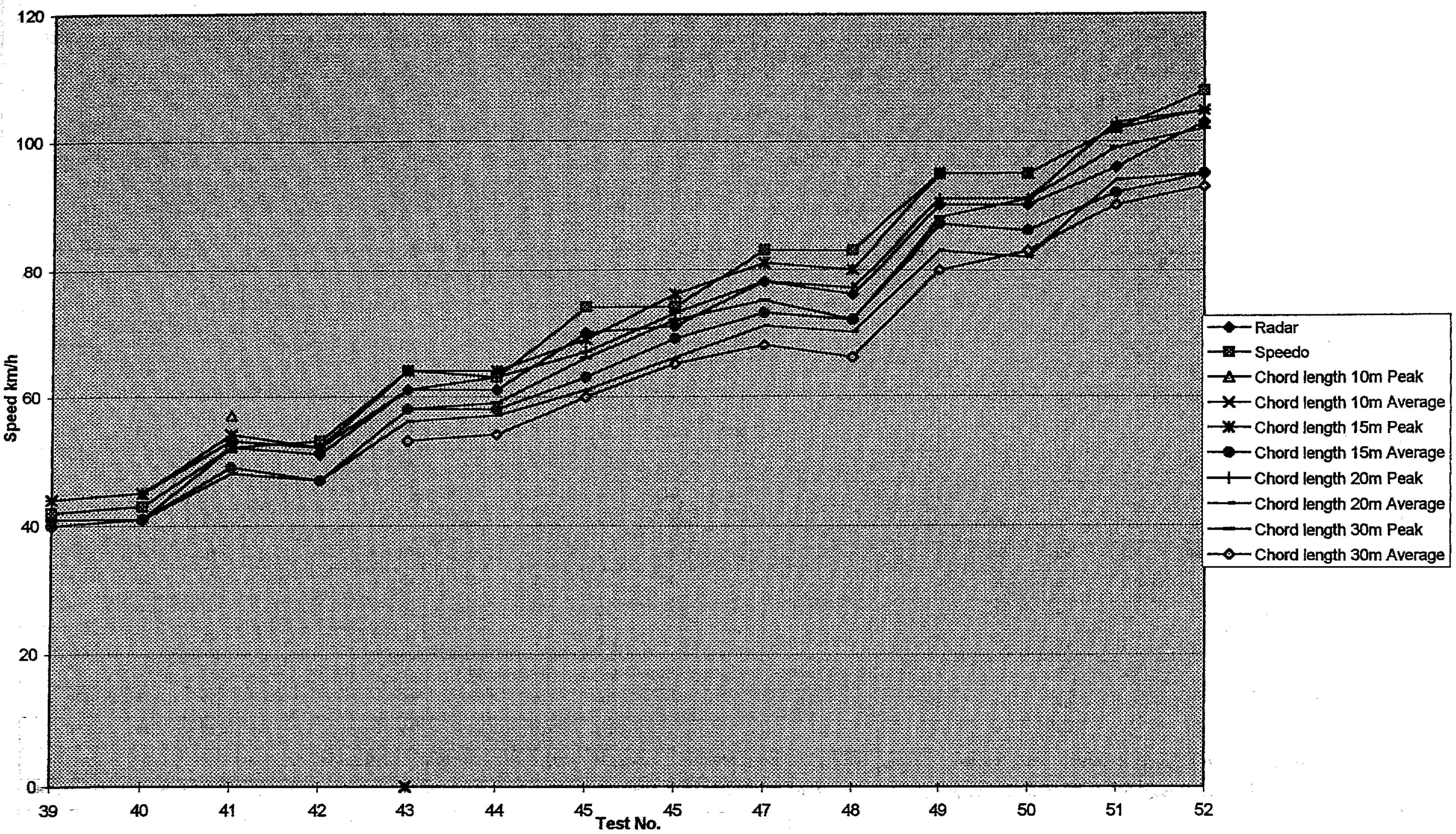
F2236F.XLS

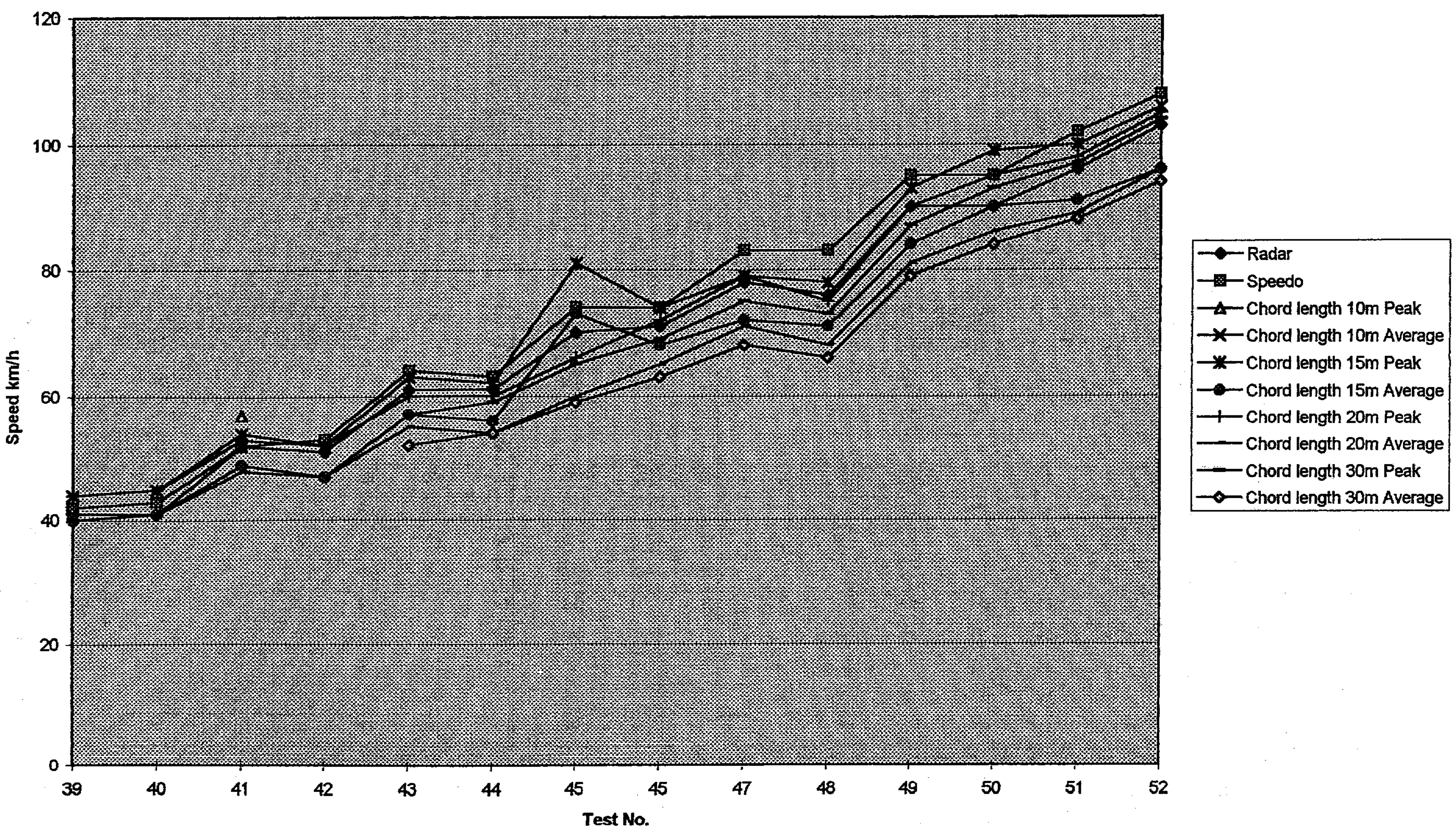

Ford, Tests 23-36, Rear Tyre
Speed km/h
0
20
40
60
80
100
120
23
24
25
26
27
28
29
30
31
32
33
34
35
36
Test No.
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average

F2236R.XLS

MITSUBISHI MAGNA V6 AUTO. TYRE PRESSURE AT 28PSI. GOODYEAR EAGLE GA TYRES P205/65R15.14 February 1996.																				
			Front Tyre								Rear Tyre									
Test No	Radar	Speedo	Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Comments	
			Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak %	Average
39	41	42			44	40	44	40					44	40	44	40			6.80%	<
40	41	43			45	41	45	41					45	41	45	41			9.70%	<
41	52	52	57	52	54	49	53	48			57	52	54	49	53	48			9.60%	<
42	51	53			52	47	52	47					52	47	52	47			1.90%	<
43	61	64			64	58	61	56	58	53			63	57	60	55	57	52	4.90%	<
44	61	63			64	58	63	57	59	54			62	56	60	54	*59	54	4.90%	<
45	70	74			69	63	67	61	66	60			81	73	66	60	65	59	15.70%	<
45	71	74			76	69	73	66	72	65			74	68	72	65	69	63	7%	<
47	78	83			81	73	78	71	75	68			79	72	79	71	75	68	3.80%	<
48	76	83			80	72	77	70	72	66			78	71	75	68	73	66	5.20%	<
49	90	95			95	87	91	83	88	80			93	84	90	81	87	79	5.50%	<
50	90	95			95	86	91	82	91	83			99	90	95	86	93	84	10%	<
51	96	102			102	92	103	94	99	90			100	91	98	89	97	88	7.30%	<
52	103	108			105	95	105	95	102	93			106	96	105	96	104	94	2.90%	<
																	*25m Chord			

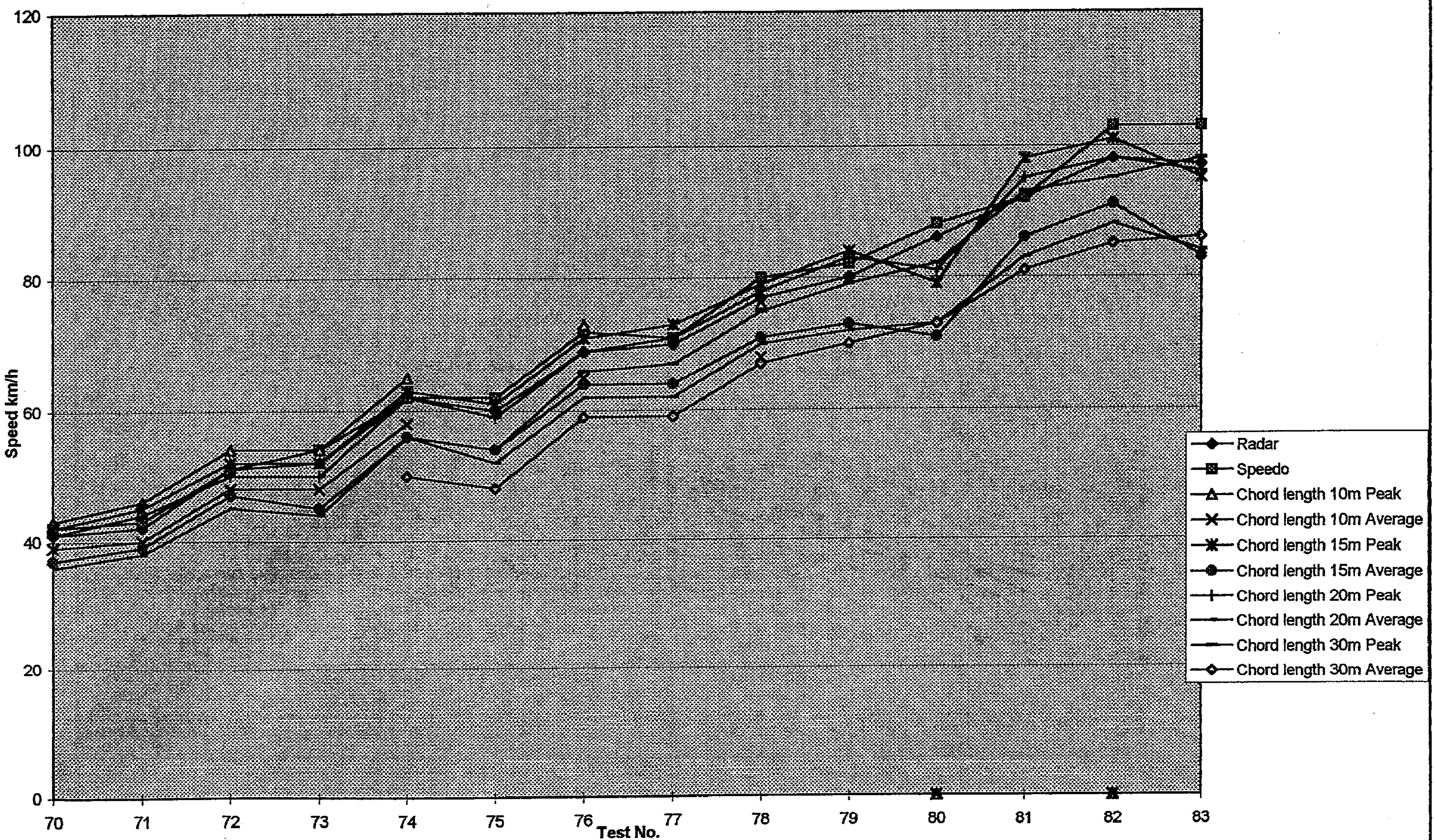
YAWMAGNA.XLS

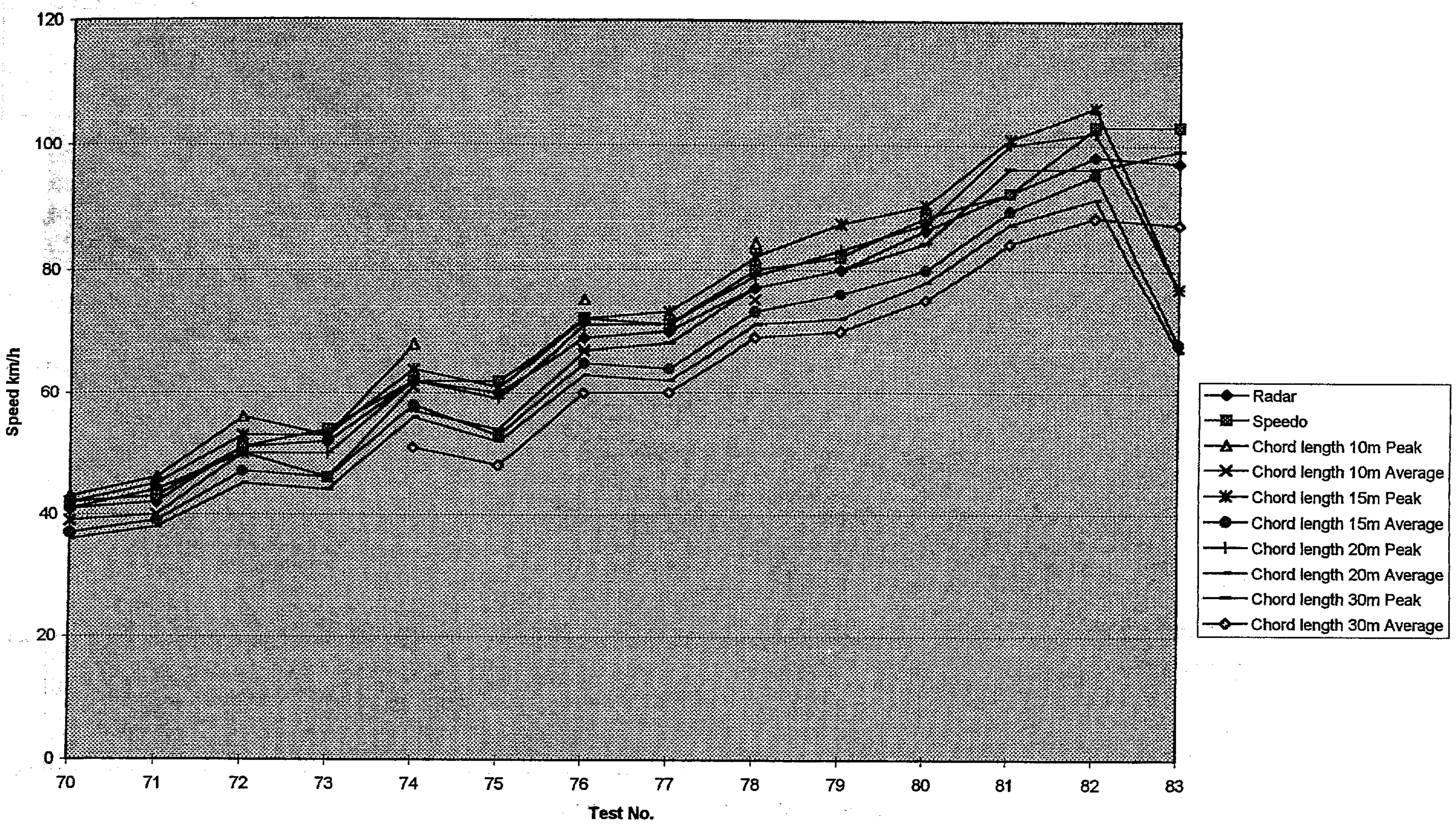

Magna, Tests 39-52, Front Tyre
Speed km/h
120
100
80
60
40
20
0
Test No.
39
40
41
42
43
44
45
45
47
48
49
50
51
52
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average


Magna,Tests 39-52,Rear Tyre
Speed km/h
120
100
80
60
40
20
0
39
40
41
42
43
44
45
45
47
48
49
50
51
52
Test No.
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average

TOYOTA COROLLA SECA. 4 CYLINDER. TYRE PRESSURE AT 28PSI. 15 February 1996																					
			Front Tyre								Rear Tyre										
Test No	Radar	Speedo	Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Comments	
			Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak %	Average
70	41	42	43	39	42	37	41	36			43	39	42	37	41	36			4.80%	<
71	42	43	46	40	45	39	44	38			46	40	45	39	44	38			9.50%	<
72	51	51	54	48	52	47	50	45			56	50	53	47	50	45			9.80%	<
73	52	54	54	48	52	45	50	44			53	46	53	46	50	44			3.50%	<
74	62	62	65	58	63	56	62	56	56	50	68	61	64	58	62	56	57	51	9.60%	<
75	60	62			61	54	59	52	54	48			61	53	59	52	54	48	1.60%	<
76	69	72	73	65	71	64	69	62	66	59	75	67	72	65	71	63	67	60	8.60%	<
77	70	71			73	64	71	62	67	59			73	64	71	62	68	60	4.20%	<
78	77	80	76	68	79	71	78	70	75	67	84	75	82	73	79	71	77	69	9.10%	<
79	80	82			84	73	83	72	79	70			87	76	83	72	80	70	8.70%	<
80	86	88			79	71	81	73	82	73			90	80	87	78	84	75	4.60%	<
81	92	92			98	86	95	83	93	81			101	89	100	87	96	84	9.80%	<
82	98	103			101	91	98	88	95	85			106	95	102	91	96	88	8.10%	<
83	97	103			95	83	96	84	98	86			77	68	77	67	99	87	1%	<

YAWSECA.XLS


Toyota Seca, Test 70-83, Front Tyre
Speed km/h
120
100
80
60
40
20
0
70
71
72
73
74
75
76
77
78
79
80
81
82
83
Test No.
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average



Toyota Seca, Test 70-83, Rear Tyre
Speed km/h
0
20
40
60
80
100
120
Test No.
70
71
72
73
74
75
76
77
78
79
80
81
82
83
Radar
Speedo
Chord length 10m Peak
Chord length 10m Average
Chord length 15m Peak
Chord length 15m Average
Chord length 20m Peak
Chord length 20m Average
Chord length 30m Peak
Chord length 30m Average

VR COMMODORE 34 PSI, STEER RIGHT. 14 February 1996																							
			Front Tyre								Rear Tyre												
Test N	Radar	Speedo	Chord length 10		Chord length 15		Chord length 20		Chord length 30		Chord length 10		Chord length 15		Chord length 20		Chord length 30		Comments				
			Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak %	Average			
53	61	65			61	55	59	53					59	53	58	52			<	<	constant throttle		
54	77	78			79	72	77	71	74	68			78	72	71	64	72	65	2.60%	<	"	"	
55	79	81			79	74	79	73	76	71			79	74	78	72	76	60	<	<	"	"	
56	91	93			94	86	91	83	90	82			92	84	82	84	88	80	3%	<	"	"	
57	101	106			103	94			100	91			100	91	98	90	98	89	2%	<	"	"	
58	69	74			71	64	71	63					72	64	69	62			4.30%	<	50% brake		
59	72	76			74	67	73	66					72	66	73	66			2.80%	<	50% brake		
60	70	75			66	59	65	58	63	56			66	59	65	58	64	57	<	<	50% brake		
61	74	76			75	68	71	65					74	68	72	66			1.40%	<	50% brake		
62	71	74			71	64	70	63					72	64	70	63			1.00%	<	lock up		
63	74	76			74	67	74	66	73	66			76	68	76	68	72	65	3%	<	accelerate		
64	75	78			79	72	76	69	74	67			76	69	75	68	72	65	5%	<	accelerate		
65.1	97			*	110	98							101	91					######	>1%	steer le	rear tyr	flicks 28psi
65.2	91			*	97	87	96	86	93	83			93	83	92	83	93	83	6.60%	<	steer right		"
66.1	102			*	141	129							109	100					######	>26%	steer le	rear tyr	"
66.2	91			*	105	96	104	95	101	92			97	89	99	91	99	90	######	>5.5%	steer ri	raer tyr	"
67.1	104			*	112	102	115	105					105	95	106	97			8.60%	>1%	steer le	rear tyr	"
67.2	98			*	103	94	100	91	99	90			101	92	100	91	98	90	5.10%	<	steer right		"
68.1	103			*	106	95	99	88					103	92	107	95			3.90%	<	steer left		"
68.2	96			*	108	96	105	94	96	86			96	86	87	87	96	86	12.5	<	steer ri	rear tyr	"
69.1	107			*	113	104	116	106					112	103	113	103			8.40%	<	steer left		"
69.2	101			*	104	95	103	94	103	94			101	92	101	92	101	93	3%	<	steer right		"
				*flick tests																			
				rear tyre ok																			
				front tyre ?																			

YAWCOMM2.XLS

ED Falcon tests. Tyre pressure at 28PSI. Flicks.																						
			Front Tyre								Rear Tyre											
Test N	Radar	Speedo	Chord length 10		Chord length 15		Chord length 20		Chord length 30		Chord length 10		Chord length 15		Chord length 20		Chord length 30		Comments			
			Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak %	Average		
84.1	73																					
84.2	69				76	69	74	68					72	65	71	64			#####	<=		
84.3	69						74	68	72	65					73	66	72	65	7.20%	<		
85.1	73	75																		<		
85.2	67				80	72	76	69					71	65	71	64			19%	>7.4%	Rear tyre<	
85.3	67	*115			75	68	74	67	73	66			74	67	74	67	73	66	#####	>7.4%	rear tyr	
ED Falcon tests. Tyre pressure at 26PSI. Steer right.																						
86	74	75			75	67	73	65	70	62			73	66	71	64	68	61	1.35%	<		
87	70	75			70	61	69	60	65	57			69	60	67	59	63	55	<=%	<		
88	83	84			82	73	78	70	78	69			82	74	80	72	77	69	<%	<		
89	86	84			85	75	86	76	85	74			89	78	86	75	83	73	3.50%	<		
90	93	93			95	84	93	82	93	81			91	80	91	80	90	79	2%	<		
91	100	100			101	90	97	87	96	86			98	87	96	86	96	86	1%	<		
ED Falcon tests. Tyre pressure at 28PSI. Flicks.																						
92.1	84	82	85	76	*96	86					82	73	74	66					#####	<2.4%	Rear ty	half left
92.2	73	67			72	65	75	67	73	65			76	68	74	66	70	63	2.70%	<		half rig
93.1	82	82	84	74	*97	85					88	77	88	77					#####	<3.6%	rear tyr	half left
93.2	73	73	82	72	79	69	78	69	75	66	80	70	80	70	78	68	7500%	66	#####	<		half rig
94.1	90	94	97	87	95	86					97	87	92	82					7.80%	<		half rig
94.2	78				83	74	83	74	85	76			82	73	83	74	84	75	9.00%	<		half left

*N.B. 1 drive wheel off ground

*front tyre becoming straighter changing from left steer to right steer

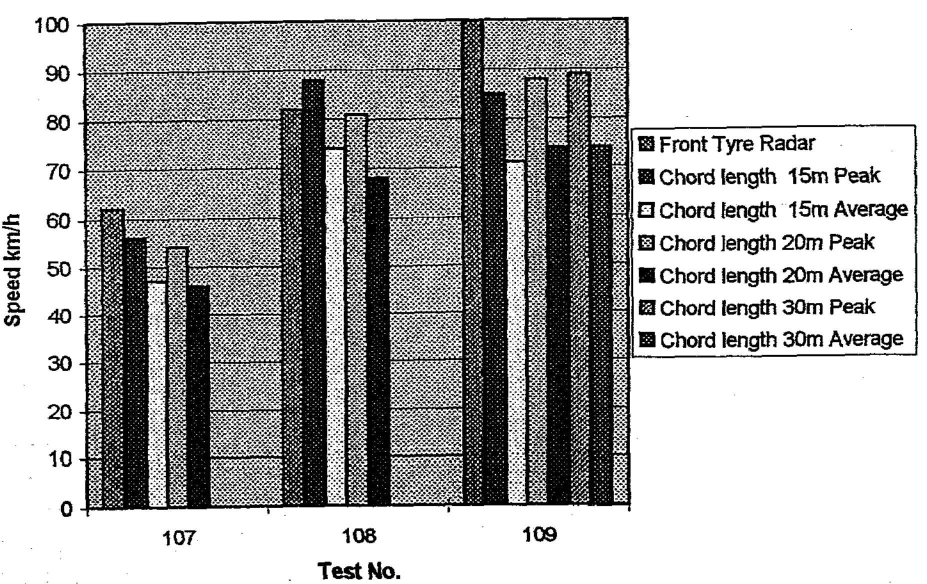
YAWFORD2.XLS

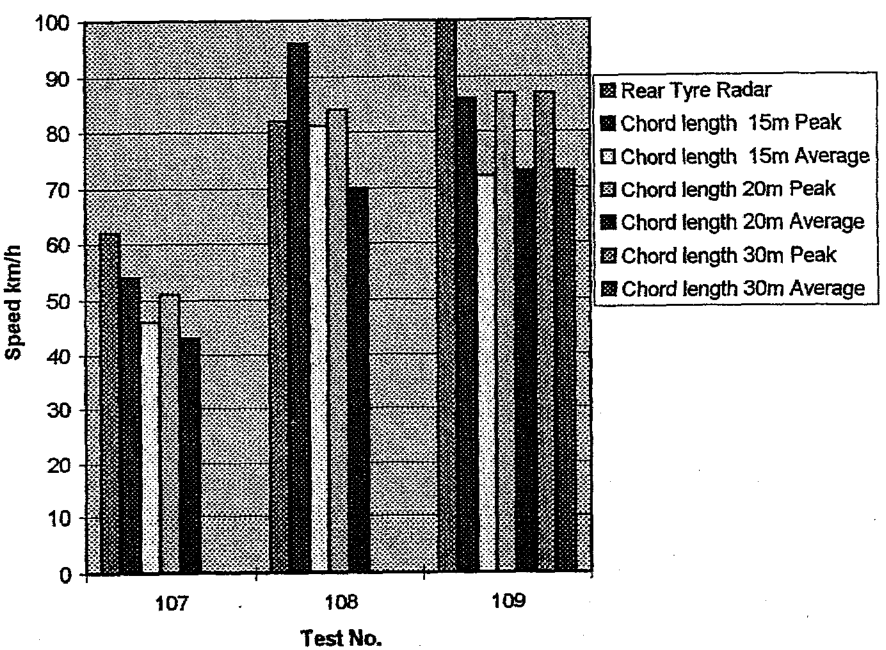
VR COMMODORE 28 PSI, CORNERING AT SPEED. 28 February 1996																					
			Front Tyre								Rear Tyre										
Test N	Radar	Speedo	Chord length 10		Chord length 15		Chord length 20		Chord length 30		Chord length 10		Chord length 15		Chord length 20		Chord length 30		Comments		
			Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak	Averag	Peak %	Average	
*93	110				110	100	109	99	107	97			103	94	111	101	107	98	1%	<	
94	125		NO FRONT TY												105	96	108	98	######	<23%	
95	75	76	NO FRONT TY										81	72	79	71	75	67	800%	<	CONSTANT STEER
96	81	84			91	81	88	79	86	77			96	86	90	81	88	79	19%	15M=>	ACC HALF
97	81						97	87	89	80					100	89	92	82	######	20M=>	HESITATION ACCELERATION
98	88	92					97	87	93	84					97	86	91	81	######	<	ACCEL
*CRITICAL SPEED OF CORNER 111KPH (PEAK)																					
			101KPH (AVERAGE)																		

ED Falcon tests. Tyre pressure at 28PSI. Cornering at speed 28 February 1996.																						
			Front Tyre								Rear Tyre											
Test No	Radar	Speedo	Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Chord length 10m		Chord length 15m		Chord length 20m		Chord length 30m		Comments			
			Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak	Average	Peak %	Average		
99	76	80					74	66	70	63					72	64	69	62	-5%	<	const. throttle	
100	80	80					91	81	85	76					91	81	85	76	13.80%		acc.	
101	89	92					104	93	96	86					102	91	96	86	16.80%		acc.	
102	88	90					97	86	89	80					95	85	88	79	10%	<	const. thr	
103	100						104	93	99	88					106	95	100	90	3%	<	const. thr	
104	97						99	88	97	87					101	90	97	87	4.10%	<	acc.	
105	97						87	78	86	77					91	82	90	81	-11.30%	<	emerg. b	
106	98						91	81	90	81					95	85	93	83	-7.10%	<	emerg. braking AB	

YAWFORD3.XLS

COMMODORE, SPLIT FRICTION TESTS, LEFT SIDE GRASS, RIGHT SIDE BITUMINOUS SURFACE (2/3 WEIGHT SHIFT TO GRASS SURFACE)							
Front Tyre		Chord length 15m		Chord length 20m		Chord length 30m	
Test No.	Radar	Peak	Average	Peak	Average	Peak	Average
107	62	56	47	54	46		
108	82	88	74	81	68		
109	100	85	71	88	74	89	74
Rear Tyre		Chord length 15m		Chord length 20m		Chord length 30m	
Test No	Radar	Peak	Average	Peak	Average	Peak	Average
107	62	54	46	51	43		
108	82	96	81	84	70		
109	100	86	72	87	73	87	73

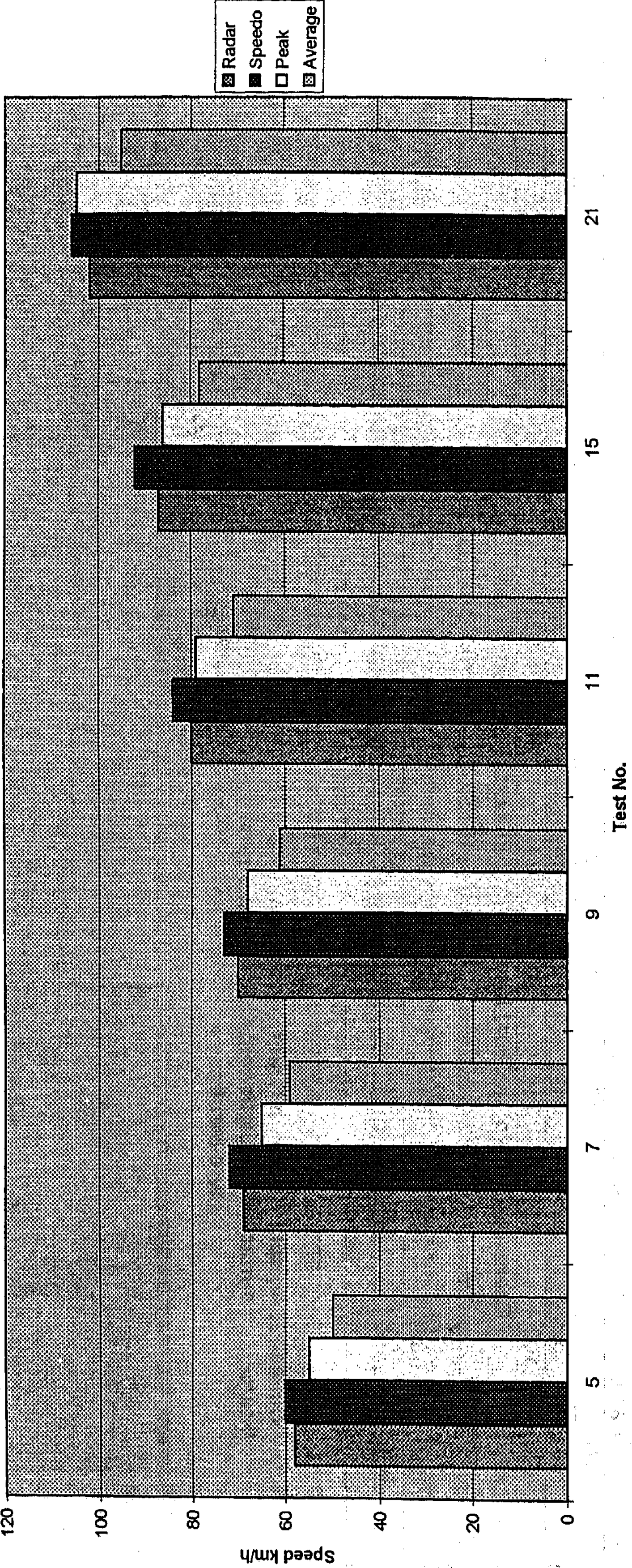




YAWSPLIT.XLS

APPENDIX 2
CENTRE OF MASS PLOT METHOD, SPEED RESULTS AND GRAPHS

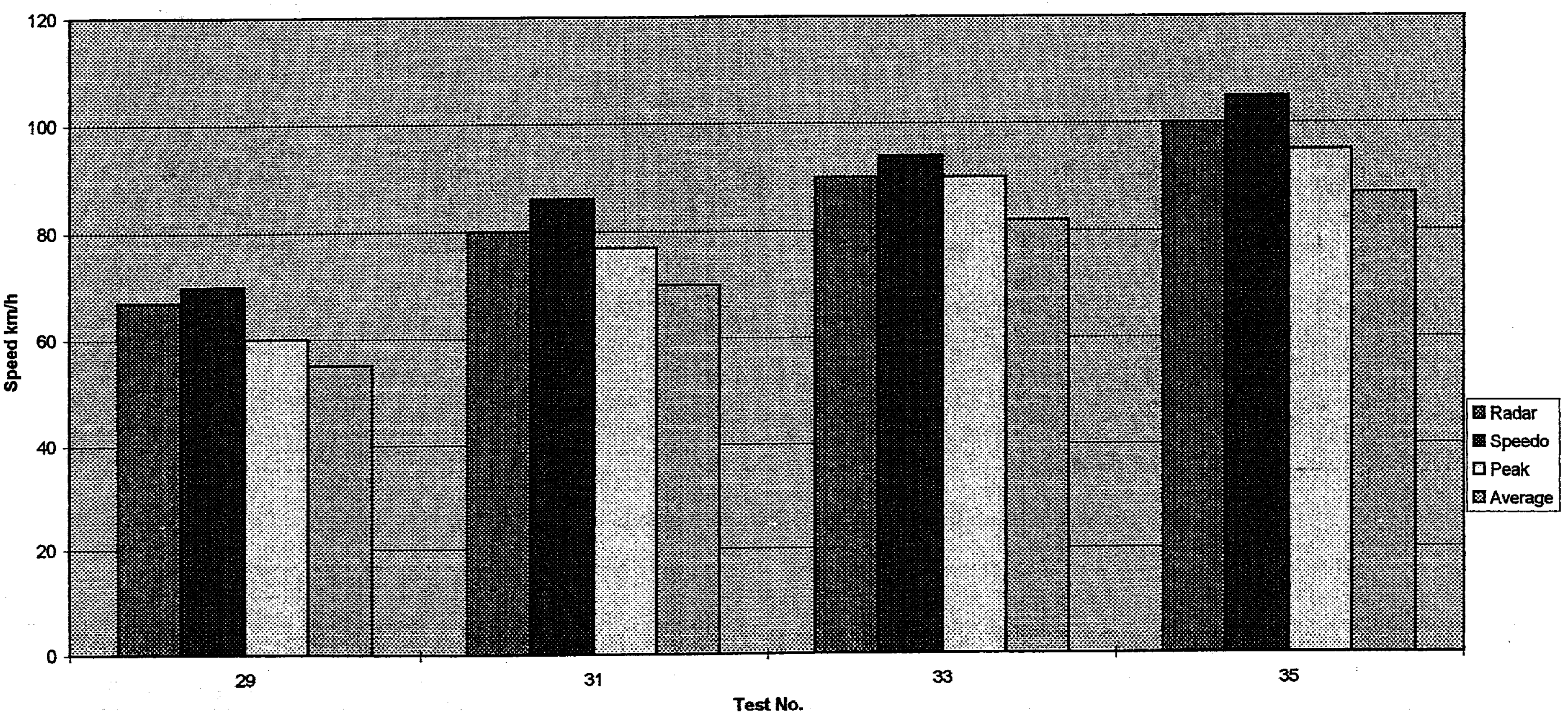
CENTRE OF MASS PLOTS, SPEED CALCULATION RESULTS.				
Commodore, straight then oversteer.				
Test No	Radar	Speedo	Peak	Average
5	58	60	55	50
7	69	72	65	59
9	70	73	68	61
11	80	84	79	71
15	87	92	86	78
21	102	106	105	95



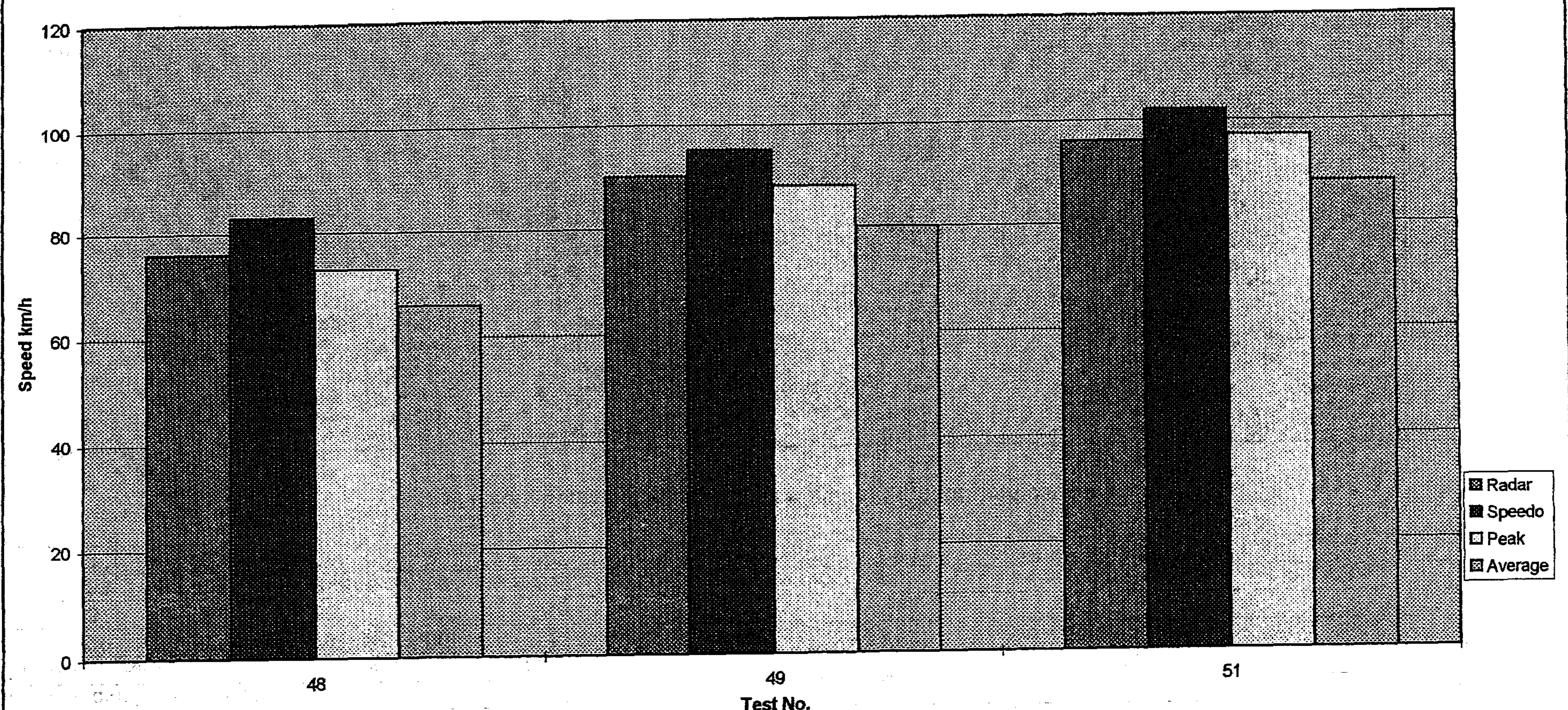
CENTRE OF MASS PLOTS, SPEED CALCULATION RESULTS

Ford, straight then oversteer.

Test No	Radar	Speedo	Peak	Average
29	67	70	60	55
31	80	86	77	70
33	90	94	90	82
35	100	105	95	87



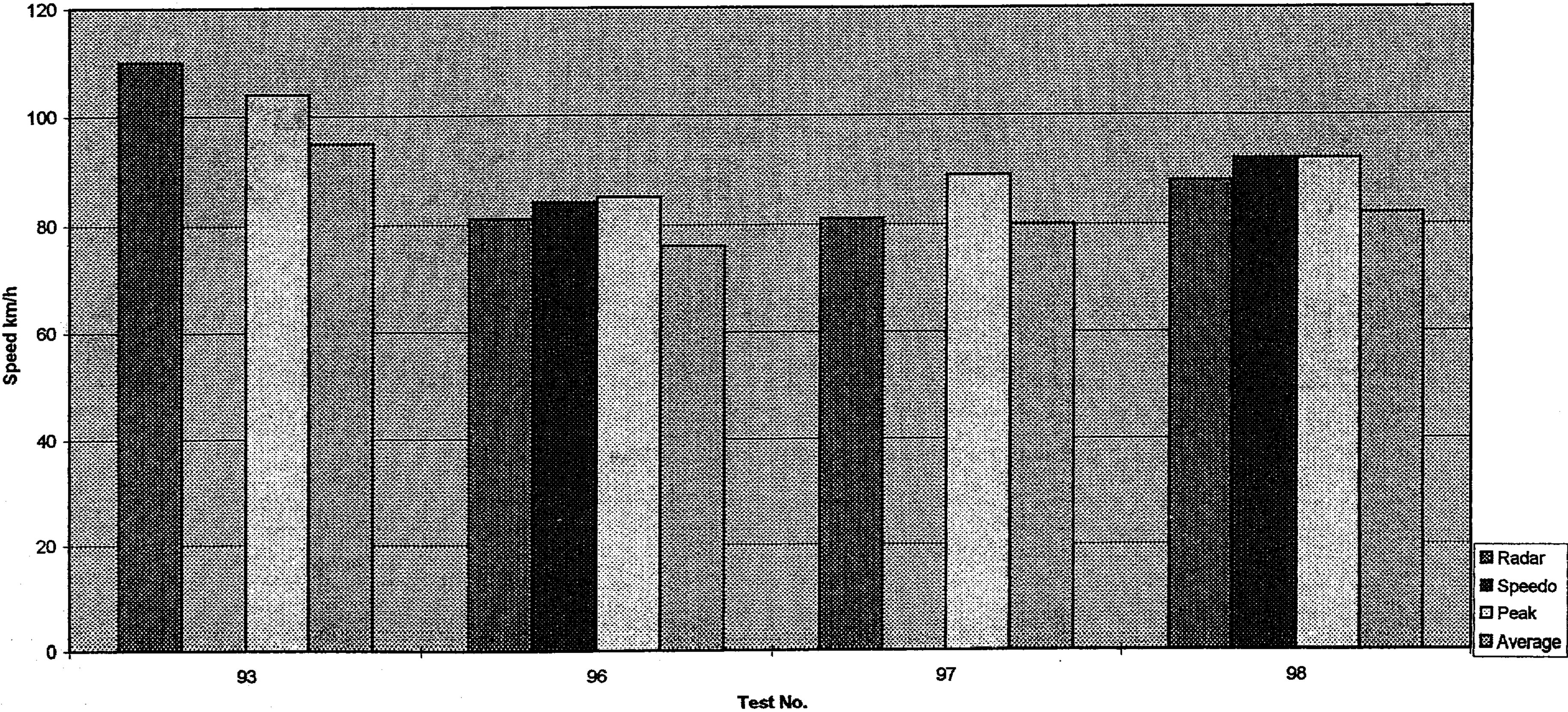
CENTRE OF MASS PLOTS, SPEED CALCULATION RESULTS				
Magna, straight then oversteer.				
Test No	Radar	Speedo	Peak	Average
48	76	83	73	66
49	90	95	88	80
51	96	102	97	88



CENTRE OF MASS PLOTS, SPEED CALCULATION RESULTS

Commodore, cornering at excessive speed.

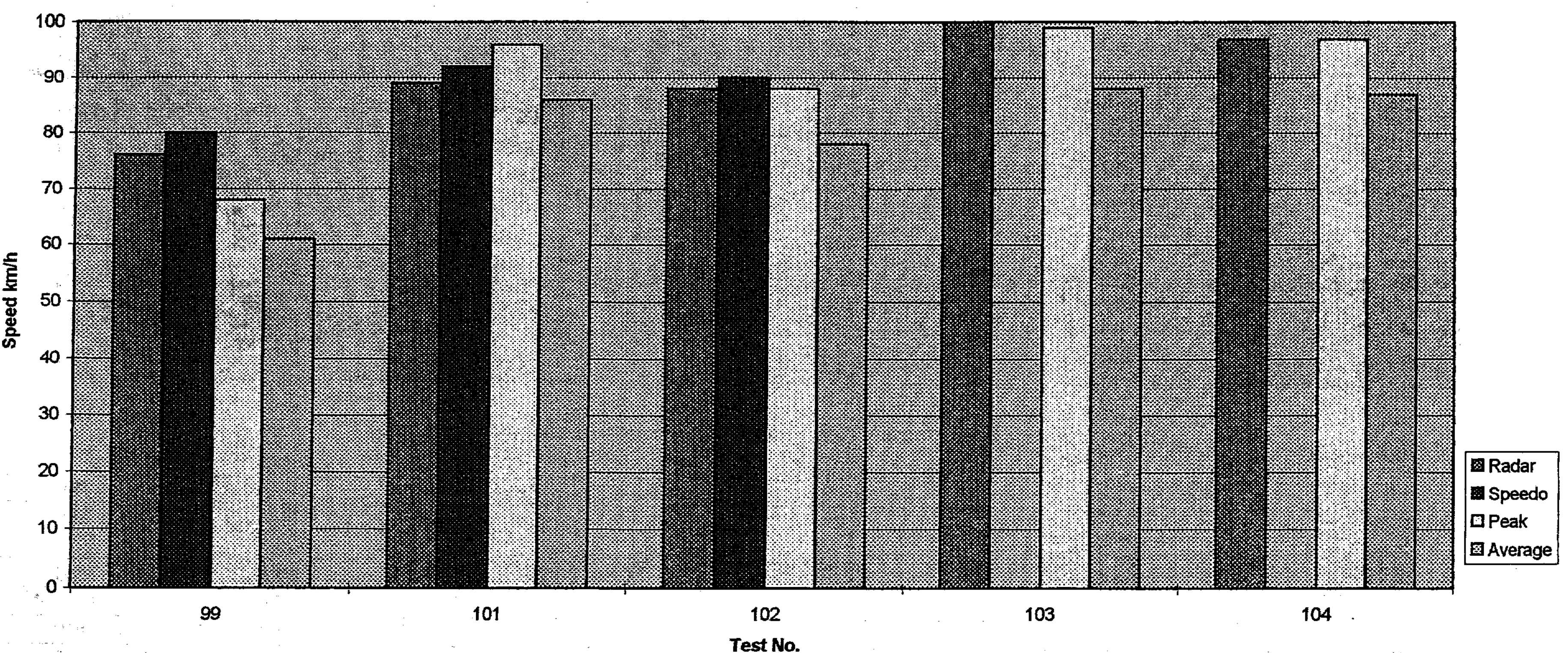
Test No	Radar	Speedo	Peak	Average
93	110		104	95
96	81	84	85	76
97	81		89	80
98	88	92	92	82



CENTRE OF MASS PLOTS, SPEED CALCULATION RESULTS

Falcon, cornering at excessive speed.

Test No	Radar	Speedo	Peak	Average
99	76	80	68	61
101	89	92	96	86
102	88	90	88	78
103	100		99	88
104	97		97	87



APPENDIX 3
GENERAL TEST PHOTOGRAPHS

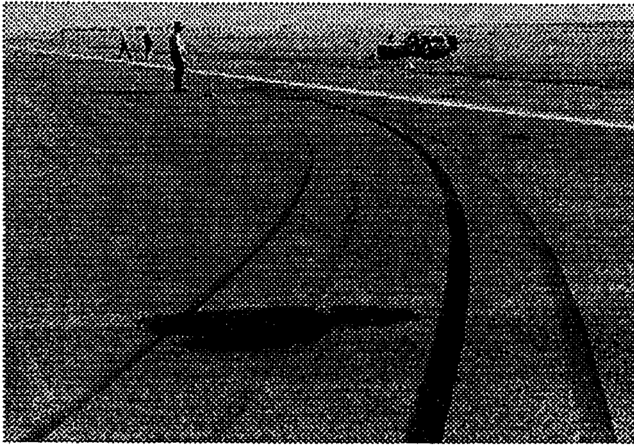

Photo 5, Yaw Marks from Commodore Oversteer to Right


Photo 8, Measuring Yaw Marks with the 'Geodimeter'

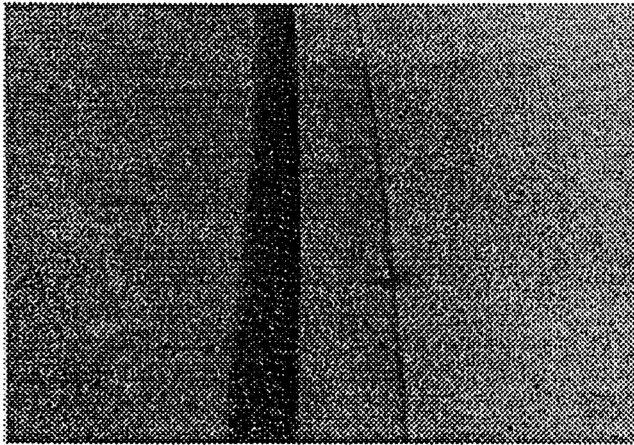

Photo 6, Deceleration Striations in a Yaw Mark


Photo 9, Yaw into Skid, Commodore, Test 62


Photo 7, Yaw Marks, Over Steering Left then Right


Photo 10, Test 62, Close Up of Yaw Mark becoming Skid Mark

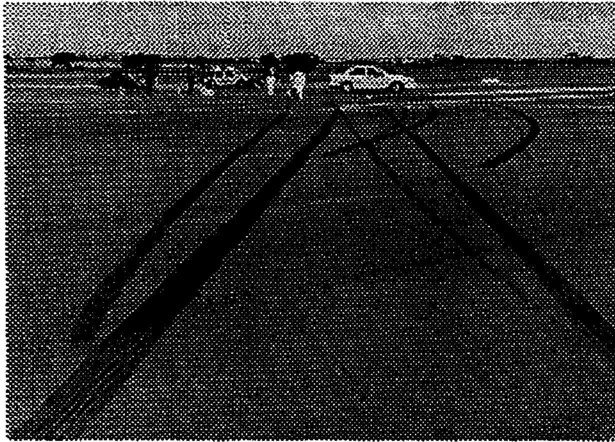

Photo 11, Comparison, Yawing then Braking Heavily, with & without ABS


Photo 12, Close Up of Striation Marks, Yawing & Braking, with & without ABS

SAE #970955
Reviewer's Discussion
by Lynn B. Fricke, Fricke Cooper Engineering
Project Y.A.M. (Yaw Analysis Methodology), Vehicle Testing and Finding of Victoria Police Accident Investigation Section
Peter Bellion, Author

The author has summarized 110 tests that were done by four vehicles to create yaw marks. These data provide an excellent source of information that others may want to use in further research on the subject of speed estimates from yaw marks. The report gives guidelines that appear to be quite sound when estimates of speed are made from yaw marks. The results obtained by the author are similar to other data that has been collected by other researchers.

SAE #970955
Reviewer's Discussion
by Charles P. Dickerson, Collision Engineering Associates, Inc.
Project Y.A.M. (Yaw Analysis Methodology), Vehicle Testing and Finding of Victoria Police Accident Investigation Section
Peter Bellion, Author

The author presents the results of a large body of testing performed to test the validity of the Yaw Analysis Methodology (YAM) which is also referred to as the Critical Speed Formula. The speed at the beginning of the maneuver, throttle and brake application, tire pressure, and maneuver direction are presented. The magnitude or phasing of the steering input is not reported. It would be valuable if the slip angle of the vehicle during the maneuver was documented and presented.

The test results show that the analysis is sensitive to the coefficient of friction value.

In the series, the runs that resulted in the most error in the estimated speed were those in which a "flick" steering input is used. This maneuver is described as being more severe than others and the author states that none of the analysis methods rendered "conservative results". The results show that the calculation method tends to predict a value higher than the actual speed.

The reviewer would suggest that the problem with the "Flick" runs, is that the resulting vehicle slip angle is higher for these maneuvers. Testing performed as Arndt and Associates, Ltd. shows that at higher slip angles (>30 to 35 degrees), error increases in the calculated speed. This is a result of the force component that produces the curved path being less than the total force acting on the vehicle. It is important to determine the vehicle slip angle before applying this method.

The paper is rich in test data to support its discussions and conclusions.

970956

Determination of Bumper Characteristics Using Prototype Moving Barriers

David J. King, James J. Bowler, and Stephen J. Ptucha
MacInnis Engineering Associates

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Accurate assessment of the severity of a low speed impact between two vehicles can sometimes only be accomplished through staged collisions with the actual or exemplar vehicles. However, the cost of obtaining, colliding, then repairing the vehicles often precludes this option.

For this paper, two prototype moving barriers were constructed to test three different bumper assemblies separate from their vehicles. Candidate bumper assemblies were mounted to the moving barriers for low speed impact testing with a stationary barrier and three other vehicles.

Forty three test series of 701 total impacts were done to compare bumper performance in moving barrier tests with their in-situ counterpart. Vehicle-to-fixed barrier, vehicle-to-vehicle, moving barrier-to-fixed barrier and moving barrier-to-vehicle tests were done using four different vehicles. The actual vehicle and moving barrier results were statistically compared.

The hypothesis that a moving barrier can be used in place of a motor vehicle in low speed impacts was tested.

INTRODUCTION

There has been increased research in assessing the severity of low speed collisions[1-7]. This paper focuses on the technique of reconstructing collision severity using a moving barrier in place of the vehicle.

The severity of a collision can be quantified in a number of ways[8]. In low-speed rear-end collisions, the velocity change (ΔV or delta-V) is the customary descriptor of severity. In this paper, velocity change and speed change are used interchangeably.

Without a doubt, the most accurate means of predicting collision severity is to recreate the collision using the actual or exemplar vehicles. This method will produce a virtually unassailable answer providing the tests were conducted properly.

When testing cannot be done, the accident reconstructionist must sometimes rely on empirical relationships such as ΔV versus average isolator compression, or energy change or absorption, and restitution versus isolator compression. The change in system energy for a collision is assumed to be the energy absorbed.

More theoretical, as opposed to empirical, methods have been pursued. Attempts to model the kinetics of bumper isolators[9] and other bumper constructions[10] have had little application in reconstruction. Some methods have shown promise, but their application requires not readily available knowledge of bumper properties. Acquiring those bumper properties requires the same testing that is required for application of the empirical methods outlined above.

The hypothesis of this study is that bumpers from our test vehicles attached to the prototype moving barriers will behave the same as if attached to the actual vehicles in low speed collisions.

MOVING BARRIER DESIGNS

MOVING BARRIER 1 (M1) - To achieve a wide variety of masses, mass centers, and mass moments of inertia, M1 was constructed mainly of rectangular aluminum tube stock with welded joints to be as light as possible. At a tare mass of 740 kg,

M1 is lighter than all but the lightest occupied vehicles. This allows in most cases for ballast to be added to duplicate the mass distribution and total mass for a specific vehicle. The center of mass position can be varied by altering the position of the ballast. The wheelbase was not made adjustable, so the center of mass to axle distance for both axles cannot be duplicated. It was determined that the suspension on the axle nearest the struck end would experience the bulk of the motion, so that having an adjustable wheelbase would not be of primary importance.

Load cells are attached to two vertical members at the front of the frame. The location of these load cells can be adjusted laterally and vertically by means of sliding tracks. Attached to the load cells is a hollow structural steel rectangular tube. To the front of this tube are attached laterally adjustable plates to which different bumper isolators can be attached. The portion of the isolator extending rearward of the isolator flange passes through slots in the steel tube.

Figure 1 shows the unladen M1. Table 1 lists the pertinent dimensions.

The M1 suspension consists of solid front and rear axles (both steerable) mounted on leaf springs attached to the barrier's main longitudinal members. The suspension stiffness can be adjusted by adding or subtracting leaves. The axles and spring assemblies were taken from the front end of two 1984 four wheel drive Suzuki SJ410's. Gas shocks were mounted vertically at the centers of all four leaf springs to provide suspension damping. The tires are P205/75/R15 on steel rims.

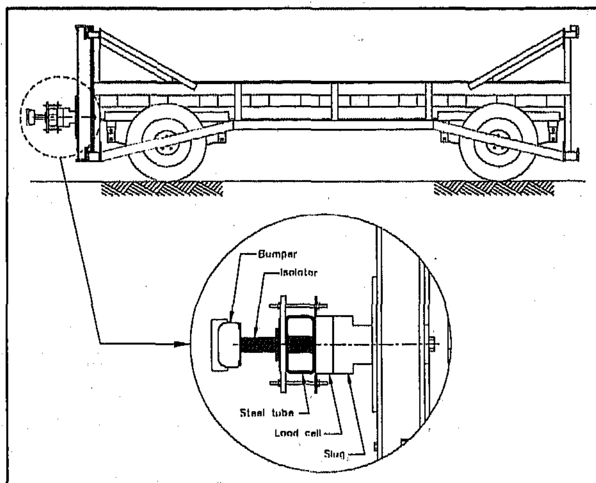


Figure 1. Unladen M1 and bumper detail.

Table 1. Moving barrier dimensions.

Empty mass	740 kg
Wheelbase	277 cm
Track width	122 cm
Bumper height	21 to 100 cm
Overall width	188 cm
Overall length	419 cm*
Overall height	125 cm

* without bumper

MOVING BARRIER 2 (M2) - The second moving barrier was made from a passenger car. A 1985 Chrysler LeBaron GTS four door was stripped of its engine (and associated components), transmission, fuel tank, rear hatch, rear doors, and interior structures. A plate was welded into the engine compartment for supporting ballast. This reduced the vehicle's mass from 1220 kg to 710 kg and likely did not significantly change the vehicle's front end stiffness for low speed purposes. The stock Chrysler suspension and wheels were not modified. The vehicle's front isolators were replaced by a pair of collapsed and welded isolators. These were used as mounts for the same load cells and rectangular steel tube structure used on M1. M2 is shown in Figure 2.

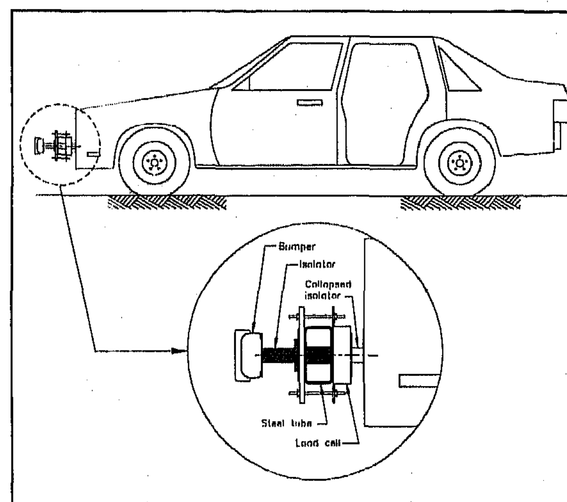


Figure 2. M2 and bumper detail.

TEST DESCRIPTION

EXPERIMENTAL METHOD - Forty three series of 701 total collisions were staged. Measured quantities were the vehicle mass, bumper isolator compression, pre-impact and post-impact speeds, and bumper force. Calculated quantities were speed change (ΔV), energy change (E), and coefficient of restitution (e).

The experimental procedure used for each test is described in detail elsewhere[2]. Briefly, the tests were staged on dry, level concrete and each vehicle was in neutral with the parking brake disengaged. A fine coat of Double*Glo Sno Spray™ was applied to each bumper isolator prior to each test in order to

distinguish isolator compression between tests. This procedure is necessary after the first few collisions since the paint on the isolator piston tubes, which leaves a good indication of compression distance, is scraped away. Tire air pressure was maintained according to the manufacturer's data placard except for M1. Its tire pressure was reduced to 20 psi from 32 psi to bring the rolling resistance in line with the other test vehicles.

VEHICLES - The four test vehicles are described in Table 2. The vehicles were weighed on a certified full-vehicle scale. Isolator data for each of the so-equipped vehicles is given in Table 2b. The approximate suspension stiffness in Table 2c was measured by placing known masses directly over the axle nearest to the impact surface and measuring the suspension deflection.

INSTRUMENTATION - Vehicle speed and force were measured in the x-direction. Speed data were acquired with an MEA 5th Wheel at 128 or 256 Hz for about 1 second before and 3 seconds after impact. Speed resolution in this sampling frequency range is about 0.01 m/s. Maximum isolator compression was measured to the nearest millimeter after each test.

Two Sensotec model 41 load cells were used to measure the impulse for most collisions. Each load cell's capacity was 44 480 N (10,000 lb.) and the total acquisition system resolution was ±30 N.

DATA INTERPRETATION - Figure 3a shows a sample of the MEA 5th Wheel and load cell output for one collision. Note that the target vehicle (Chevette) is stationary before impact. After impact, damped oscillation of both vehicles on their suspensions is seen in the traces. Long post-impact sampling durations were necessary to allow the oscillations of some vehicles to settle. For the collision shown in Figure 3a, the post-impact speeds were determined to be 0.24 m/s for the bullet vehicle (LeBaron) and 0.68 m/s for the target vehicle. The bullet vehicle's impact speed was 0.79 m/s.

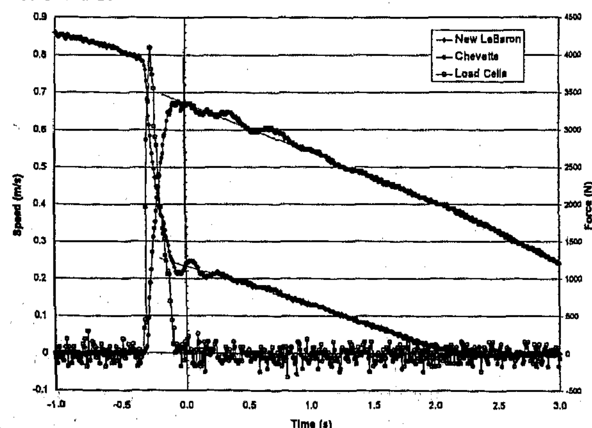


Figure 3a. Sample MEA 5th Wheel and force output.

Figure 3b shows a time-expanded view of the impact. The peak force was 4100 N and the collision duration was 0.25 s.

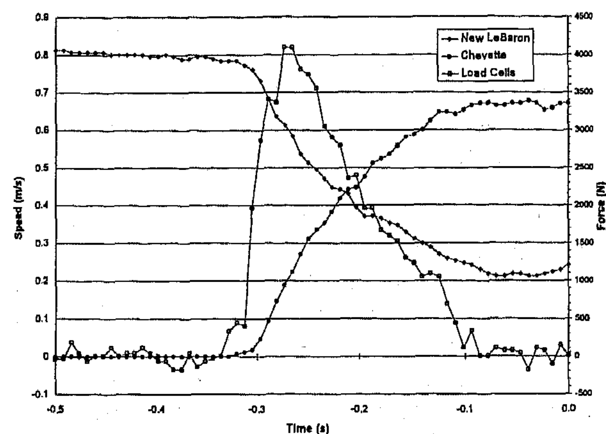


Figure 3b. Time-expanded view of collision interval.

ISOLATOR FAILURE - The performance of the LeBaron isolators became erratic after 330 tests (and a maximum speed change of 2.86 m/s). The average isolator stroke was 6 mm greater for a given speed change indicating an internal failure. The overall length of these isolators did not change. The relationship between isolator compression and speed change noted before the "failure" was observed with replacement isolators.

Static isolator compression tests results indicate that the initial compression force for the failed isolators was about 600 N lower than the replacement isolators. In addition, the force at full compression was 990 to 2311 N lower for the failed isolators. Sectioning of the failed isolators revealed that the metering pin was slightly bent to the outside of the annulus. General isolator descriptions are given in Reference 1. Both isolators exhibited the same behavior even though the static force at maximum compression for the left isolator was 1321 N greater than the right isolator. This is the only instance of isolator failure from repetitive testing that we have encountered in over 4000 tests.

BUMPER MOUNTING CONFIGURATION - It is important to mount candidate vehicle bumpers to the moving barrier in a manner as similar as possible to the actual vehicle. This requires that any significant load bearing points on the bumper that could transmit a collision force to the vehicle frame also have to be attached to the moving barrier face.

The front bumper from the Chrysler LeBaron is mounted on two hydraulic isolators and has a wrap-around cover. The ends of the cover are parallel to the vehicle's long axis and attach to the leading edge of the fenders. Insertion of load cells was necessary for force measurement in subsequent vehicle-to-vehicle tests. Figures 4 and 5 compare the results of front-to-barrier tests between the stock

Table 2 (a). Test vehicle data.

Year	Manufacturer	Model	VIN	Surface	Bumper Type
78	Chevrolet	Chevette, 4 door	1B68E8Y200647	Rear	isolator
89	Chevrolet	Sprint, 2-door	JG1MT2166KK772042	Rear	foam core
85	Mercury	Topaz, 4-door	2MEBP75R2FB644464	Rear	isolator
85	Chrysler	LeBaron GTS 4-door	1C3BH48D2EN167119	Front	isolator

Table 2 (b). Test Vehicle Isolator data.

Model	EPTL* (mm)	Max. Observed Compression Dynamic (mm)	Max. Compression Static (mm)	Force to Initiate Compression (N)	Force at full Compression (N)	Numbers on Isolator (Left Isolator First)
Chevette	52	47	49 L 48 R	1982 L 1982 R	6605 L 6605 R	3407771 / 22008563 3407771 / 22008563
Topaz	59	53	53 L 51 R	2642 L 2642 R	9247 L 9247 R	E53B 17D864-AA E53B 17D788-AA
LeBaron 1st set of Isolators	76	55	73	1321 FR 1321 FL	9908 FR 8587 FL	4270749 /// 3388441 4270749 /// 3388441
LeBaron 2nd set of Isolators	76	50	73	1982 NR 1982 NL	10898 NR 10898 NL	4270749 /// 2868871 4270749 /// 1098971

* (EPTL= Exposed Piston Tube Length; / = new line of numbers; /// = remainder of number located elsewhere on isolator; L,R = Left and Right; F = failed isolator; N = new isolator)

Table 2 (c). Test vehicle dimension data and approximate suspension stiffness.

Model	Wheel base (cm)	Mass (kg)	Suspension Stiffness (N/m)
M1	277	740	168500
M2	255	710	67000
Chevette	243	970	76000
Sprint	227	760	55000
Topaz	254	1150	79000
LeBaron	255	1220	67000

LeBaron front bumper and after load cell modification. Insertion of the load cells created a 4 cm gap between the wrap-around bumper cover and the fender leading edges. The gaps were spanned with aluminum channels to support the ends of the cover as when installed on the vehicle.

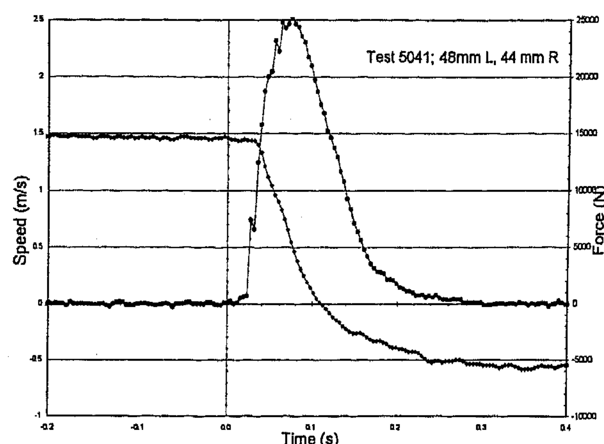


Figure 4(a). Force measured for entire bumper.

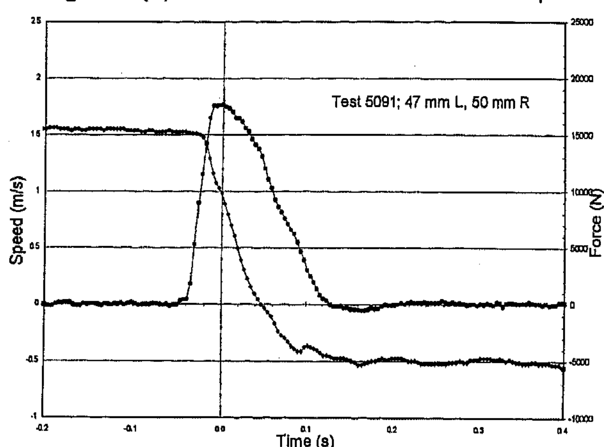


Figure 4(b). Force measured at isolators.

In Figure 4(a) the impact speed was 1.49 m/s and the peak load measured by the load cells, which is the load applied to the face of the bumper, was 25 kN. In Figure 4(b), the impact speed was 1.5 m/s. The load cells, between the isolators and the bumper, measured a peak force of only 17.6 kN. No other parameters were changed except that the load cells were inserted between the bumper bar and the isolators. With the load cells mounted to the bumper, they were not measuring the load applied through the wrap-around bumper cover to the fenders. Since the isolator compression and speed change were similar, the load cells did not alter the bumper performance. Comparison of these two tests shows that the wrap-around part of the bumper cover carried a peak load of approximately 7.4 kN, or about 30% of the collision force. Figures 5(a) and (b) show the same trend for a 0.6 m/s impact speed.

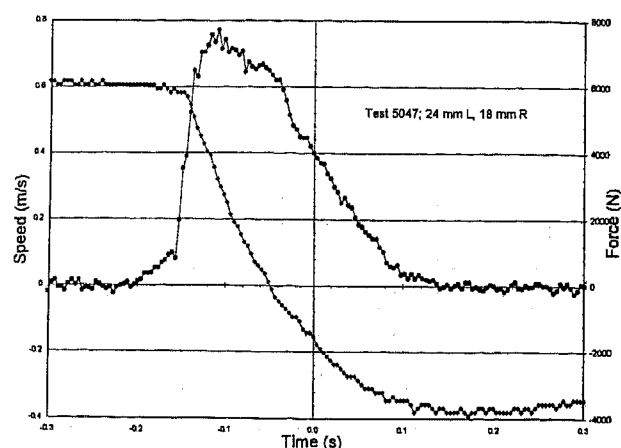


Figure 5(a). Force measured for entire bumper.

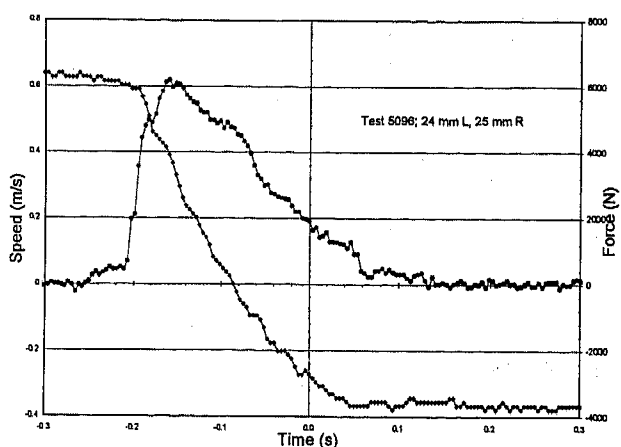


Figure 5(b). Force measured at isolators.

These tests highlight the concept that isolator compression is empirically correlated to speed change. While the isolators are not carrying all of the force, their compression can still be used to gauge impact severity. In addition, these tests demonstrate the importance of duplicating the bumper mounts to achieve comparable results. Had the spacers between the bumper wrap-around and fenders not been present, then more force would have been carried by the isolators and the correlation between isolator compression and severity would have been altered, giving more isolator compression for a given impact speed. The force comparison is shown in Figure 6.

RESULTS

TEST SUMMARY - The test data are shown in Figures 7 to 13, organized into the groups shown in Table 4.

STATISTICAL EVALUATION - The hypothesis of this study was tested statistically from the collected data.

For the test series where the actual test vehicles were used and not the moving barriers, a trend line was plotted through the speed change (ΔV), energy absorption (E), and coefficient of restitution (e) versus isolator compression

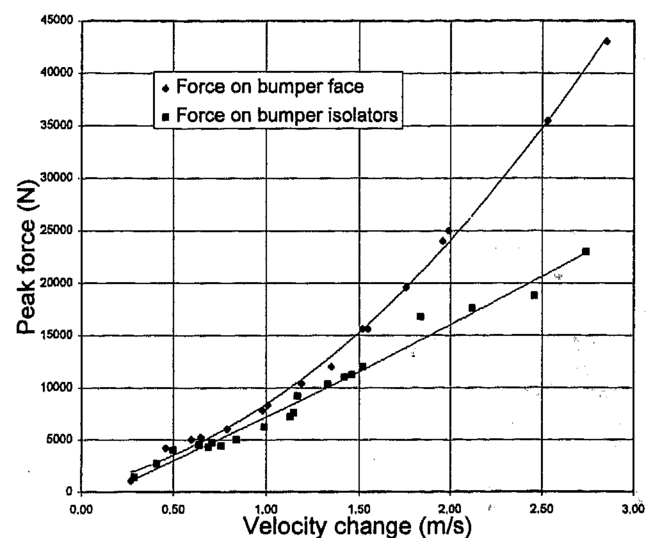


Figure 6. Comparison of force on the bumper face with the force on the isolators.

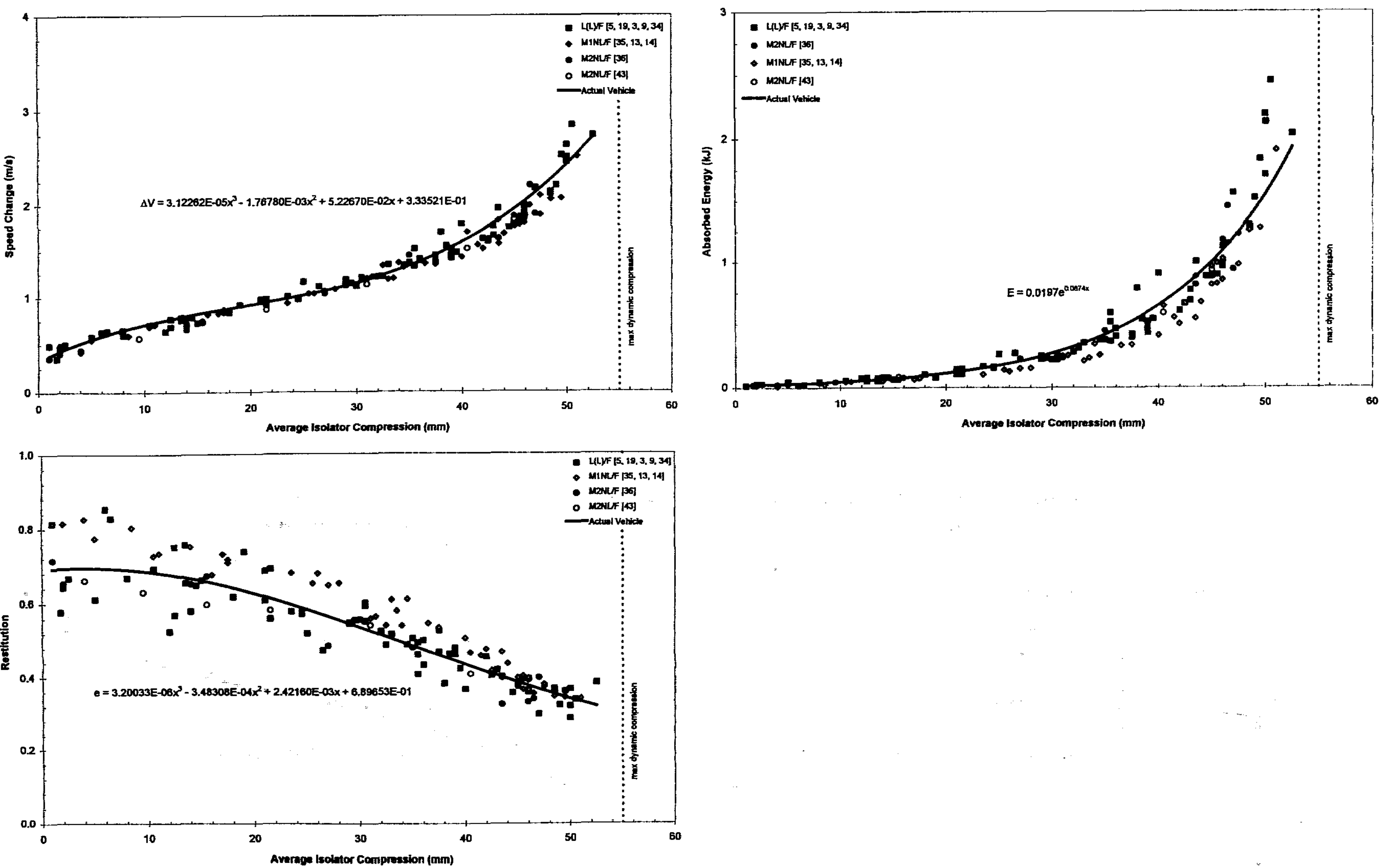
Table 4. Test Summary.

Group	Series	Vehicles	Symbol	# Tests	Comments
A	2	Chevette / Fixed Barrier with Load Cells	C/F(L)	25	
	20	M1 with Chevette / Fixed Barrier	M1C/F	26	
	41	M2 with Chevette / Fixed Barrier	M2C/F	12	
B	3	LeBaron / Fixed Barrier with load cells	L/F(L)	16	
	5	LeBaron with load cells / Fixed Barrier	L(L)/F	21	
	9	M1 with LeBaron / Fixed Barrier	M1L/F	20	
	10	M1 with LeBaron / Fixed Barrier	M1L/F	23	Mass distribution on M1 varied
	11	M1 with LeBaron / Fixed Barrier	M1L/F	15	Mass distribution on M1 varied; Springs replaced
	12	M1 with LeBaron / Fixed Barrier	M1L/F	14	M1 tire pressure reduced
	13	M1 with LeBaron / Fixed Barrier	M1L/F	13	2 of 4 front leaves and front shocks removed from M1
	14	M1 with LeBaron / Fixed Barrier	M1L/F	30	M1 bumper height adjusted to lower impulse height
	19	LeBaron with load cells / Fixed Barrier	L(L)/F	12	Isolator function tested
	27	LeBaron with load cells / Fixed Barrier	L(L)/F	9	LeBaron isolators tested for function - both failed
	34	New LeBaron with load cells / Fixed Barrier	NL(L)/F	14	
	35	M1 with New LeBaron / Fixed Barrier	M1NL/F	8	
	36	M2 with New LeBaron / Fixed Barrier	M2NL/F	11	
	43	M2 with New LeBaron / Fixed Barrier	M2NL/F	6	No beam, no load cells
C	4	Topaz / Fixed Barrier with load cells	T/F(L)	31	
	31	M1 with Topaz / Fixed Barrier	M1T/F	25	
	39	M2 with Topaz / Fixed Barrier	M2T/F	11	
	42	M2 with Topaz / Fixed Barrier	M2T/F	13	No beam, no load cells
D	6	LeBaron with load cells / Chevette	L(L)/C	15	
	16	M1 with LeBaron / Chevette	M1L/C	18	
	18	M1 with LeBaron / Chevette	M1L/C	7	Continuation of series 16
	22	LeBaron with load cells / Chevette	L(L)C	9	
	23	M1 with LeBaron / Chevette	M1L/C	19	
	24	M1 with LeBaron / Chevette	M1L/C	13	Rubber spacers added to M1 bumper mounts
	25	M1 with LeBaron / Chevette	M1L/C	3	Rubber spacers added to M1 isolator mounts
	21	M1 with Chevette / LeBaron	M1C/L	16	
	26	LeBaron with load cells / Chevette	L(L)/C	16	LeBaron isolators failed at test 443
	28	LeBaron with load cells / Chevette	L(L)/C	7	
	33	New LeBaron with load cells / Chevette	NL(L)/C	13	
	37	M2 with New LeBaron / Chevette	M2NL/C	20	
E	1	Sprint / Fixed Barrier with load cells	S/F(L)	15	
	7	LeBaron with load cells / Sprint	L(L)/S	11	
	15	M1 with LeBaron / Sprint	M1L/S	30	
	38	M2 with New LeBaron / Sprint	M2NL/S	11	
F	8	LeBaron with load cells / Topaz	L(L)/T	13	
	17	M1 with LeBaron / Topaz	M1L/T	23	
	44	M2 with New LeBaron / Topaz	M2NL/T	20	
G	29	Topaz / Chevette	T/C	23	
	30	M1 with Topaz / Chevette	M1T/C	6	
	32	M1 with Topaz / Chevette	M1T/C	25	Continuation of series 30
	40	M2 with Topaz / Chevette	M2T/C	13	

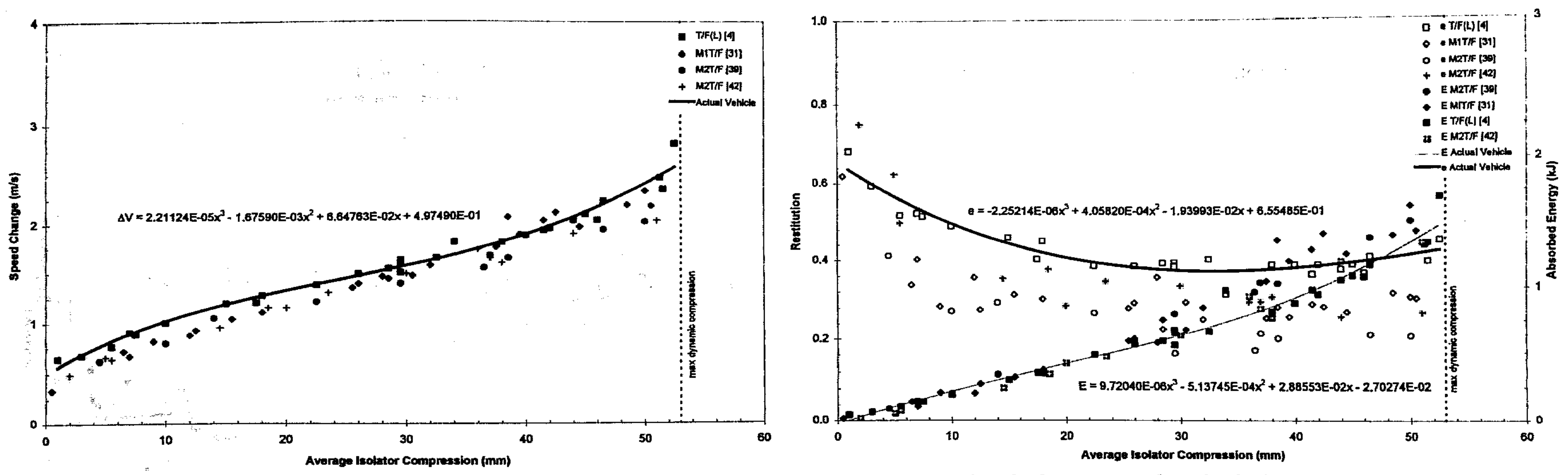
Striking Vehicle / Struck Vehicle

Figures 7a - 7b. Group A- Chevette / Fixed Barrier (speed change, absorbed energy, and restitution)

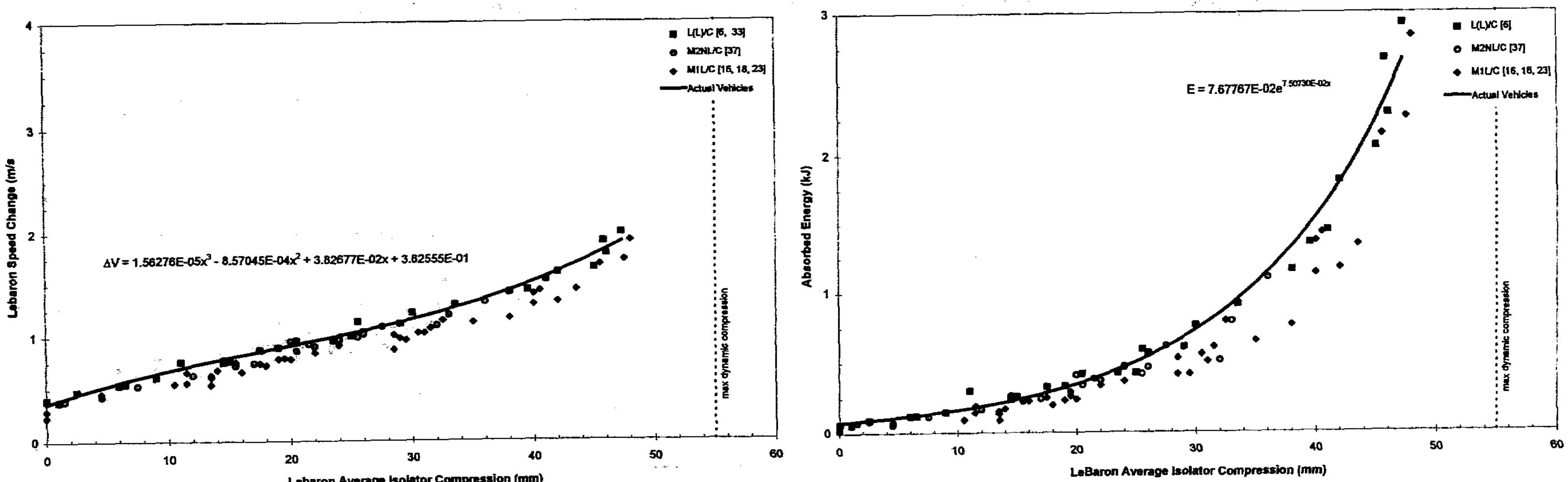


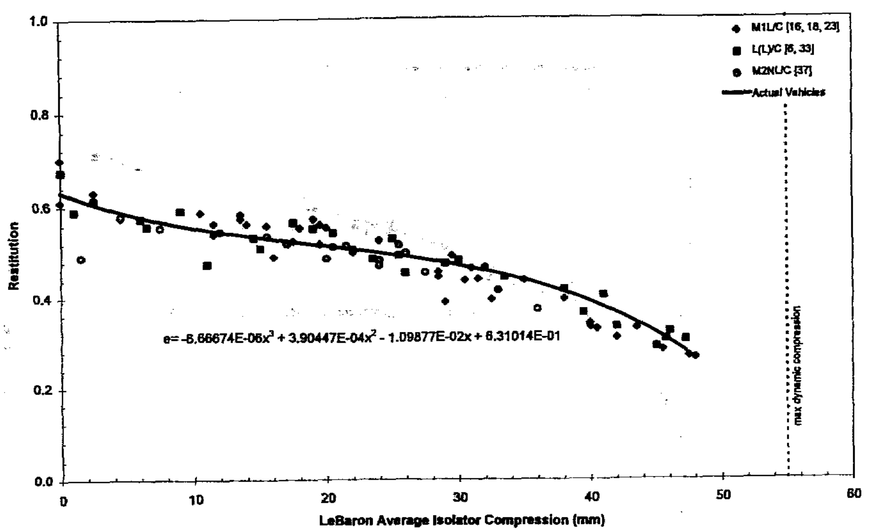


Figures 8a - 8c Group B - LeBaron / Fixed Barrier (speed change, absorbed energy, and restitution)

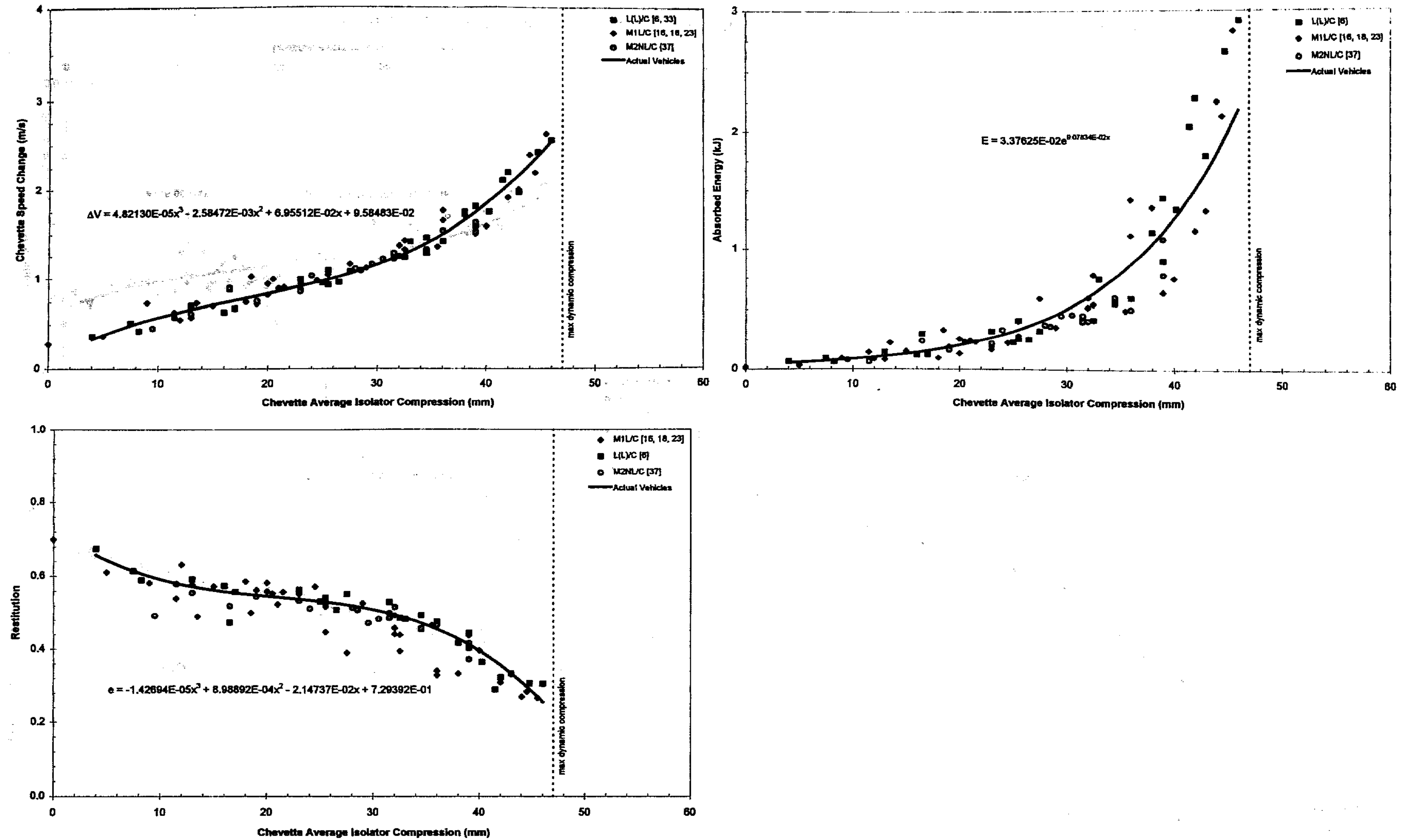


Figures 9a - 9b Group C - Topaz / Fixed Barrier (speed change, absorbed energy, and restitution)

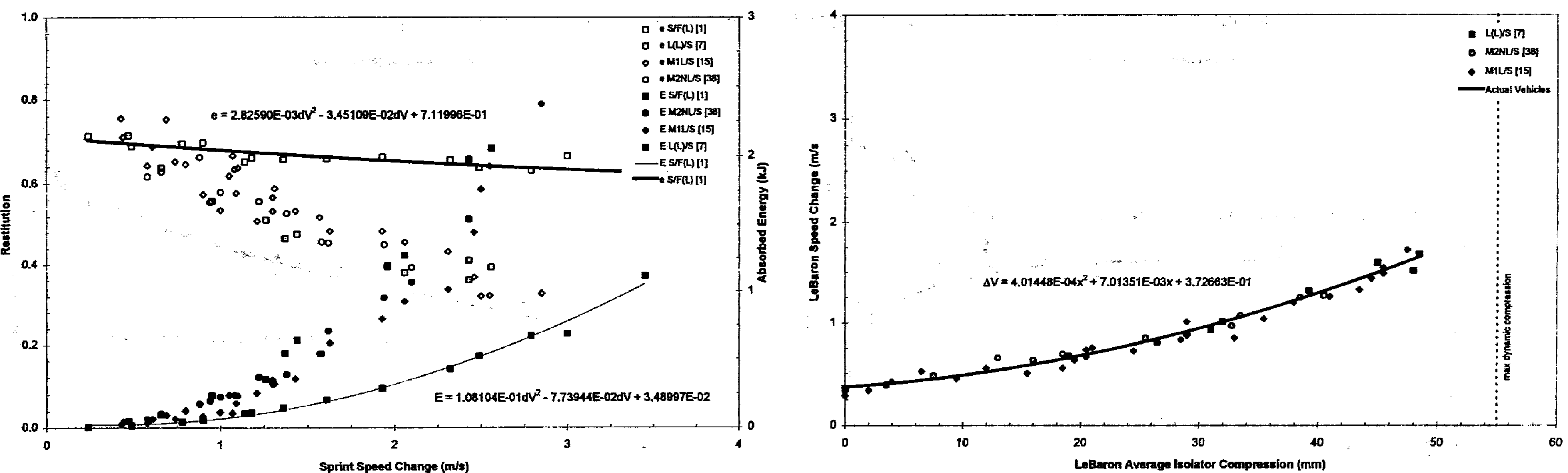


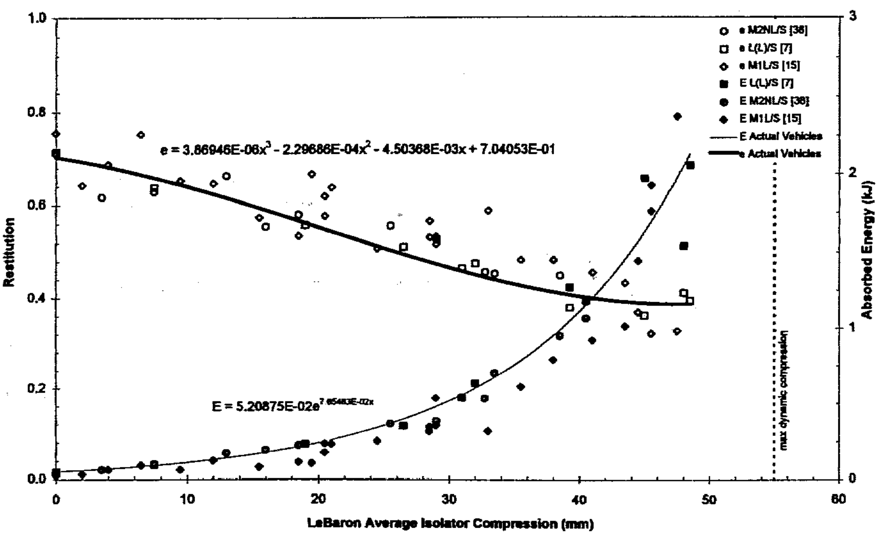


Figures 10a - 10c Group D - LeBaron / Chevette (speed change, absorbed energy, and restitution)

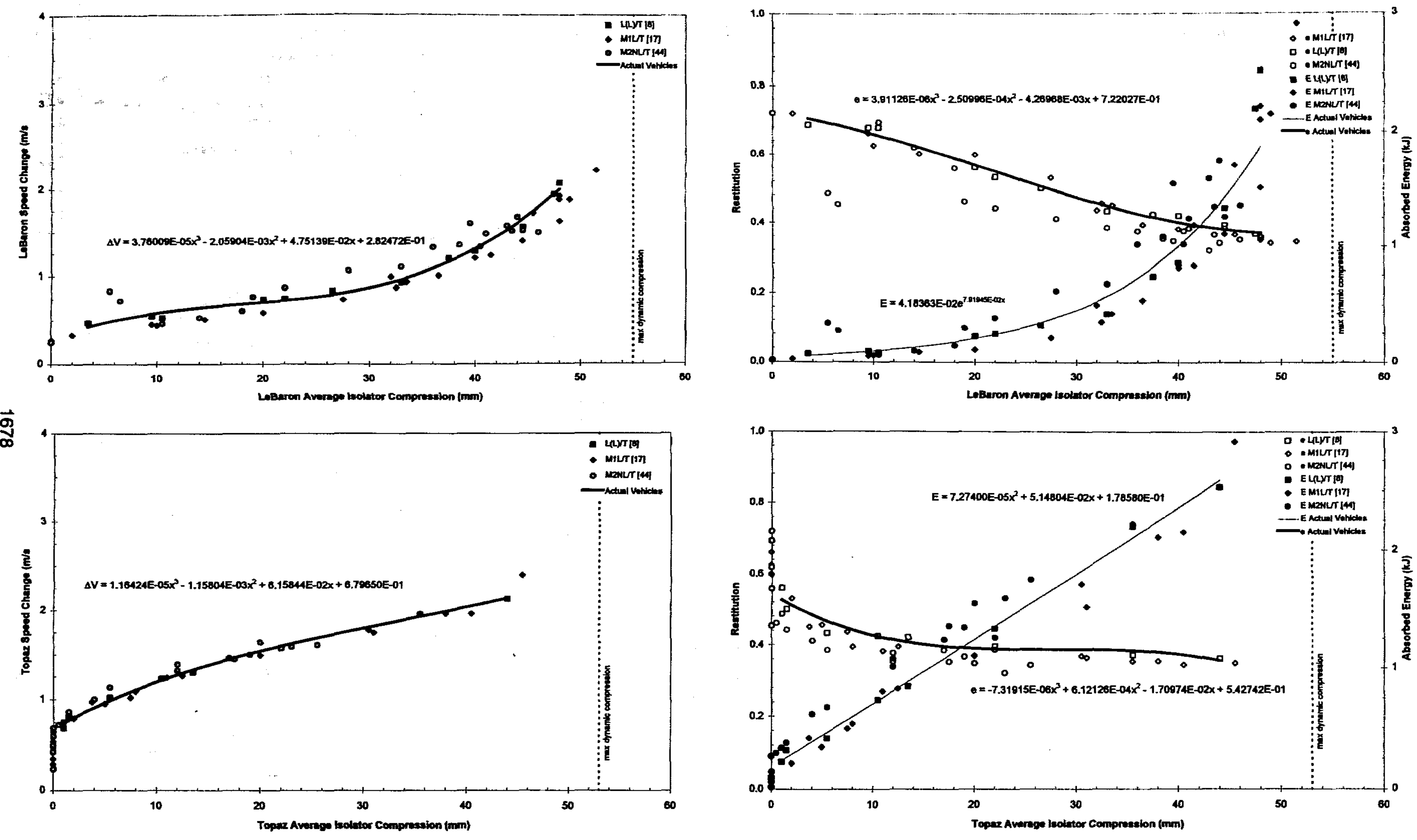


Figures 10d - 10f Group D - LeBaron / Chevette (speed change, absorbed energy, and restitution)

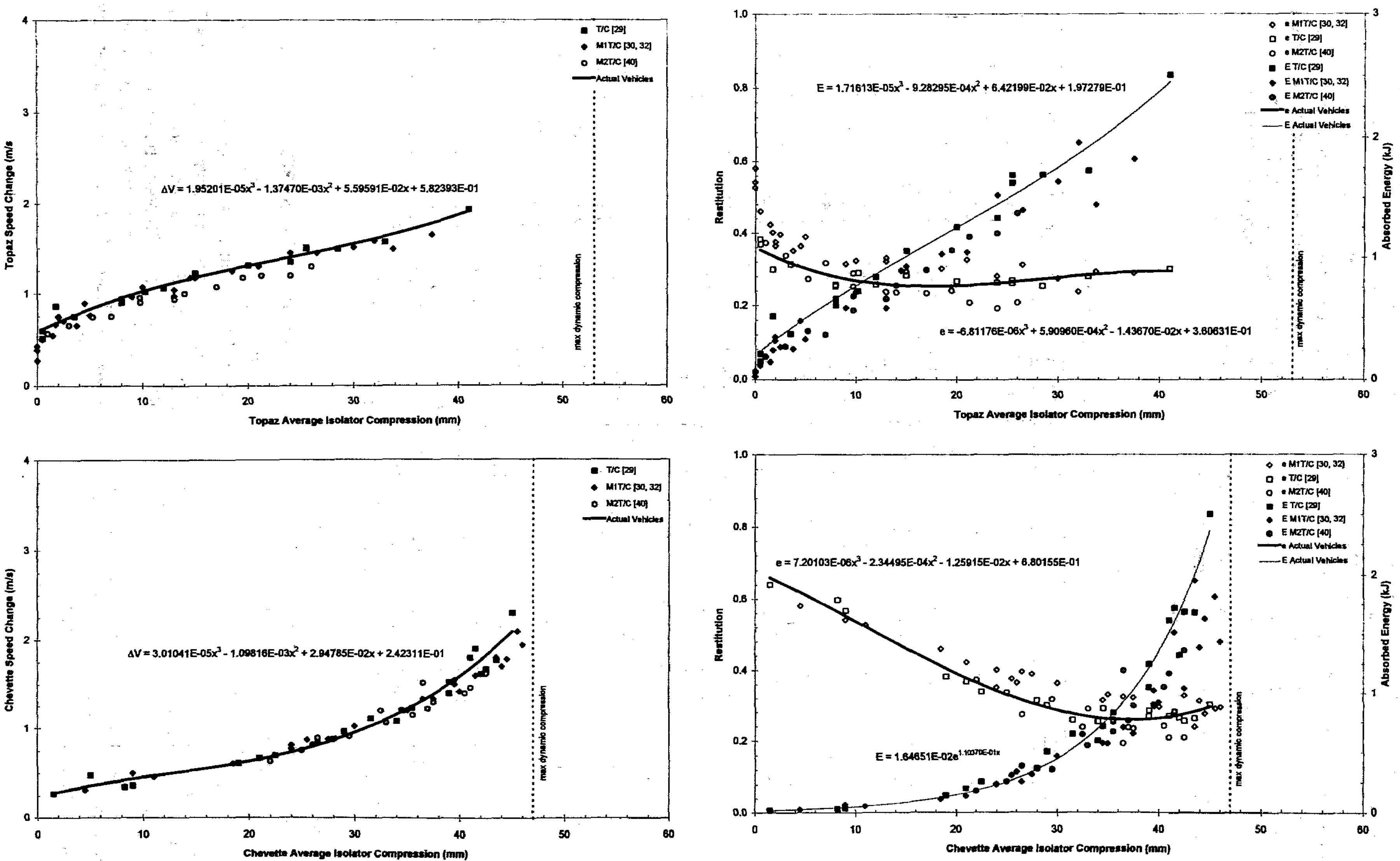




Figures 11a - 11c Group E - Sprint / Barrier and LeBaron / Sprint (speed change, absorbed energy, and restitution)



Figures 12a - 12d Group F - LeBaron / Topaz (speed change, absorbed energy, and restitution)



Figures 13a - 13d Group G - Topaz / Chevette (speed change, absorbed energy, and restitution)

relationships. Each data point from the moving barrier tests was plotted against the value calculated from the actual vehicles' trend line. These plots have the actual vehicle on the x-axis and the moving barrier on the y-axis. The actual speed change was compared to the moving barrier speed change and a best-fit line was plotted as shown in Figures 14 to 20. Similar comparisons were made for the energy and restitution values. The dashed lines represent a confidence interval of 67% (± 1 standard deviation) about the best-fit line using the t-distribution. If the bumpers on M1 and M2 were behaving identically to the actual vehicle, the slope of the best fit line would be close to unity and the unity line shown on the graphs would be contained within the ±1 standard deviation confidence interval.

A simple scoring of each data set was used to assess the barrier performance. The performance scores are summarized in Table 5. The slope of the best-fit line was weighted between 0.9 and 1.1 based on the difference from unity. A slope of 1 is given a slope factor of 1, and slopes of 0.925 or 1.075 were given slope factors of 0.25, and so on. Slopes less than or equal to 0.9 or greater than or equal to 1.1 were given factors of zero. The fraction (0 to 1) of the confidence interval which contained the unity line was also calculated. The fraction is shown under the Unity heading in Table 5. The slope factor and the overlap fraction scores were summed to obtain a total score. A maximum score would be 2 for a best fit line lying on the unity line. A performance score greater than 1 is considered acceptable.

Table 5. Performance Scoring Results.

Group A	Slope	Factor	Unity	Total	Group E	Slope	Factor	Unity	Total
M1dV	0.885	0.00	0.41	0.41	dV - M1 L	0.997	0.97	1.00	**1.97**
M2dV	0.942	0.42	0.29	0.70	e - M1 L	1.016	0.84	1.00	**1.84**
M1E	0.905	0.05	0.15	0.20	En - M1 L	1.041	0.60	1.00	**1.60**
M2E	0.860	0.00	0.07	0.07	dV - M2 L	0.911	0.11	1.00	**1.11**
M1e	1.132	0.00	0.00	0.00	e - M2 L	0.788	0.00	1.00	1.00
M2e	0.789	0.00	0.73	0.73	En - M2 L	0.915	0.15	0.64	0.79
Group B	**Slope**	**Factor**	**Unity**	**Total**	**Group F**	**Slope**	**Factor**	**Unity**	**Total**
M1dV	0.922	0.22	0.62	0.83	dV - M1 L	0.922	0.22	0.27	0.49
M2dV	0.983	0.83	1.00	**1.83**	dV - M1 T	1.023	0.77	1.00	**1.77**
M1E	0.930	0.30	0.44	0.74	e - M1 L	1.042	0.58	1.00	**1.58**
M2E	1.023	0.77	1.00	**1.77**	e - M1 T	1.051	0.49	0.50	0.99
M1e	1.274	0.00	0.40	0.40	E - M1 L	1.118	0.00	1.00	1.00
M2e	1.057	0.43	1.00	**1.43**	E - M1 T	0.999	0.99	1.00	**1.99**
Group C	**Slope**	**Factor**	**Unity**	**Total**	dV - M2 L	0.853	0.00	1.00	**1.00**
M1dV	1.082	0.18	0.49	0.67	dV - M2 T	0.907	0.07	0.76	0.84
M2dV	0.899	0.00	0.00	0.00	e - M2 L	0.780	0.00	0.62	0.62
M1E	1.219	0.00	0.51	0.51	e - M2 T	0.749	0.00	0.00	0.00
M2E	1.194	0.00	0.24	0.24	E - M2 L	0.884	0.00	1.00	1.00
M1e	0.876	0.00	0.00	0.00	E - M2 T	1.011	0.89	0.00	0.89
M2e	1.099	0.01	0.00	0.01					
Group D	**Slope**	**Factor**	**Unity**	**Total**	**Group G**	**Slope**	**Factor**	**Unity**	**Total**
dV - M1 L	0.934	0.34	0.00	0.34	dV - M1 T	0.946	0.46	1.00	**1.46**
dV - M1 C	0.945	0.45	1.00	**1.45**	dV - M2 T	0.894	0.00	0.00	0.00
e - M1 L	1.197	0.00	1.00	1.00	M1 - E T	0.911	0.11	0.43	0.54
e - M1 C	1.055	0.45	1.00	**1.45**	M2 - E T	0.932	0.32	0.00	0.32
En - M1 L	0.860	0.00	0.20	0.20	M1 - e T	1.359	0.00	0.00	0.00
En - M1 C	1.071	0.29	1.00	**1.29**	M2 - e T	1.634	0.00	1.00	1.00
dV - M2 L	1.028	0.72	0.29	**1.01**	M1 - dV C	0.867	0.00	0.58	0.58
dV - M2 C	0.934	0.34	1.00	**1.34**	M2 - dV C	0.781	0.00	0.67	0.67
e - M2 L	0.807	0.00	1.00	1.00	M1 - E C	0.681	0.00	0.31	0.31
e - M2 C	0.813	0.00	1.00	1.00	M2 - E C	0.705	0.00	0.48	0.48
En - M2 L	0.862	0.00	0.55	0.55	M1 - e C	0.834	0.00	1.00	1.00
En - M2 C	0.725	0.00	0.34	0.34	M2 - e C	1.340	0.00	1.00	1.00

VEHICLE TO FIXED BARRIER TESTS

GROUP A - CHEVETTE BUMPER INTO FIXED BARRIER - Figure 14 compares the actual speed change, energy absorption, and coefficient of restitution for the Chevette into the fixed barrier with M1 and M2. The Chevette bumper was attached to M1 and M2 for the barrier impacts in the manner shown in Figures 1 and 2. The test data are shown in Figure 7.

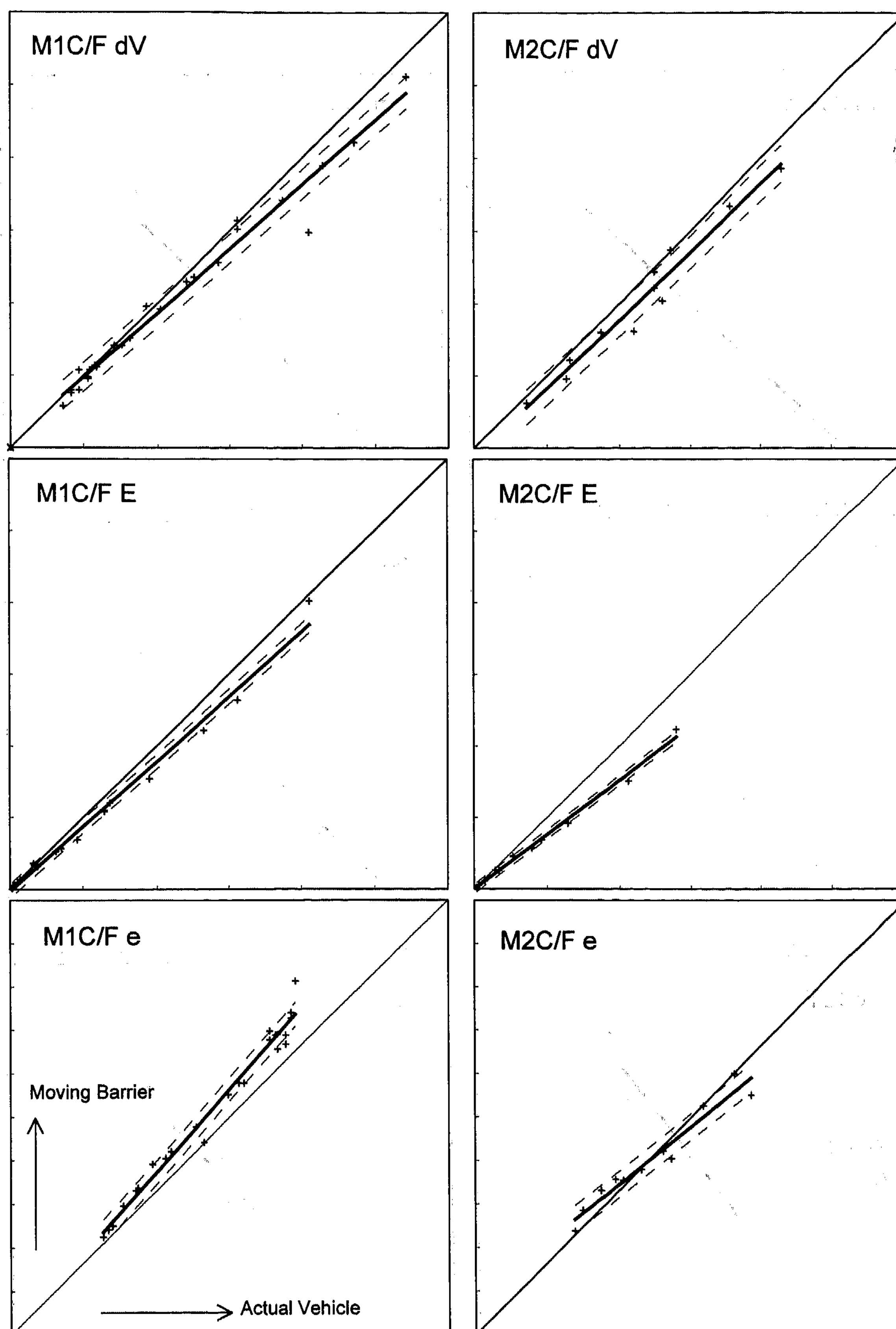


Figure 14. Group A Statistical Comparison- Chevette/Fixed Barrier (ΔV, E, and e).

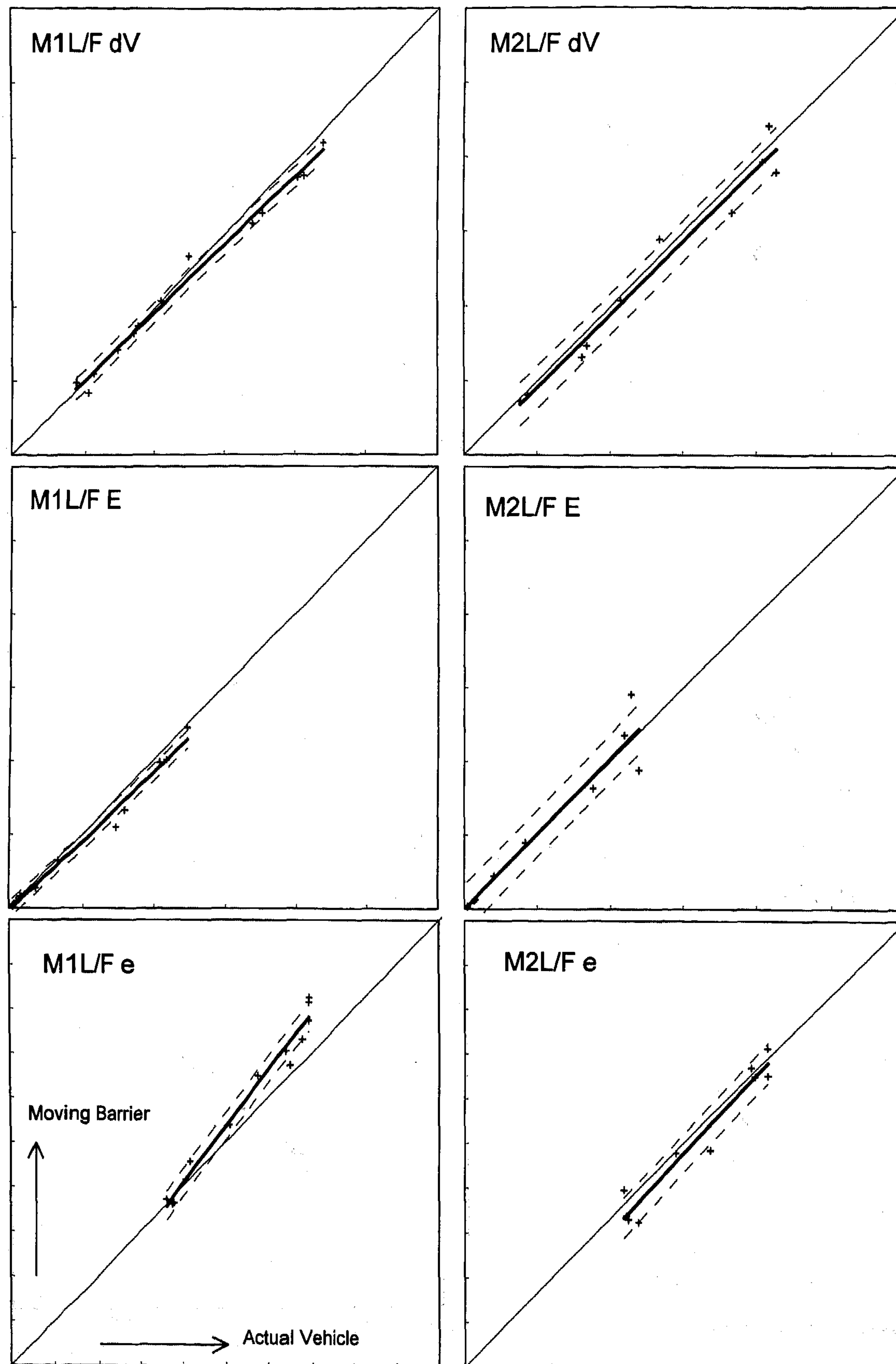


Figure 15. Group B Statistical Comparison -LeBaron/Fixed Barrier (ΔV, E, and e).

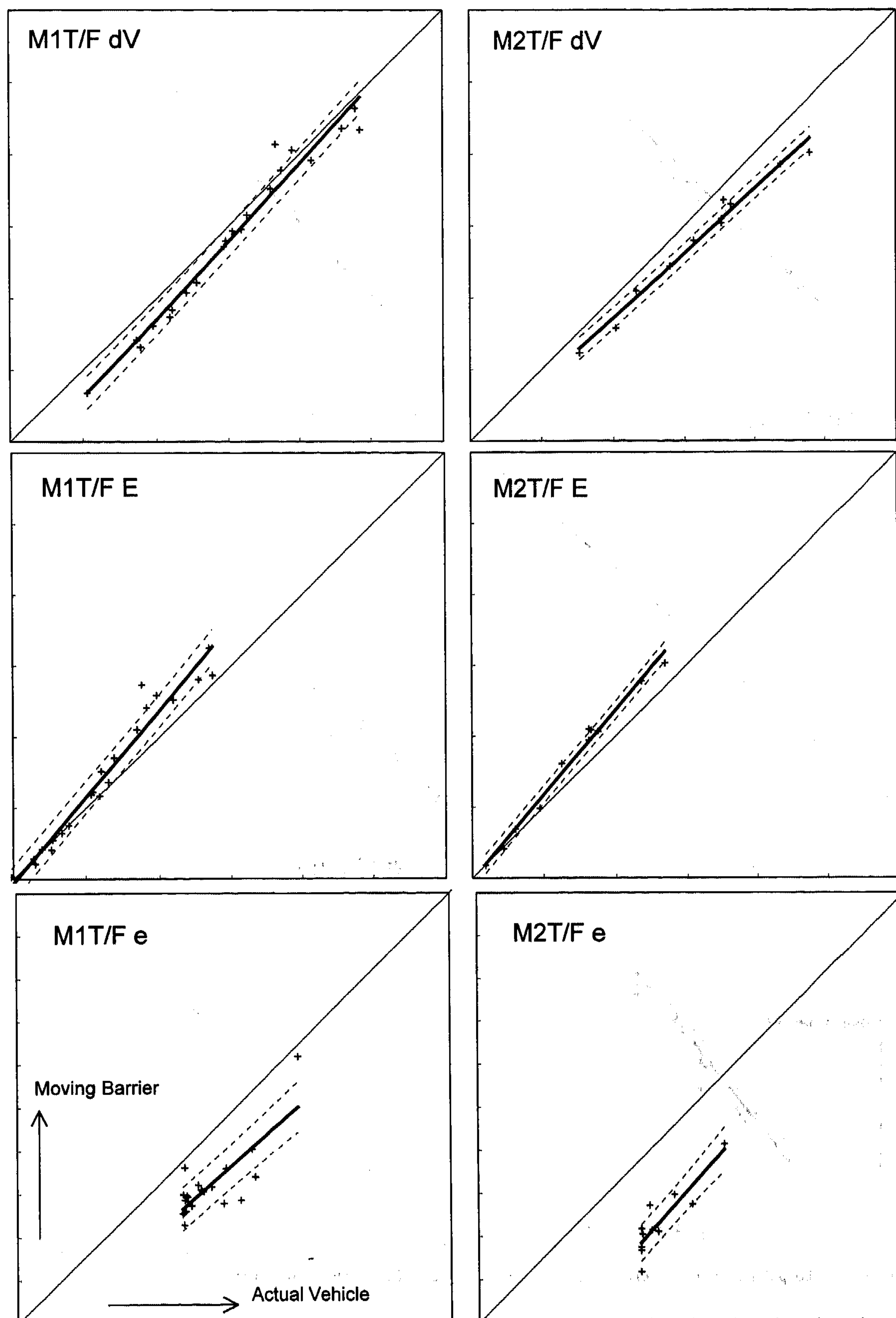


Figure 16. Group C Statistical Comparison -Topaz/Fixed Barrier (ΔV, E, and e).

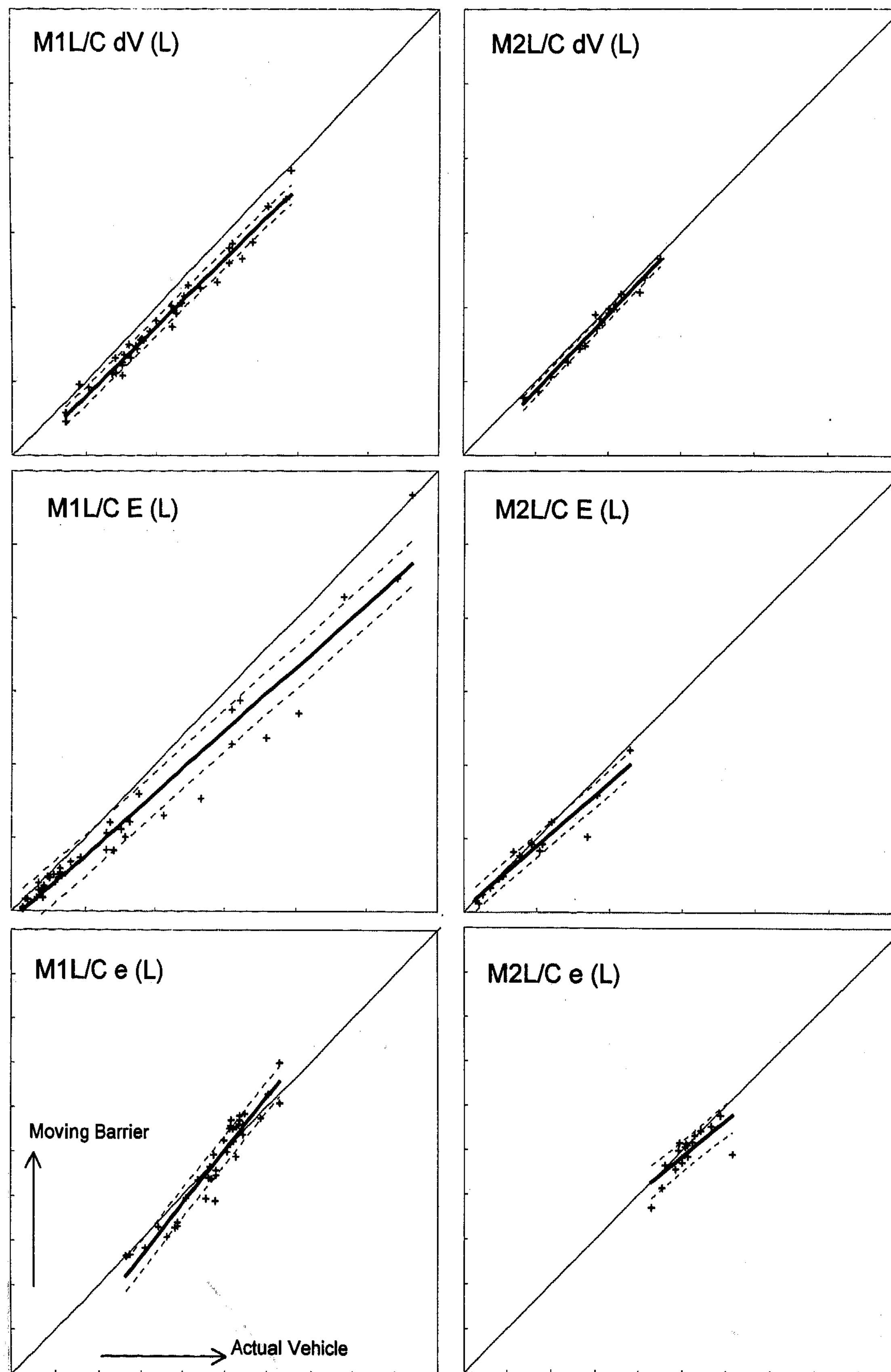


Figure 17a. Group D Statistical Comparison -LeBaron/Chevette (ΔV, E, and e) referenced to LeBaron.

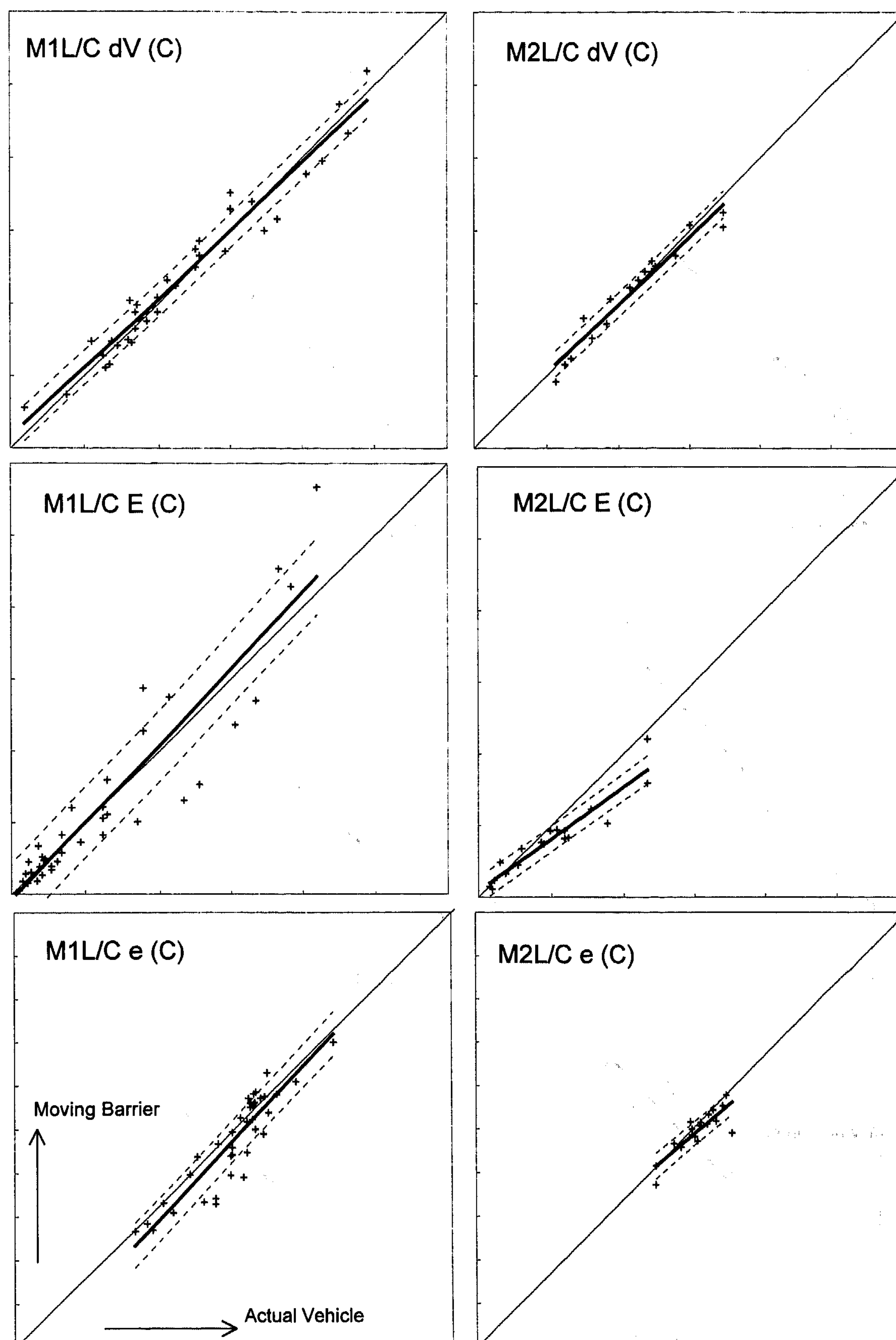


Figure 17b. Group D Statistical Comparison -LeBaron/Chevette (ΔV, E, and e) referenced to Chevette.

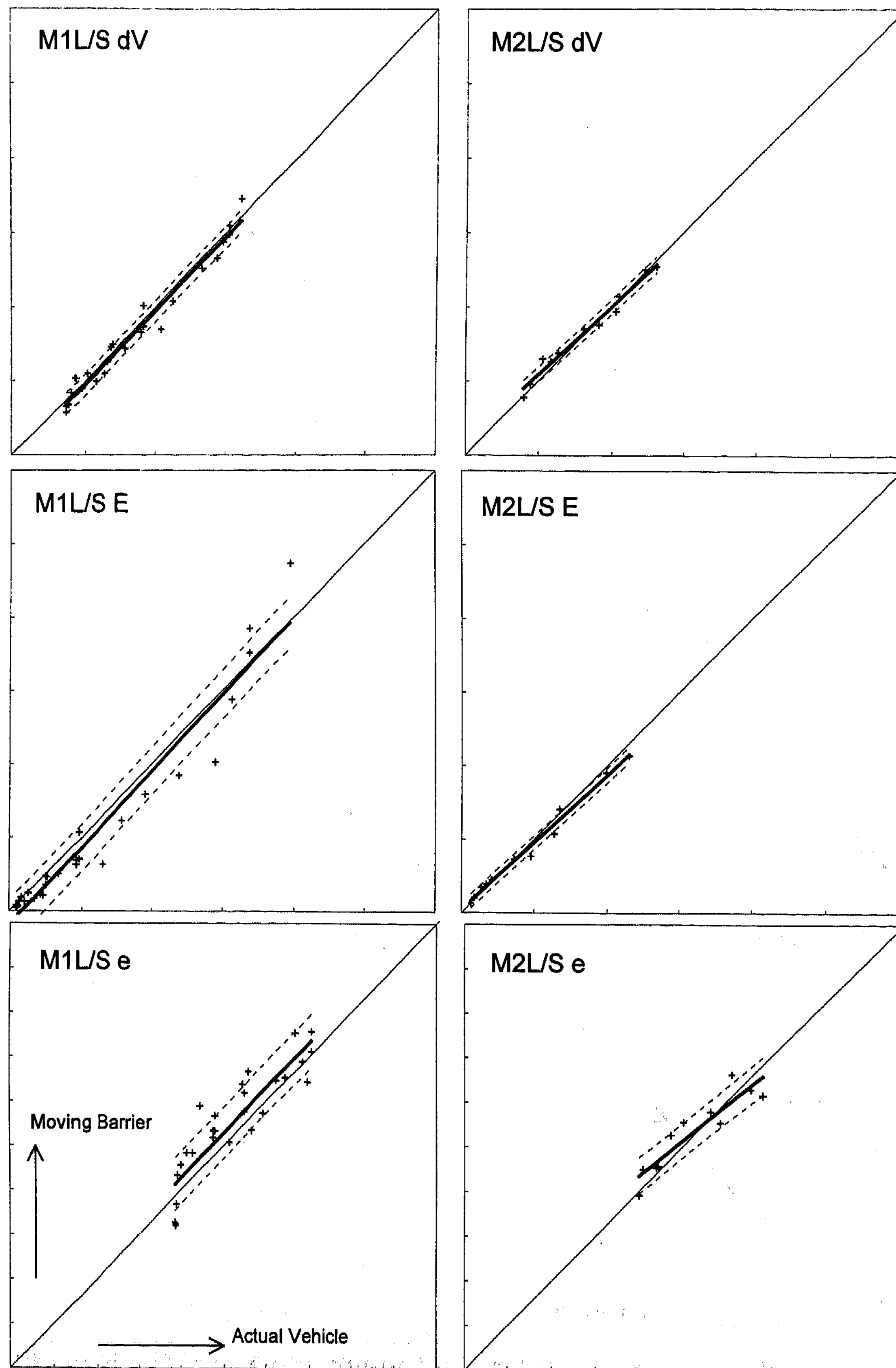


Figure 18. Group E Statistical Comparison -LeBaron / Sprint (ΔV, E, and e).

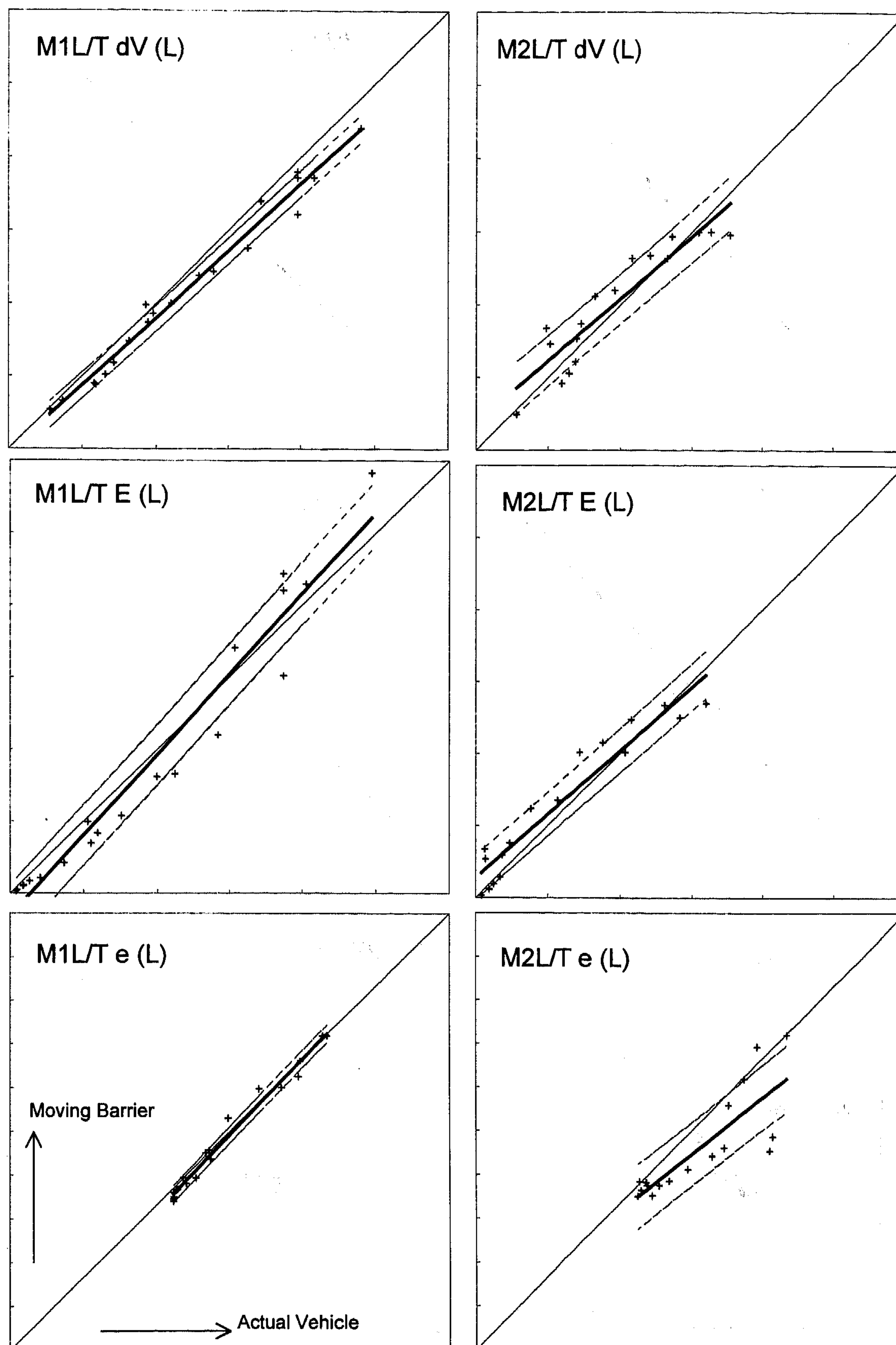


Figure 19a. Group F Statistical Comparison -LeBaron / Topaz (ΔV, E, and e) referenced to LeBaron.

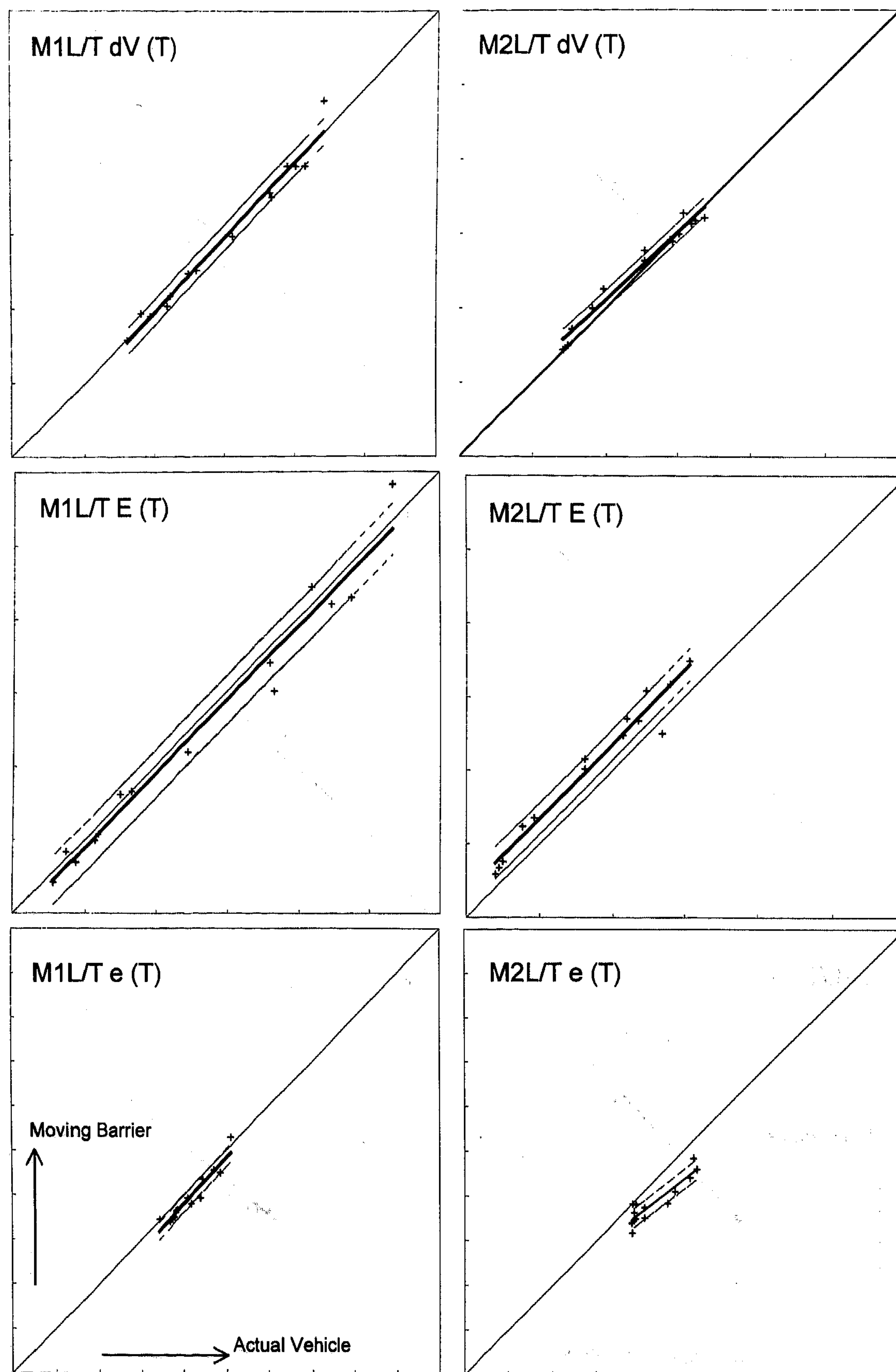


Figure 19b. Group F Statistical Comparison -LeBaron / Topaz (ΔV, E, and e) referenced to Topaz.

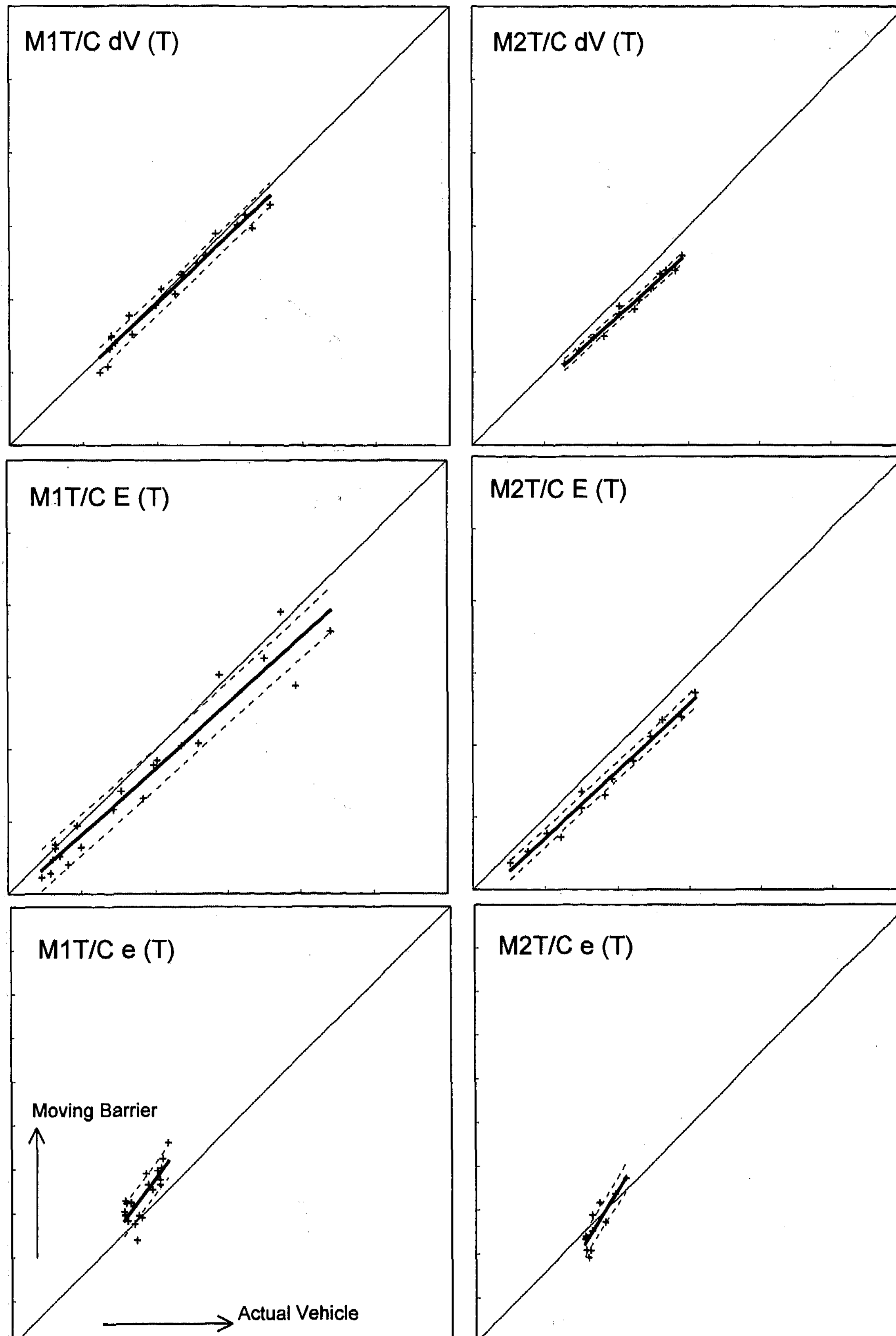


Figure 20a. Group G Statistical Comparison -Topaz/Chevette (ΔV, E, and e) referenced to Topaz.

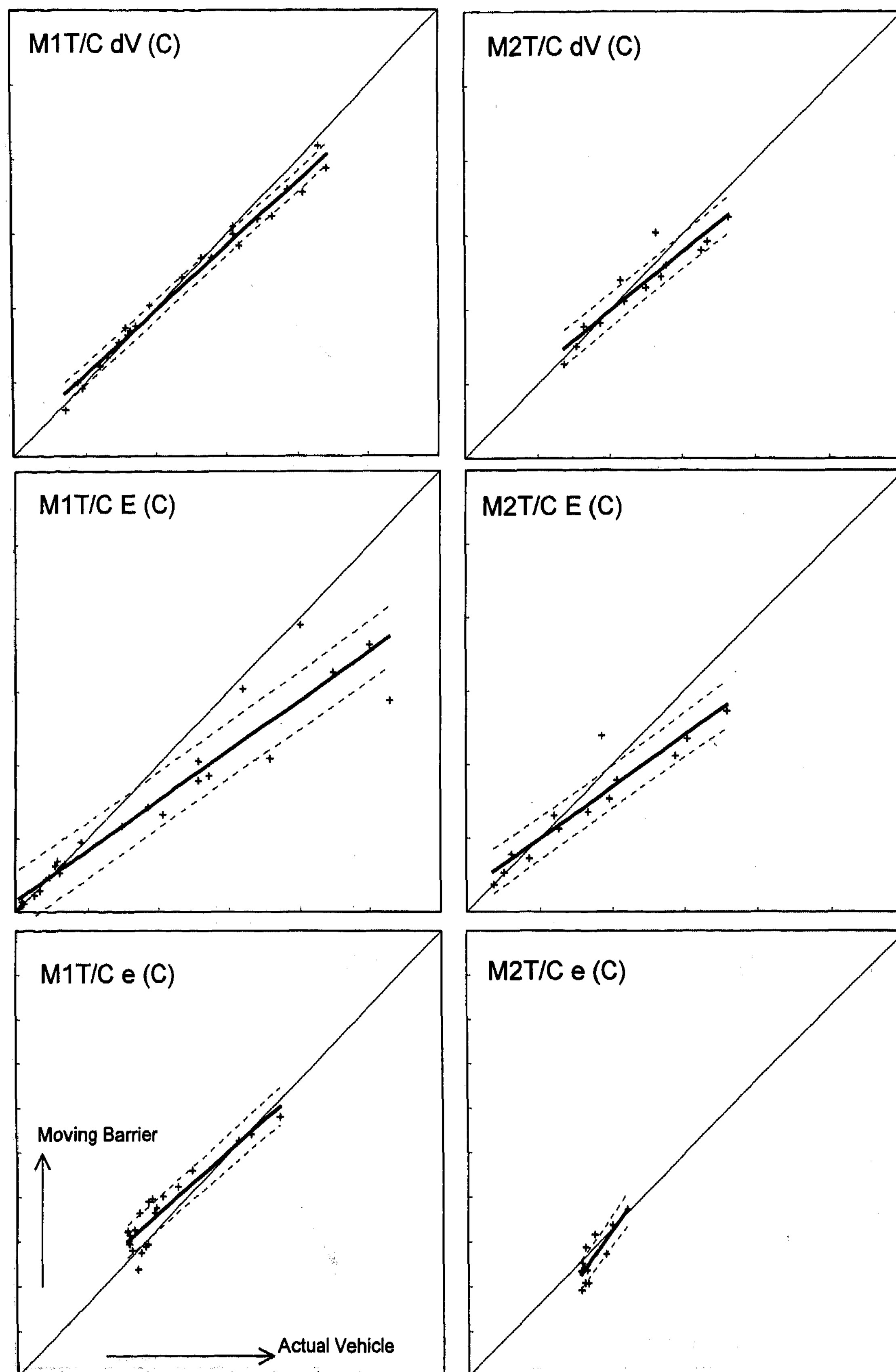


Figure 20b. Group G Statistical Comparison -Topaz/Chevette (ΔV, E, and e) referenced to Chevette.

The speed change comparison charts indicate that both M1 and M2 underestimated the actual Chevette tests except for very low speed changes.

The energy absorption comparison charts show that the energy absorbed in the M1 tests more closely resembles the actual Chevette tests than the M2 tests. However, both M1 and M2 underestimated the actual energy absorption. As the severity increased, the M1 and M2 results drifted further from the actual Chevette tests, more so with M2 than M1.

The coefficient of restitution comparison charts indicate that the M2 tests follow the actual tests slightly better than the M1 tests, which tend to overestimate the actual.

The highest performance scores were for M2's speed change and restitution comparisons, but all were less than 1.

GROUP B - LEBARON BUMPER INTO FIXED BARRIER - Figure 15 shows the comparison charts between the actual LeBaron into the fixed barrier and M1 and M2 with the LeBaron bumper into the fixed barrier. The test data are shown in Figure 8.

The M1 tests underestimated the speed change of the actual LeBaron tests except for speed changes below approximately 1 m/s.

The M1 tests slightly underestimated the energy absorption and overestimated the coefficient of restitution of the actual tests.

In the speed change, energy absorption, and coefficient of restitution comparison charts, M2 was much closer to the actual LeBaron into the fixed barrier than the M1 tests. This is expected since M2 was built from the LeBaron.

The performance scores for M2 were all acceptable.

GROUP C - TOPAZ BUMPER INTO FIXED BARRIER - The results of the fixed barrier impact comparisons for the Topaz, and M1 and M2 with the Topaz bumper attached are shown in Figure 16. The test data are shown in Figure 9.

The speed change from the M1 tests were closer to the actual Topaz at higher speed changes. Conversely, the M2 tests were further from the speed change of the actual Topaz at higher speed tests.

Both the M1 and the M2 tests overestimate the energy absorption compared with the actual Topaz except for the very low speed tests.

M1 and M2 both significantly underestimate the coefficient of restitution compared with the actual Topaz into the fixed barrier.

The highest performance score was for M1's speed change comparison, but at 0.67 the relation is unacceptable.

VEHICLE-TO-VEHICLE TESTS

In the vehicle-to-vehicle tests, two sets of comparison charts were examined. In the first set, the speed change, energy change, and coefficient of restitution of M1 and M2 were compared with that of the actual vehicles based on the isolator compression of the bullet vehicle. The second set of comparison charts were based on the isolator compressions of the target vehicle (if applicable).

GROUP D - LEBARON BUMPER INTO CHEVETTE - Figure 17 compares the actual LeBaron into the Chevette with M1 and M2 with the LeBaron bumper into the Chevette. The test data are shown in Figure 10.

The M1 tests underestimated the speed change compared to the actual LeBaron tests when the comparison was based on the LeBaron isolator compression. M1 compared closely to the actual when the comparison was based on the Chevette isolator compression. The speed change of the M2 tests slightly underestimated the speed change of the actual tests when the comparisons were based on the LeBaron or the Chevette isolator compression. In general the speed changes compared better in the M2 tests than the M1 tests.

The M1 and M2 tests underestimated the energy absorbed in the actual LeBaron tests based on the LeBaron isolator compression. However, M2 was closer to the actual LeBaron than M1. The comparison graphs referenced to the LeBaron isolator compression are shown in Figure 17a.

Based on the Chevette isolator compression, (Figure 17b) the energy absorption in the M1 tests was very close to the actual LeBaron tests and the M2 tests underestimated the energy absorption. This result is significantly different from the energy absorbed referenced to the LeBaron isolator compression.

The coefficient of restitution data from the M1 and the M2 tests were not significantly different than the actual LeBaron tests. This was the case whether the comparisons were based on the LeBaron isolator compression or the Chevette isolator compression.

The highest scores were for the M1 speed change, energy, and restitution comparisons referenced to the Chevette bumper isolator compression. These were all acceptable.

GROUP E - LEBARON BUMPER INTO SPRINT - Figure 18 shows the comparison charts between the actual LeBaron into the Sprint and M1 and M2 with the LeBaron bumper into the Sprint. The test data are shown in Figure 11.

The speed change and energy absorption charts in Figure 18 show that the M1 and M2 tests compare very closely with the actual LeBaron tests. The coefficient of restitution was slightly overestimated in the M1 tests.

The highest scores were for M1's speed change, energy, and restitution, all of which were acceptable.

GROUP F - LEBARON BUMPER INTO TOPAZ - Figure 19 shows the comparison charts between the actual LeBaron into the Chevette and M1 and M2 with the LeBaron bumper into the Chevette. The test data are shown in Figure 12.

M1 slightly underestimated the speed change when the comparison was based on the LeBaron isolator compression (Figure 19a). However, the speed change in the M1 tests closely resembled the actual LeBaron tests when the comparison was based on the Topaz isolator compression (Figure 19b). The speed change in the M2 tests was slightly overestimated at the low speeds and was closer to the actual at higher speeds. This relationship was approximately the same whether the comparison was based on the LeBaron isolator compression or the Topaz isolator compression.

The energy absorption in the M1 and M2 tests was close to the actual LeBaron when the comparison was based on the LeBaron's isolator compression. The energy absorbed in the M1 tests also closely resembled the actual LeBaron tests when the comparison was based on the Topaz isolator compression. However, the M2 tests overestimated the energy absorbed when the comparison was based on the Topaz isolator compression.

The coefficients of restitution for M1 were close to the actual LeBaron. However, the M2 tests underestimated the coefficient of restitution, especially at low speeds. A similar trend is observed for the Topaz referenced charts.

The highest scores were for the speed change and energy relations referenced to the Topaz. The only acceptable LeBaron-referenced group was the M1 restitution.

GROUP G - TOPAZ INTO CHEVETTE - Figure 20 compares the actual Topaz into the Chevette with M1 and M2 with the Topaz bumper into the Chevette. The test data are shown in Figure 13.

The speed change compared closely with the actual Topaz in the M1 tests referenced to the Topaz isolator compression. In the M2 tests the speed change was underestimated, especially when the comparison was based on the Topaz isolator compression. The comparison charts referenced to the Topaz are shown in Figure 20a. Figure 20b shows the test results referenced to the Chevette.

The energy absorption was underestimated using M1 and M2. However, M2 had slightly better results than M1 when the comparison was based on the Chevette isolator compression. M1 had slightly better results when the comparison was based on the Topaz isolator compression.

The coefficient of restitution was difficult to compare since there was little variance in restitution between low speed collisions and higher speed collisions. The M1 tests tend to overestimate the coefficient of restitution slightly. The M2 tests slightly underestimate restitution at the low end and overestimate at the high end.

The only acceptable score was for the M1 speed change referenced to the Topaz isolator compression.

DISCUSSION

The tests did not clearly distinguish one moving barrier over the other. M2 simulated the LeBaron into the fixed barrier better than M1, while M1 simulated the Topaz into the fixed barrier better than M2. Unexpectedly, M1 better simulated the LeBaron during LeBaron bumper into Topaz tests even though M2 was constructed from the LeBaron.

The bullet and target vehicle speed changes were consistently better approximated than the absorbed energy or the restitution. The restitution approximation was rarely acceptable.

The reason for the poor correlation of the collision restitution is unclear, although the two main factors are likely vehicle suspension and stiffness of the isolator mounting on the vehicle.

The absorbed energy of a collision is related to the collision restitution and varies proportionally as shown:

$$E \propto 1 - e^2 \qquad (1)$$

When the restitution was overestimated, the absorbed energy was an underestimate. This is shown in Figure 11 where M1 restitution values lie above the actual vehicle values, while the absorbed energy data points for M1 fall below the actual. The trend line for the absorbed energy data is below unity in Figure 18, and the trend line for restitution is above unity.

Although the data did not always correlate for all three of the collision characteristics (speed change, absorbed energy and restitution), the findings are encouraging. The speed change correlation was on average better than the energy and restitution. Since the energy change is proportional to restitution, modifying the moving barriers to better estimate the restitution of the actual vehicles should result in an overall improvement.

ADDITIONAL FINDINGS

Additional findings that were not the main focus of this paper were examined and are discussed.

SUSPENSION EFFECTS - Test series 22 consisted of 9 tests in which the Chrysler LeBaron with its suspension blocked struck the Chevrolet

Chevette. From the initial test of M1 with the LeBaron bumper into the Chevette, it was hypothesized that the difference in collision behavior between M1 and the actual LeBaron was related to M1's high suspension stiffness. It was noted that there was lower energy absorption and a higher isolator compression for a given speed change during collisions involving the LeBaron bumper mounted on M1. Additional tests were performed with the LeBaron suspension blocked, allowing it to behave more like M1. The results showed that with the modified suspension less energy was absorbed during the collision and the LeBaron had a higher isolator compression for a given speed change than with the suspension fully functional. Although the LeBaron with blocked suspension did not behave exactly like M1 during these collisions, it did show that vehicle suspension is a significant factor during collisions.

CHEVETTE BUMPER ON M1 - In test series 16, M1 with the LeBaron bumper struck the Chevette. In test series 21, M1 with the Chevette bumper struck the LeBaron. These two series were compared in order to determine if there would be differences in speed change, energy absorbed and restitution when the LeBaron bumper was mounted on M1 compared to when the Chevette bumper was mounted on M1. The correlation between isolator compression and speed change did not change between these two series. With the Chevette bumper mounted on M1, the collisions exhibited a lower restitution and a slightly higher energy absorption over the collisions in which the LeBaron bumper was mounted on M1.

MEA 5TH WHEEL AND CALCULATED ΔV COMPARISON - The MEA 5th Wheel speed change results were compared with the speed change calculated from the load cell data. In the x-direction the relation between the impact force and the speed change can be represented by the expression:

$$F\Delta t = m\Delta V \quad (2)$$

Load cell results from the actual vehicle-to-fixed barrier tests were used with this expression to compare the theoretical speed change with the MEA 5th Wheel result. From the load cell data, the term $F\Delta t$ (the average force times the duration) was calculated using the trapezoidal rule. As shown in Figure 21, there is good agreement between the 5th Wheel and load cell methods.

The slope of the best-fit line through the points in Figure 21 is 1.0268, and the r^2 is 0.9981.

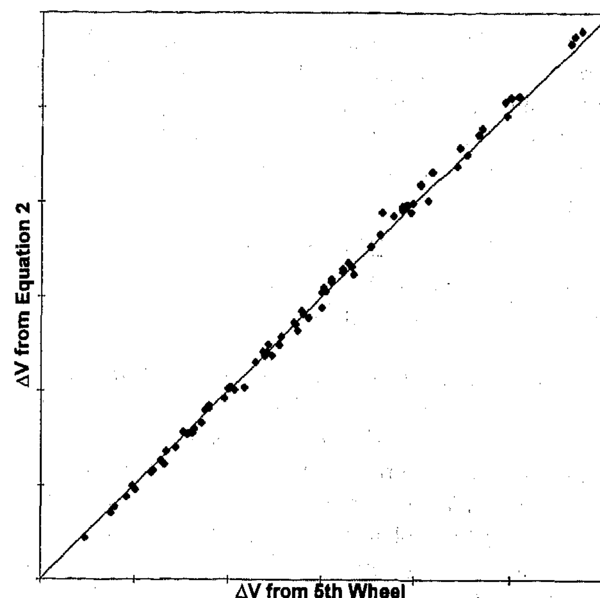


Figure 21. MEA 5th Wheel and Load Cell ΔV Comparison

BUMPER MOUNTING BEAM - Tests were conducted with M2 striking the fixed barrier without load cells and without the rectangular steel tube mounting beam attached. In these tests the bumper isolators were fitted into the frame rail ends at the front of M2 in the same manner as they would be on the actual vehicle. One test series used the LeBaron bumper and the other used the Topaz bumper.

There was no significant difference between the fixed barrier impacts with the LeBaron bumper on M2 with or without the load cells and mounting beam. These data also were not significantly different from the actual LeBaron, as shown in Figure 8.

The speed change versus isolator compression results for M2 with the Topaz bumper tests were approximately the same with or without the front mounting beam as shown in Figure 9. The energy absorption in the tests without the mounting beam were slightly lower and closer to the actual Topaz than the tests with the mounting beam. The coefficient of restitution was significantly higher in the tests without the beam than the tests with the beam but still lower than the actual Topaz.

CONCLUSIONS

FIXED BARRIER TESTS

1. M1 did not give acceptable results for speed change versus isolator compression. M2 gave acceptable results only when simulating the LeBaron.

2. M1 consistently overestimated the coefficient of restitution and underestimated the energy absorbed with the exception of the Topaz tests.

3. M2 tended to underestimate the coefficient of restitution except in the LeBaron tests. M2 also underestimated the Chevette energy change, accurately estimated the energy change in the LeBaron tests, and overestimated the Topaz energy change.

VEHICLE TO VEHICLE TESTS

1. The data presented indicate that for some vehicle pairs, a moving barrier can be an effective surrogate.

The best results were for tests between the LeBaron and Sprint. In these tests, the moving barrier results were similar to the actual LeBaron. It is not clear which Sprint parameters caused this test set to be different from the other tests with the LeBaron as the bullet vehicle. It may be due to the foam core bumper on the Sprint. The Sprint also had the softest suspension and was the lightest vehicle.

2. In most of the vehicle-to-vehicle tests the moving barrier speed change versus isolator compression results are similar to the actual vehicle tests.

3. A single conclusion cannot be made about the use of moving barriers because the data were dependent on which vehicle was being simulated as well as which vehicle was struck. For example, both moving barriers gave close results when simulating the LeBaron into the Sprint. However, M2 did not give acceptable results simulating the LeBaron into the Topaz even though M2 is a modified version of the actual LeBaron.

4. Blocking the suspension on the LeBaron caused it to behave more like M1. This suggests that part of the reason for the differences between M1 and the actual LeBaron was the high suspension stiffness of M1.

5. Care must be exercised when mounting candidate bumpers to the moving barrier in order to duplicate as closely as possible the actual bumper mounting conditions. In particular, bumper wrap-arounds may carry a significant portion of the bumper collision loads. Suspension characteristics, center of mass location, mass distribution, and axle to bumper distance may be influential variables in achieving fidelity between a vehicle and the vehicle's bumper mounted on the moving barrier.

RECOMMENDATIONS

1. The suspension on M1 should be redesigned so that its spring stiffness can be adjusted to closer resemble a passenger car suspension instead of the current truck-like suspension.
2. A new system for mounting bumpers to M2 should be developed to allow more vertical and horizontal adjustability. The new mounting system should not affect the front stiffness of M2.
3. A series of tests should be conducted to determine how well the moving barriers simulate pickup trucks, vans, and sport utility vehicles equipped with rigidly mounted steel beam bumpers.
4. Further research into the role played by the moving barrier suspension and the steel tube mounting beam is needed.

ACKNOWLEDGMENT

The authors wish to thank Mark Bailey, Colin Blair, Craig Brown, Robert Dyck, Michael Gardiner, Kurt Ising, John Miles, Gunter Siegmund, Robert Steves, David Switzer and the rest of the staff at MacInnis Engineering for their observations and assistance. Staging over 700 tests does not occur without the devotion and assistance of many individuals.

REFERENCES

1. King, D.J., Siegmund, G.P. and Bailey, M.N., "Automobile Bumper Behavior in Low-Speed Impacts", SAE 930211.
2. Siegmund, G.P. Bailey, M.N. and King, D.J., "Characteristics of Specific Automobile Bumpers in Low-Velocity Impacts", SAE 940916.
3. Bailey, M.N., Wong, B.C., and Lawrence J.M., "Data and Methods for Estimating the Severity of Minor Impacts", SAE 950352.
4. Szabo, T.J. et al, "Human Occupant Kinematic Response to Low Speed Rear-End Impacts", SAE 940532.
5. Malmsbury, R.N., Eubanks, J.J., "Damage and/or Impact Absorber (Isolator) Movements Observed in Low Speed Crash Tests Involving Ford Escorts", SAE 940912.
6. McConnell, W.E., et al, "Analysis of Human Test Subject Kinematic Responses to Low Velocity Rear End Impacts", SAE 930889.
7. McConnell, W.E., et al, "Human Head and Neck Kinematics after Low Velocity Rear-End Impacts - Understanding "Whiplash"", SAE 952724.
8. Bailey, M. "Assessment of Impact Severity in Minor Motor Vehicle Collisions", Journal of Musculoskeletal Pain.
9. Romilly, D.P., et al, "Low Speed Rear Impacts and the Elastic Properties of Automobiles", Proceedings 12th International Conference of Experimental Safety Vehicles, 1989, Gothenburg, Sweden.
10. Santella et al. "Energy Transfer Characteristics of Thermoplastic Bumper Beams in a Pendulum Impact", SAE 920523.

SAE #970956
Reviewer's Discussion
by Ernest Klein, Collision Research & Analysis, Inc.
Determination of Bumper Characteristics Using Prototype Moving Barriers
David J. King, Mark Bailey, Authors

We use moving barriers in the study of side and rear-end collision studies at magnitudes where significant crush damages are induced. This paper examines the application of proposed moving barriers to low speed rear-end collisions. The authors present their proposed moving barriers and their applications to measure test vehicle bumper isolator compression in low speed rear end collisions.

While reasonable results relating to Delta V versus isolator compression were obtained by the authors, additional work is needed for the improvement of such systems.

One may also question the reasoning behind the development of a moving barrier system to study such low speed collisions, when using actual vehicles for such studies is most representative. The cost associated is only the replacement of the bumper or the bumper and the EA units will restore those vehicles to pre-test conditions.

SAE # 970956
Reviewer's Discussion
by Wayne McCracken, Research Engineers, Inc.
Determination of Bumper Characteristics Using a Prototype Moving Barrier
David J. King, Mark Bailey, Authors

This paper is very helpful in extending the research into bumper shock compression versus Delta V in rear impacts. The authors have done very well documented study which looks at this relationship on many different vehicles.

970960

Effects of Restitution in the Application of Crush Coefficients

Raymond R. McHenry and Brian G. McHenry
McHenry Consultants, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Effects of restitution on damage interpretations are compounded by the fact that restitution acts to reduce the amount of residual deformation, for a given maximum dynamic crush, while also acting to increase the total impact speed change. This paper presents a revised analytical procedure to include restitution effects for the CRASH program and refinements to the restitution modeling within the SMAC program. The conversion of vehicle impact test results into inputs for the two revised programs is also included. The effects of the refinements to the damage analysis procedures on reconstruction results are illustrated by direct comparisons with corresponding results produced by the original SMAC and CRASH programs and with measured data from full scale vehicle impact tests.

INTRODUCTION

During a motor vehicle collision, the maximum dynamic deformation generally exceeds the residual deformation. Subsequent to the peak dynamic deformation, the collision partners begin a restitution phase as the deformed structures restore kinetic energy, or "spring" back. The restitution force level and duration determine the impulse that acts on the collision partners during the restitution phase.

When an accident vehicle is examined, the residual, or permanent, deformation is observed and/or measured. The original form of damage analysis in CRASH does not include provisions for the effects of restitution. The original SMAC collision routine includes a simplified restitution model which is cumbersome to apply, can be sensitive to time increment size, and tends to over-predict the residual damage. The resulting effects on the accuracy of damage-based reconstructed values of ΔV, for the case of direct, central barrier collisions, ranges from approximately 10 to 30% underestimates, depending on properties of the specific vehicle and the extent of residual crush. For the case of oblique, non-central collisions, a similar range of effects is anticipated on the basis of indirect measures of corresponding restitution values [1, 2][1].

At the present time, crush coefficients for vehicle collision analysis are predominantly based on impact speeds and damage measurements from rigid, fixed barrier crash tests. The residual damage is correlated with the impact speed by means of fitted linear relationships. In general, there is no consideration given to the effects of restitution in applications of the fitted crush coefficients. However, the ignored effects of restitution on the total impact speed-change, corresponding to a given amount of residual crush, are compounded by the fact that restitution acts to reduce the amount of residual deformation, for a given maximum dynamic crush, while also acting to increase the total impact speed change. Thus, substantially different vehicles can share nearly equal slopes and intercepts in CRASH-type plots of the approach period speed-change as a function of residual crush. This can occur even though the actual exposure severity for a given residual crush may be significantly different.

The effects of restitution on damage interpretations became an important topic of interest in relation to recent efforts by McHenry Consultants, Inc. aimed at (a) reducing the sensitivity of restitution control in the original SMAC computer program [3] and (b) achieving significant improvements in the general accuracy level of damage interpretations. The related modifications of the SMAC and CRASH computer programs have retained the existing forms of the fitted crush coefficients. However, supplementary information regarding the restitution behavior of individual vehicles serves as the basis for refinements in the damage analysis aspects of the cited computer programs.

ANALYTICAL APPROACH

The analytical developments presented in this paper are based on limited test data that are available on the restitution behavior of automobiles in direct central collisions against rigid barriers. It is assumed herein that an analytical

1 Numbers in brackets [ ] indicate references at end of paper

model of the unit-width structural properties, in terms of the load-deflection characteristics during loading and unloading, that adequately correlates with the results of direct central collision tests, will also produce a reasonable approximation of restitution behavior in oblique, non-central collisions.

Direct measures of restitution in oblique, non-central collisions of automobiles are not known to be readily available. However, analytical procedures have been developed (e.g., Brach [1], Ishikawa [2]) to extract approximations of the effective coefficient of restitution from such test data. These approaches may provide a means of extending the current investigation to include comparisons with oblique, non-central collisions.

The assumed form of crush behavior, including restitution, is depicted in Figure 1A. The objective of the analytical developments has been to establish relationships among the variables depicted in Figure 1A and the existing CRASH coefficients A and B [4] that will **maintain the linear relationship depicted in Figure 1B**. On the basis of results of the analysis, it is concluded that the linear relationship depicted in Figure 1B can be maintained without changes in A and B, over wide ranges of properties of the form depicted in Figure 1A. Thus, the relationship of Figure 1B does not uniquely define a combination of crush and restitution properties. Restitution consists of two separate aspects: (1) a partial dimensional recovery and (2) a partial restoration of kinetic energy. The specific combination of the two restitution aspects is determined by three measures of restitution behavior from test data that serve as the basis for four fitted constants (K_1, K_2, ρ, Γ, see Appendices). The crush resistance during increasing loads, K_1, is determined by CRASH coefficients A and B and by the extent of dimensional recovery. The unit-width crush resistance during unloading, K_2, is determined by K_1 and the extent of restoration of kinetic energy. The nature and extent of effects on damage interpretations are outlined in the following paragraphs:

In Figures 2 and 3, full-frontal crush properties are depicted for two substantially different equal-width hypothetical vehicles which share identical crush coefficients (i.e., A and B in the CRASH format). Yet, at 20 inches of residual crush, the total impact speed-changes of the two vehicles differ by approximately 23% (i.e., approximately 7.3 MPH) as shown in Figure 4 and 5.

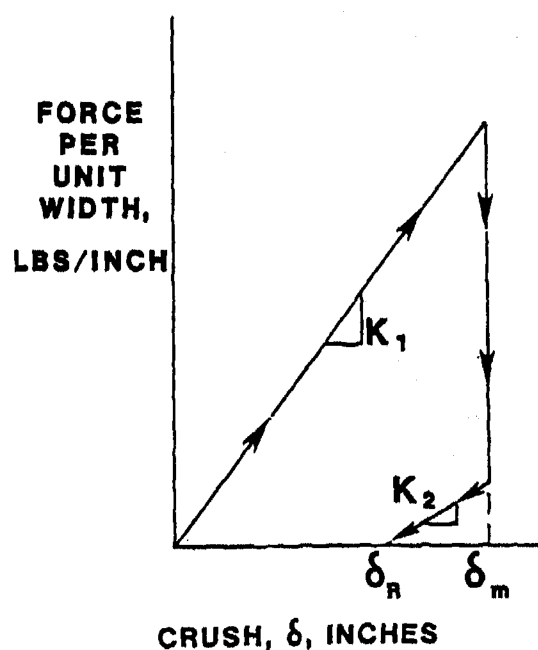


Figure 1A: Force v. Crush

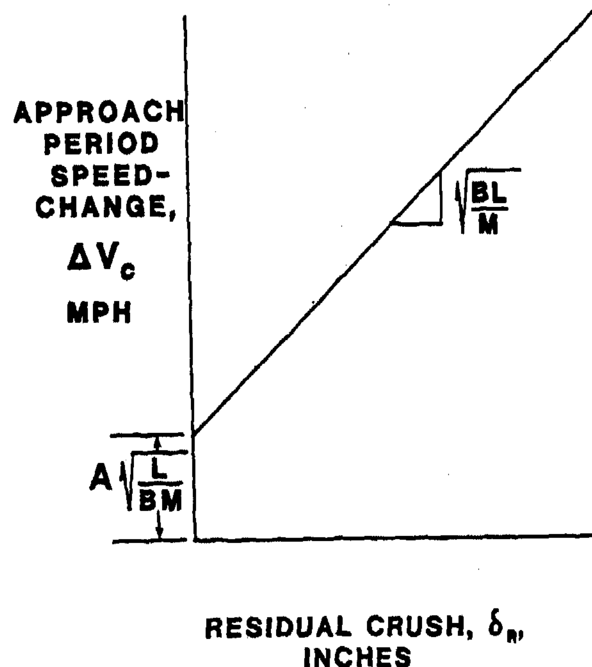


Figure 1B: Approach-Period Speed-Change vs. Residual Crush, Symmetrical Central Impacts

Figure 1 Assumed Form of Crush Behavior

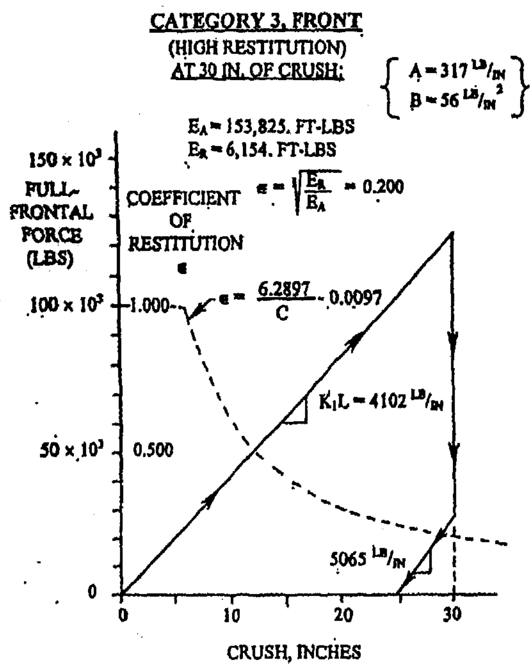


Figure 2 Full-Frontal Crush Properties

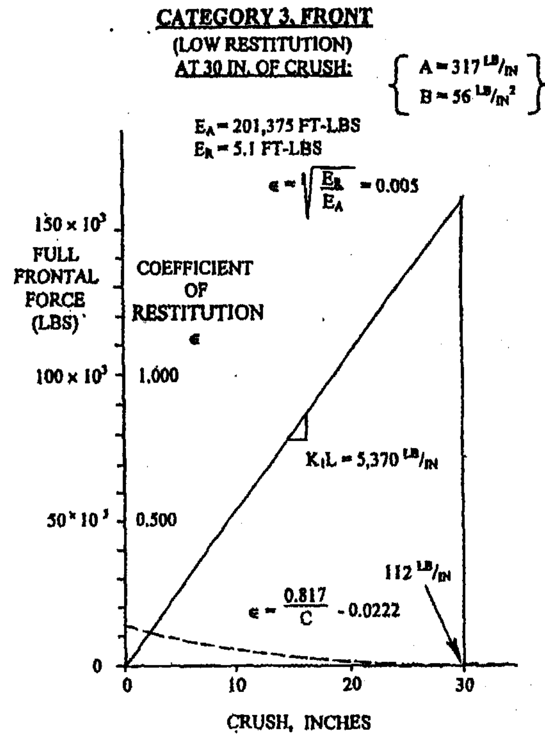


Figure 3 Full-Frontal Crush Properties

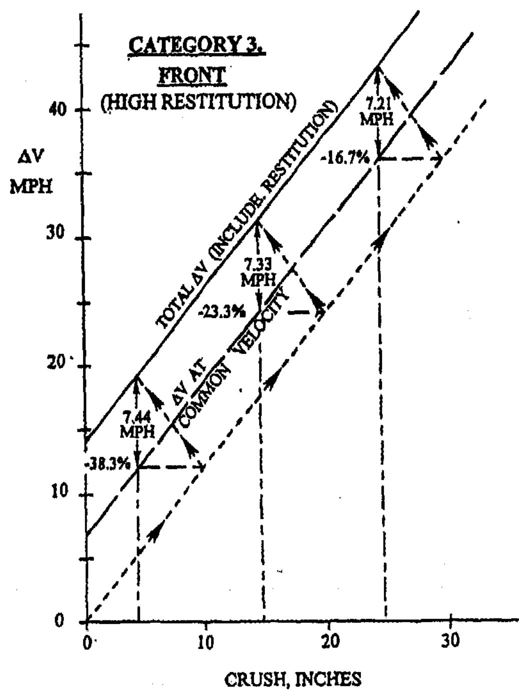


Figure 4 Impact Speed Change v. Crush, Category 3 Frontal, High Restitution

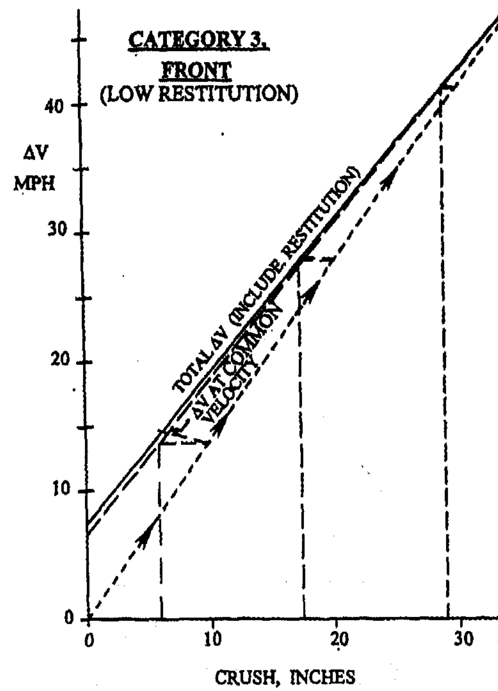


Figure 5 Impact Speed-Change v. Crush, Category 3 Frontal, Low Restitution

The vehicle properties depicted in Figures 2 and 3 are based on theoretical requirements to achieve a linear relationship between ΔV_o (i.e., the impact speed-change during the approach period) and the residual crush. To the extent that such a linear relationship constitutes a reasonable approximation of actual behavior, the underlying crush resistance and restitution definitions of Figures 2 and 3 are supported. It should be noted that the low-restitution vehicle of Figure 3 is approximately 31% stiffer than the high-restitution vehicle of Figure 2. Yet the crash behavior of each vehicle is defined by identical values of the current crush coefficients:

per reference [5]:

$$A = 317 \text{ LB/IN}$$

$$B = 56 \text{ LB/IN}^2$$

per reference [6]:

$$b_0 = 6.741 \text{ MPH}$$

$$b_1 = 1.191 \text{ MPH/IN}$$

In view of the significantly different values for the total ΔV at given values of residual crush as demonstrated in Figures 4 and 5, it is obvious that the current crush coefficients must be supplemented with restitution information. It should be noted that an absence of restitution throughout the entire range of deformation would produce zero values for the fitted coefficients A and b_0.

COMPARISON OF RESTITUTION IN ORIGINAL AND REVISED SMAC

At the time of development of the original SMAC computer program, (1972, Reference [3]) emphasis was on demonstration of the feasibility of the overall concept. The selected analytical approach to restitution at that time (presently retained in both original SMAC and EDSMAC) included the use of extensive simplifying assumptions aimed at reducing the requirements for associated computer memory and logic. In particular, identical load-deflection rates were applied for loading and unloading of the individual radial vectors, that define the collision interface, and the unloading was implemented at force levels close to the peak values [7]. As a part of the selected simulation approach, the "unloaded" lengths of the individual vectors were adjusted at each time increment during increases in loading. The associated coding thereby avoided any needs for logic to detect the end of loading at the individual radial vectors and for additional memory to store related information. The selected original approach permits reasonable control of restitution over a limited operating range about a given value of deformation for which restitution information is available. However, the high load level during restitution makes the simulation excessively sensitive to time-increment size. It is also cumbersome to apply and it tends to over-predict the residual damage.

Since the early 1990's the extent of limitations on logic and memory, that are imposed by readily available computers, are substantially reduced. Also, the available data on restitution behavior have been increased. Therefore, the collision routine of SMAC has been modified to achieve improvements in reconstruction accuracy. Results representative of the changes in restitution and in the ratio of residual to maximum dynamic crush, each as a function of the maximum dynamic crush and the residual crush, are depicted in Figures 6A and 6B, respectively

The ratio, δ_f/δ_m, clearly must vary from zero, at zero residual crush, to 1.000 at that value of residual crush where there is no further restitution. Figure 6B shows that original SMAC (EDSMAC) does not vary the ratio of δ_f/δ_m as a function of δ_f in a realistic manner.

It is obvious from Figures 6A and 6B that the ratio of (δ_f/δ_m) from the original SMAC program (or the equivalent ratio (C_R/C_T) in [8]) should not be used as a basis for determining the appropriate K_V value for the SMAC program corresponding to the fitted CRASH coefficients A and B. If a measured value of $(\delta_f/\delta_m)_1$ for a given maximum dynamic crush value, $(\delta_m)_1$ is available from a crash test, equation (8) in Appendix 1 can serve to define the appropriate K_v for given values of A and B.

Implementation of a revised damage analysis procedure, that includes restitution effects, was outlined for the CRASH computer program in [9]. A corresponding revision, which was recently incorporated in the SMAC computer program, has been applied to generate the responses presented on the following pages. Detailed analytical relationships are presented in the Appendices.

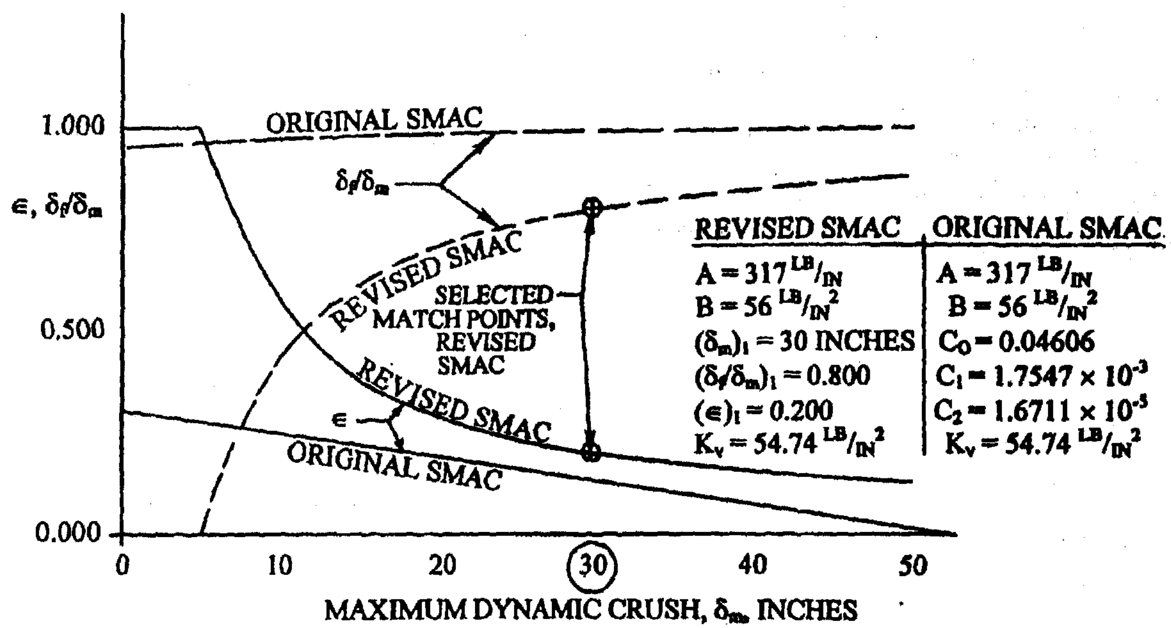


Figure 6A Plotted Against Maximum Dynamic Crush

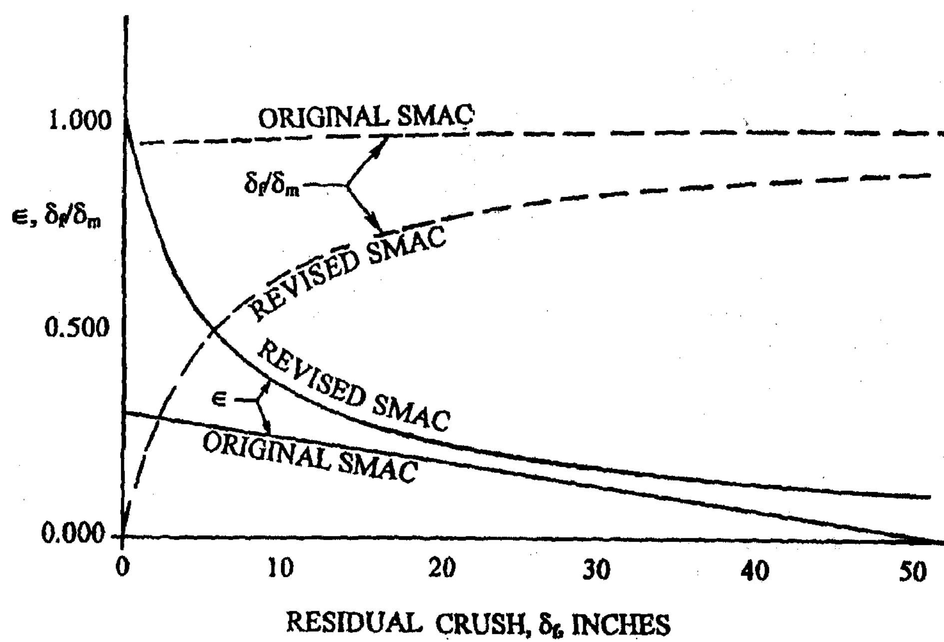


Figure 6B Plotted Against Residual Crush

Figure 6 Simulated Restitution Behavior in Original and Revised SMAC Routines

RESULTS

Damage-Based Reconstructions of Delta-V (ΔV)

In the following comparisons of (1) reconstruction results obtained with the original CRASH (EDCRASH) program and the revised CRASH program and also (2) reconstruction results obtained with the original SMAC (EDSMAC) program and the revised SMAC program, the differences in damage-based ΔV values are produced entirely by the inclusion of realistic restitution properties in a fully defined hypothetical vehicle. Neglect of restitution effects constitutes an analytical error since it involves the total omission of a significant aspect of the collision, particularly at impact speed changes below 30 mph.

In the application of any reconstruction technique to physical evidence there are many potential sources of error which must be properly taken into account. The presented comparisons isolate the changes in reconstruction results that are produced by effects of restitution.

It should be noted that meaningful direct comparisons of theoretical results with test data require that the accuracy and the repeatability of the test data be established. Also, such comparisons yield the total error from the combined sources.

A comparison of results obtained with the original CRASH program and the revised CRASH program for a defined hypothetical vehicle is presented in Figure 7. The figure illustrates the fact that an application of the original CRASH damage analysis procedure can produce ΔV errors (i.e., underestimates) in the range of 10-30%.

A comparison of results obtained with the NHTSA SMAC (EDSMAC) program and the revised SMAC program for the same hypothetical vehicle is presented in Figure 8. The figure illustrates the fact that an application of the NHTSA SMAC (EDSMAC) program for damage analysis purposes can also produce ΔV errors (i.e., underestimates) in the range of 10-30%.

The original form of restitution control in the SMAC program acts to return part of the absorbed energy but it does so with a less-than-actual dimensional recovery (Figure 9, Figure 10). As a result, a purely damage-based determination of ΔV by means of the original SMAC program tends to underestimate the true value of ΔV due to the fact that the residual damage for a given ΔV is overestimated. The cited error source in damage interpretations in original SMAC has led to some misguided adjustments in the crush stiffness (e.g., [10]) to achieve a match of predicted damage extent.

A comparison of the results obtained by the various versions of SMAC and CRASH are combined on Figure 11. It should be noted that the results for the revised versions of SMAC and CRASH are identical. The retention of an excessive amount of predicted residual damage by original SMAC makes the ΔV errors at low speeds somewhat greater than those of original CRASH.

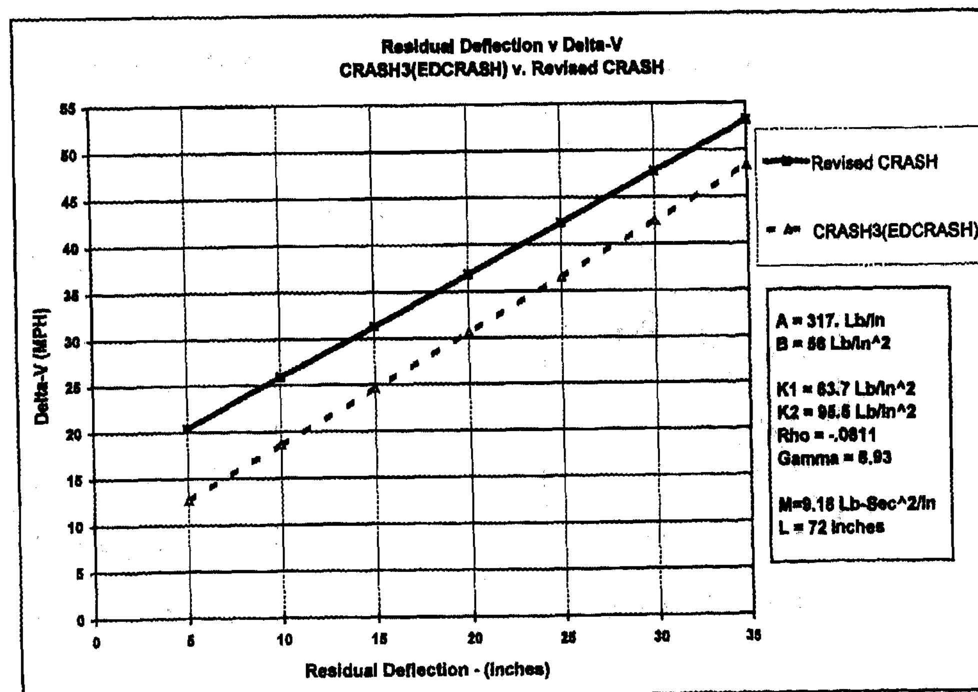


Figure 7 Comparison Residual Deflection Delta-V CRASH3 (EDCRASH) v. Revised CRASH

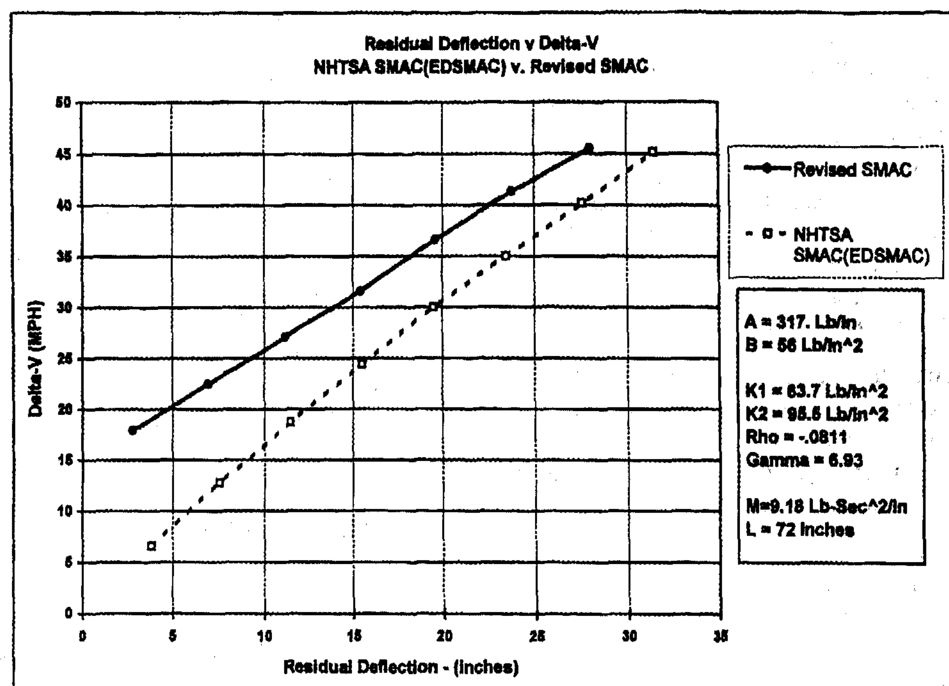


Figure 8 Residual Deflection v. DeltaV for Original SMAC (EDSMAC) v. Revised SMAC

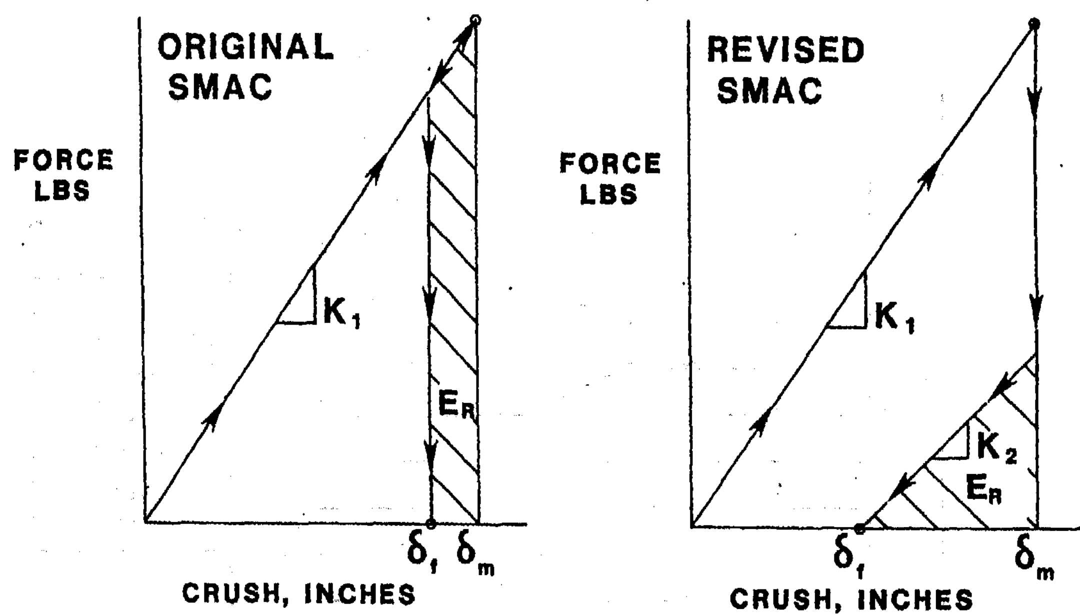


Figure 9 Simulated Force Deflection Characteristics for the original SMAC (EDSMAC) program and the Revised SMAC program

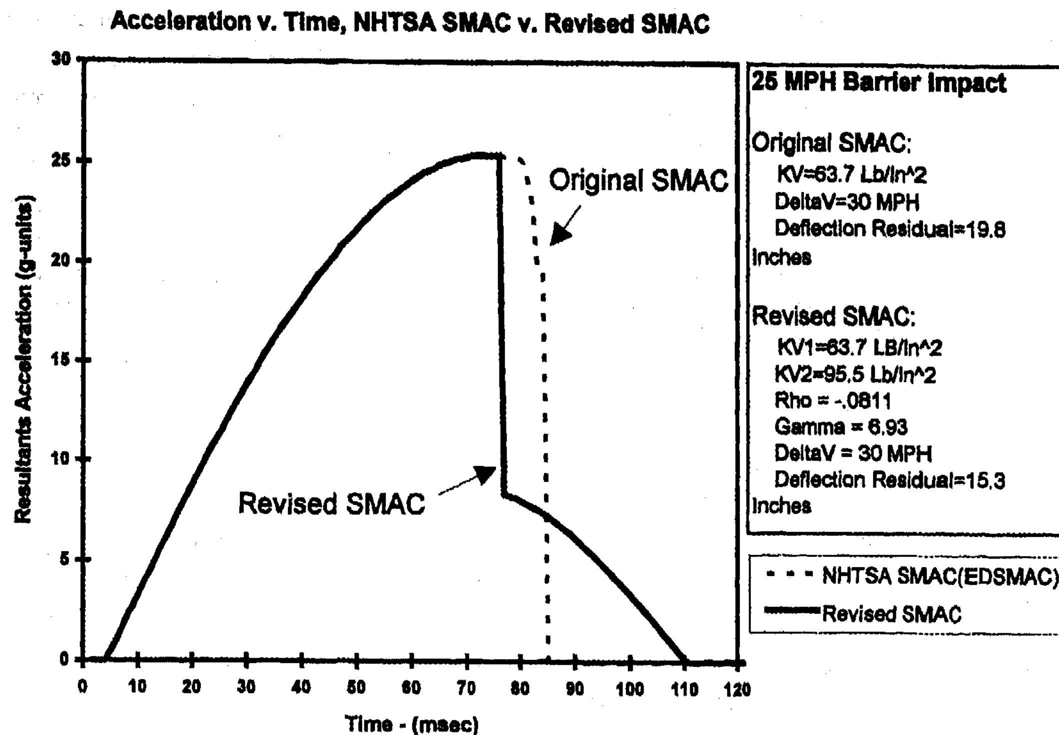


Figure 10 Acceleration v. Time for Original SMAC (EDSMAC) and Revised SMAC

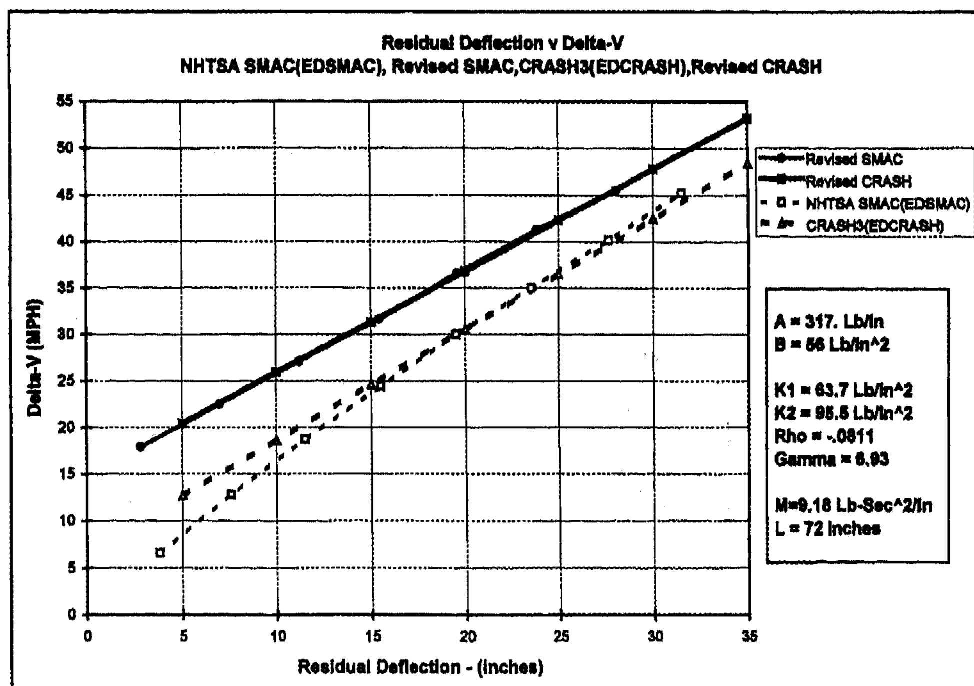


Figure 11 Residual Deflection v. ΔV for CRASH & SMAC program

Comparison of the Mathematical Model of Crush Behavior with Test Data

The lack of exact repeatability of full scale crash test results introduces "scatter" in comparisons of the detailed results of multiple tests. Measurements related to restitution behavior have been found to include an unusually large amount of scatter, particularly at low values of impact speed-changes. The effective coefficient of restitution can include significant effects of forces that are external to the two-body collision system. For example, in an SAE barrier crash, damaged running gear (e.g., jammed or impeded front wheels) can produce significant drag forces that act to reduce the rebound velocity, particularly at low levels of returned energy. Also, energy absorbers on bumpers can act to delay the return of a portion of the absorbed energy.

The mathematical model defined herein addresses the responses of the vehicle structure only. The magnitude of the effects of external forces on individual full-scale tests has not been measured. In the absence of special tests that measure the isolated responses of the structure only (e.g., tests run on casters), it is necessary to focus attention on correlation with the higher returned-energy end of the available measured responses.

While some comprehensive test data from individual tests are available, test series over a range of impact speeds with comprehensive reporting are still relatively rare.

In Figure 12, comprehensive test data from a series of four tests of '79 through '82 Ford LTDs [11] are compared with the fitted mathematical model. In Figure 13, test data for six '81-'85 Ford Escorts [12] which did not include reporting of the ratio of residual to maximum crush, are compared with the fitted mathematical model. In Figure 14, test data from a series of five tests on 1975-1979 VW Rabbits [13] which did not include reporting of the coefficient of restitution are compared with the fitted mathematical model.

Note that the measurements of different aspects of restitution behavior in crash tests that are presented in Figures 12, 13 and 14 include scatter produced at least in part by the previously cited effects.

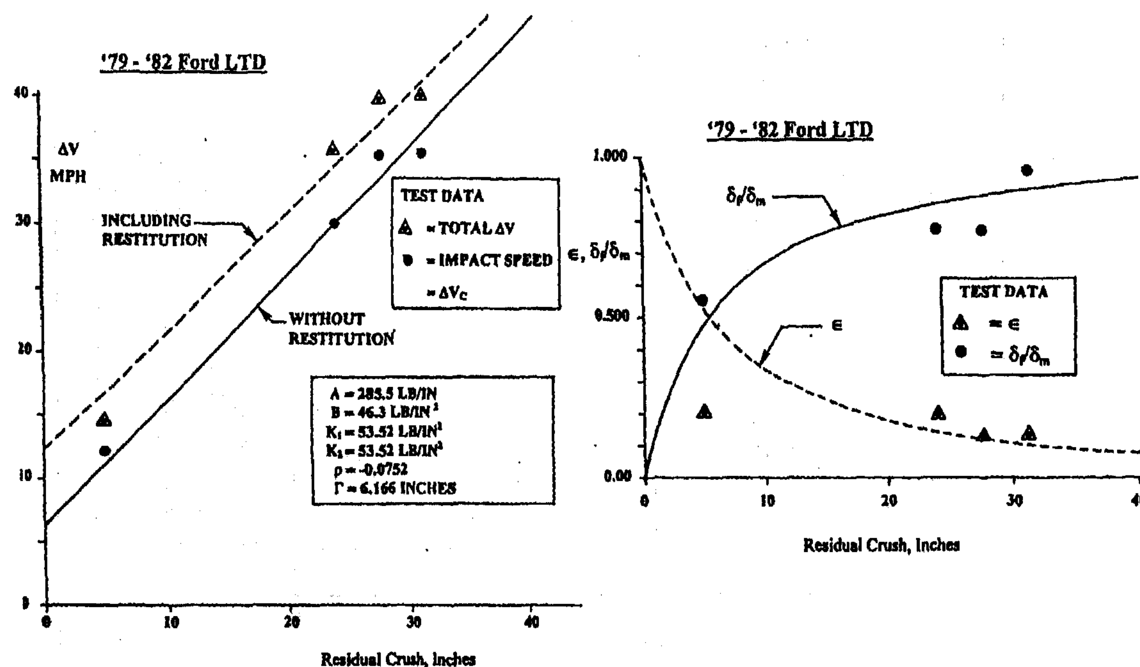


Figure 12 Comparison of Mathematical Model with Ford LTD Tests [11]

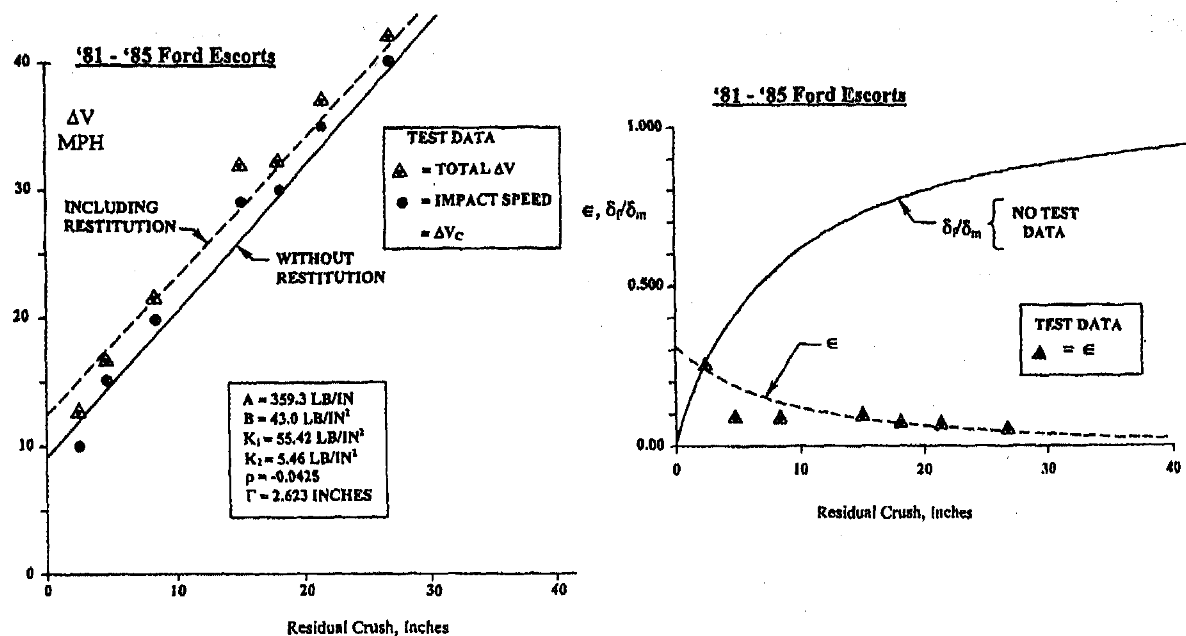


Figure 13 Comparison of Mathematical Model with Ford Escort Tests [12]

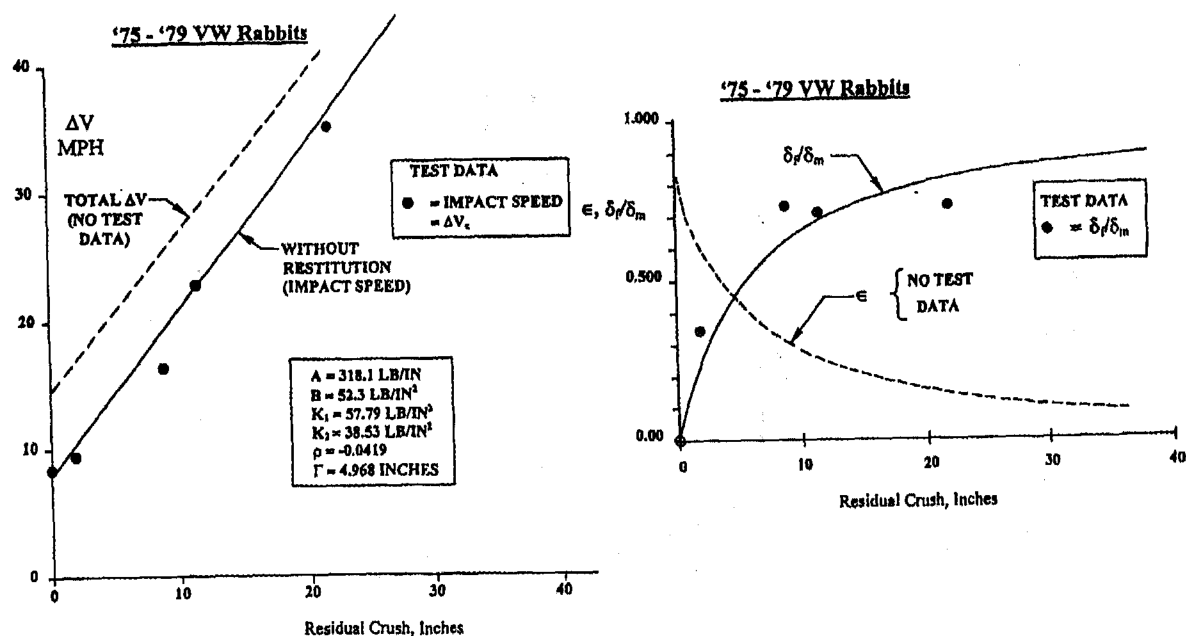


Figure 14 Comparison of Mathematical Model with VW Rabbit Tests [13]

Collisions between Vehicles with Different Restitution Properties

In Figure 15 the results of a 30 MPH head-on mirror image collision, simulated with the revised SMAC program, are displayed. Note that in the simulated collision of Figure 15, the restitution properties of the collision partners are identical. Figure 16 depicts the simulated results with the revised SMAC program of a head-on 27 MPH collision between a high restitution and a low restitution vehicle. Note that the effective restitution is an intermediate value.

In the SMAC implementation, force equilibrium is maintained between the interacting structures of the collision partners at all points within the contact zone throughout the unloading process. By this means, the two vehicles continue to interact during the unloading process until they each reach their residual values of crush at all contact points.

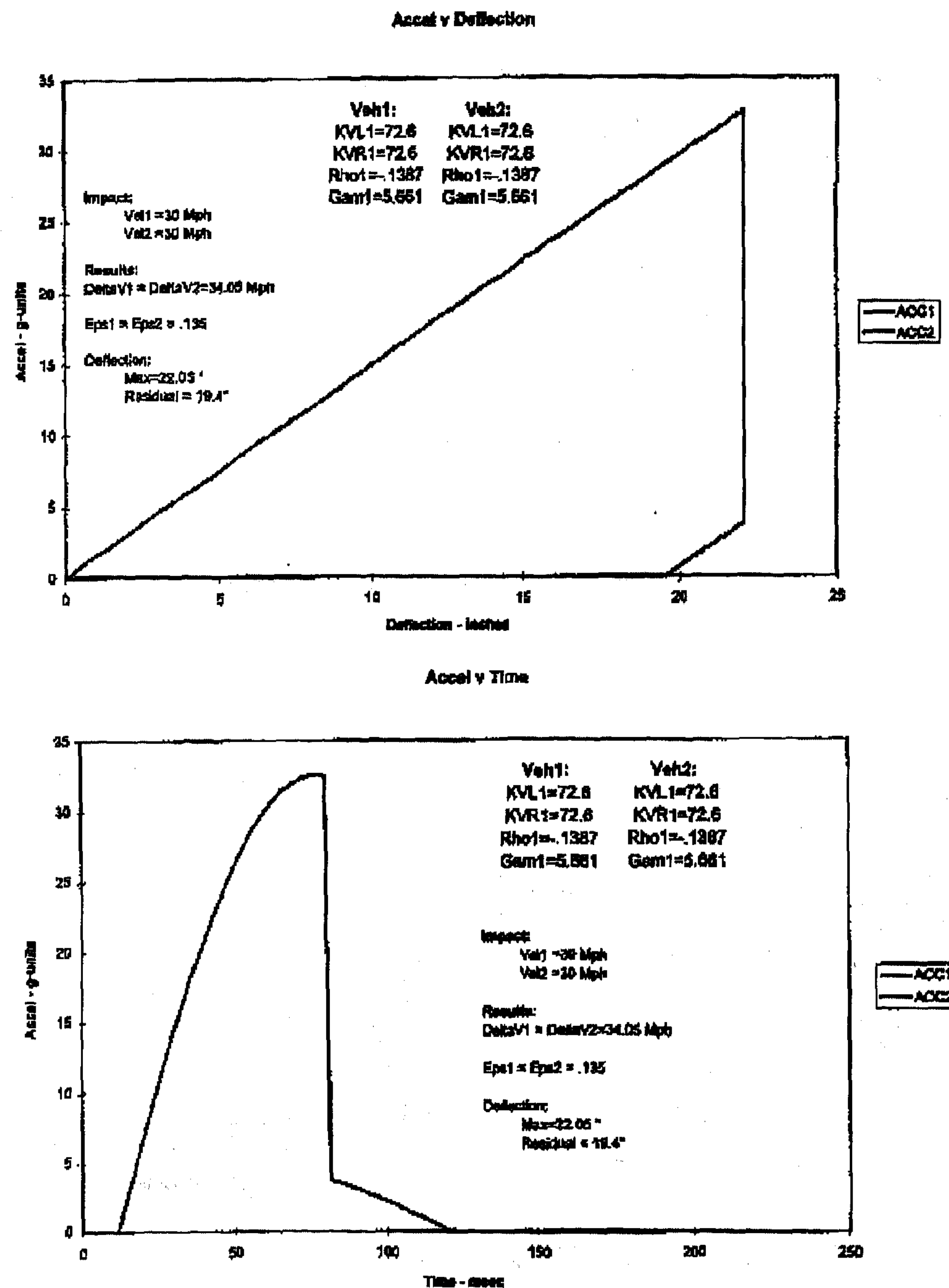


Figure 15 Example Response characteristics of revised SMAC collision routine, 30 MPH Head-On Mirror Image Vehicles

In the CRASH implementation [9], the restored energy for each of the collision partners is separately calculated by means of integrations across the damage interface. The resulting values are added together and then combined with the total absorbed energy for application in the calculation of ΔV_1 and ΔV_2.

In each form of implementation, the effective overall coefficient of restitution in a given collision includes effects of the width and location on each vehicle of the contact area, the detailed damage profiles, and the individual unit-width crush properties of the collision partners. This combination of effects is believed to constitute a realistic analytical representation of the actual physical system during the unloading process.

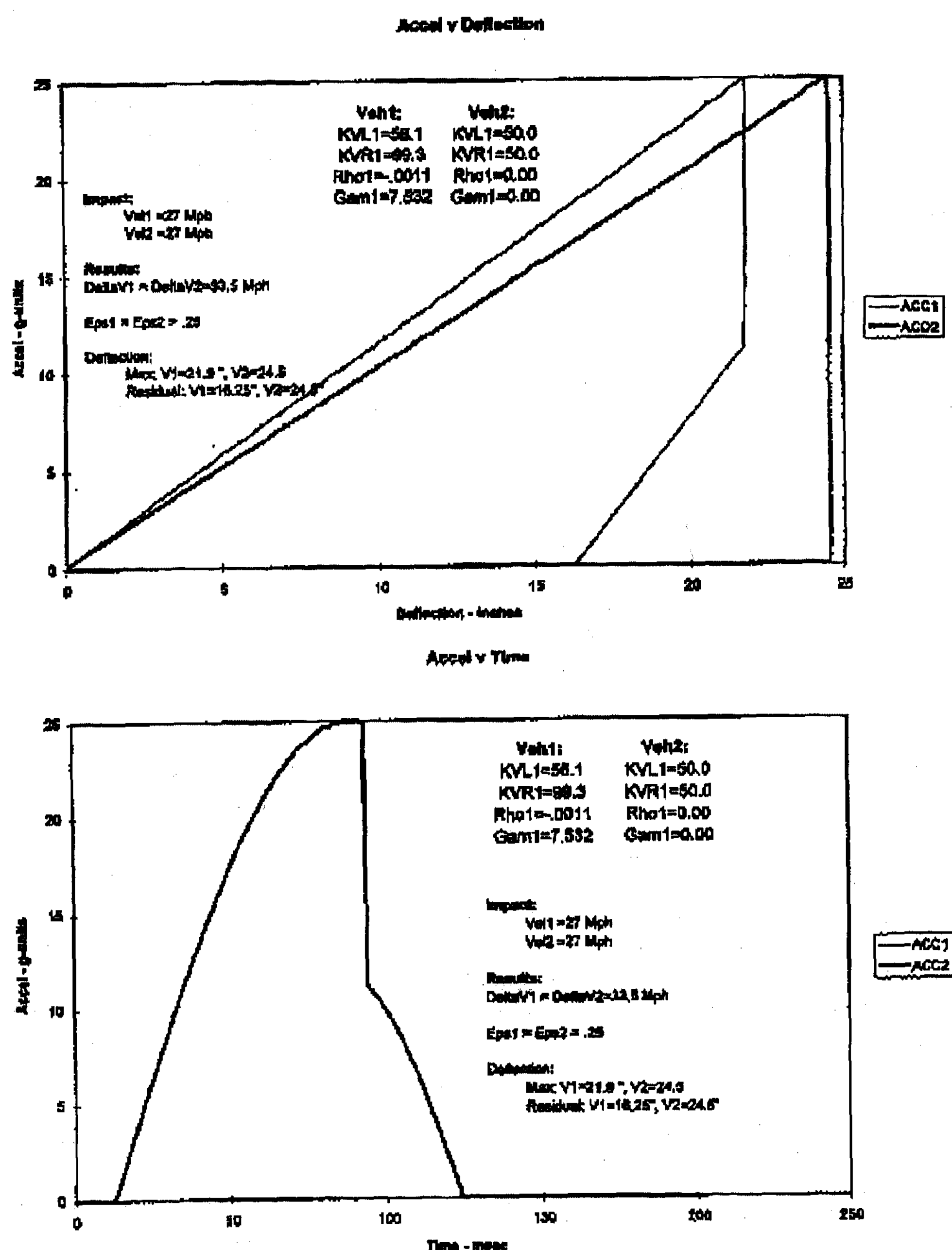


Figure 16 Example Response characteristics of revised SMAC collision routine, 27 MPH Head-On, High and Low Restitution Vehicles.

CONCLUSIONS AND RECOMMENDATIONS

1. Damage analyses which make use of either the original SMAC or CRASH programs produce lower-than-actual values for the impact speed-change, with the maximum errors occurring at low speeds where restitution is greatest. The original CRASH (EDCRASH) does not, of course, include restitution. The simplistic form of simulation of restitution in the original SMAC (EDSMAC), which restores a portion of the crush energy as needed in a time-history solution form, does so in a manner that retains an excessive amount of predicted residual damage. As a result, the error in a purely damage-based ΔV is somewhat greater for SMAC than CRASH at low speeds and somewhat less at high speeds.

2. The developed implementations for both SMAC and CRASH of a revised damage-based procedure that includes restitution have been shown to be capable of producing significant improvements in the accuracy of reconstruction results. They also can serve to insure compatibility of the inputs that define crush properties for the SMAC and the CRASH forms of analysis.

3. The revised damage analysis procedures for CRASH and SMAC provide a unique capability for entering separate definitions of the restitution properties of collision partners.

4. The limited comparisons of the mathematical model of crush behavior with test data that have been possible to date indicate a reasonable degree of correlation. The general form of the mathematical model is dictated by the assumption of a linear relationship between the Delta-V preceding restitution (i.e., to the point of a common velocity) and the residual crush. It should be noted that the specific analytical approach that is defined in Appendices 1 and 2 is not inherently limited to the case of a linear force-deflection characteristic. Modification of the modeled crush behavior to include a saturating force could be readily accomplished if corresponding test data were to become available.

5. The reported research results are considered to constitute an important demonstration of a means of achieving significant improvements in reconstruction accuracy. The restitution aspects of test data needed to fully utilize the described refinements are sometimes included in test reports. They should be made to be a routine part of crash test reports.

REFERENCES

1. Brach, R.M., "Impact Analysis of Two-Vehicle Collisions", SAE Paper No. 83-0468
2. Ishikawa, H., "Impact Center and Restitution Coefficients for Accident Reconstruction", SAE Paper No. 94-0564
3. McHenry, R. R.,"A Computer Program for Reconstruction of Highway Accidents", , SAE Paper No. 730980, Seventeenth Stapp Car Crash Conference, Oklahoma City, Oklahoma, November 12-13, 1973.
4. McHenry, R. R., ,"User's Manual for the CRASH Computer Program", Calspan Report No. ZQ-5708-V-3, Contract No. DOT-HS-5-01124, January 1976.
5. Monk, M. W., Guenther, D.A.,"Update of CRASH2 Computer Model Damage Tables, Volume 1", Vehicle Research and Test Center, National Highway Traffic Safety Administration, East Liberty, Ohio, Report No. DOT-HS-806446, March 1983.
6. Campbell, K.L.,"Energy Basis for Collision Severity", SAE Paper No. 740565, 3rd International Conference on Occupant Protection, Troy, Michigan, July 10-12, 1974.
7. McHenry, R.R., Jones, I.S. and Lynch, J.P.,"Mathematical Reconstruction of Highway Accidents", Calspan Report No. ZQ-5341-V-2, Contract No. DOT-HS-053-3-658, December 1974.
8. Nystrom, G.A., Kost, G., and Werner, S.M., "Stiffness Parameters for Vehicle Collision Analysis", SAE Paper No. 910119, February 1991.
9. McHenry, R.R., McHenry, B.G.,"A Revised Damage Analysis Procedure for the CRASH Computer Program", SAE Paper No. 861894, Thirtieth Stapp Car Crash Conference Proceedings, San Diego, California, October 27-29, 1986.
10. Day, T.D., Hargens, R.L.,"Further Validation of EDSMAC Using the RICSAC Staged Collisions", SAE Paper 90-0102.
11. Ragland, C., "Frontal Impact Protection Program - Frontal Structures: Analysis of Frontal Crash Test Data", NHTSA, Office of Vehicle Research, Sept. 1986.
12. Kerkhoff, J.F., Varat, M.S., et al,"An Investigation into Vehicle Frontal Impact Stiffness, BEV and Repeated Testing for Reconstruction", SAE Paper No. 930899.
13. Navin, F., MacNabb, M. and Miyasaki, G., "Elastic Properties of Selected Vehicles", SAE Paper No. 880223

CONTACT INFORMATION

Questions or comments on the paper are welcomed and can be addressed to the authors:

e-mail: **mchenry@interpath.com**

Postal Service Mail: **103 Brady Court, Suite 200**
Cary, NC 27511 USA

WWW: **http://www.mchenrysoftware.com**

APPENDIX 1: Fitting of Crush Properties to Crash Test Data

Definition of Symbols:

CRASH3 Crush Coefficients, Ref. [5]:

A = Intercept, Lb./Inch

B = Slope, Lb./In^2

Campbell Coefficients, Ref. [6]:

b_o = Intercept, Miles per hour

b_1 = Slope, Miles per hour/inch

E_a = Absorbed energy, inch-lb.

E_r = Restored energy, inch-lb.

Figure A1:

F = Crush resistance force, Lb.

F' = Virtual crush resistance force, Lb.

K_1 = Crush resistance per unit width for increasing loads, lb./in^2

K_2 = Crush resistance per unit width for decreasing loads, lb./in^2

K_v = SMAC crush resistance, Lb./in^2 Ref. [3]

L = Contact width, inches

M = Mass, lb.-sec^2/in

V_c = Common velocity of contact regions at end of approach period of collision, inches/sec

V_o = Initial velocity, inches/sec.

V_f = Final (separation) velocity, inches/sec

ΔV_c = Impact speed-change during the approach period of the collision, inches/sec

ΔV = Total impact speed-change, inches/sec.

δ = Crush, inches.

δ_f = Residual crush, inches.

δ_m = Maximum dynamic crush, inches

Γ = Restitution constant.

ε = Coefficient of restitution

ρ = Restitution constant.

To the extent that the approach-period impact speed change, ΔV_c, can be reasonably approximated as a linear function of the residual crush, δ_f, the corresponding relationship between the coefficient of restitution, ε, and the maximum dynamic crush, δ_m, must be defined in the form:

$$\varepsilon = \frac{\Gamma}{\delta_m} + \rho \qquad (1)$$

$\{$ if $\varepsilon > 1.00$, $\varepsilon = 1.00$ $\}$

where ρ and Γ are fitted constants.

While the fitted A and B coefficients of CRASH are retained to define ΔV_c as a linear function of δ_f, those coefficients are supplemented by measures from crash tests of the restitution behavior. The combined crash test information is applied in the form of a total of four fitted constants:

K_1, K_2, ρ, Γ

The fitting procedure consists of following:

From measured data in crash tests, a test condition is selected at which the restitution behavior is to be matched. For that selected condition the maximum dynamic crush, $(\delta_m)_1$, the ratio of residual crush to maximum dynamic crush, $(\delta_f/\delta_m)_1$, and the corresponding coefficient of restitution, $(\varepsilon)_1$, are determined to supplement A and B. The four fitted constants are then calculated:

$$K_2/K_1 = \left\{\frac{(\varepsilon)_1}{1-\left(\delta_f/\delta_m\right)_1}\right\}^2 \qquad (2)$$

$$\Gamma = \frac{A}{B}\sqrt{\frac{K_2}{K_1}} \qquad (3)$$

$$\rho = (\varepsilon)_1 - \frac{\Gamma}{(\delta_m)_1} \qquad (4)$$

$$K_1 = B\left(1-\rho\sqrt{\frac{K_1}{K_2}}\right)^2 \text{ LB/IN}^2 \qquad (5)$$

$$K_2 = K_1\left(\frac{K_2}{K_1}\right) \text{ LB/IN}^2 \qquad (6)$$

The lack of exact repeatability of crash test results makes it clearly necessary to rely on multiple crash tests for a definition of crash behavior and to deal with the scatter of the results by means of statistical procedures. Measurements related to restitution behavior have been found to include an unusually large amount of scatter, particularly at low levels of returned energy (for example, see comparisons with test data elsewhere in this paper).

In view of the preceding, it is highly desirable to use regression analysis on multiple data points, to insure that the selected test condition is representative of the overall measurements of behavior. In the case of multiple data points, linear regression is first used to define the coefficients of equation (18) of Appendix 2, and, thereby, A and B. Next, the coefficient of δ_m in equation (9) of Appendix 1 is established by means of a fit to the reported values of δ_f and δ_m with an intercept at -A/B. A value for $(\delta_m)_1$ is selected and the corresponding values of $(\delta_f)_1$ and $(\delta_f/\delta_m)_1$ are calculated from the fitted equation (9). The measured value of $(\varepsilon)_1$, corresponding to $(\delta_m)_1$, is established and entered.

As more measurements of restitution behavior become available, it may become possible to establish patterns in the fitted constants for different vehicle types.

The physical significance of the fitted constants may be seen in Figures A1 and A2.

In Figure A1, it should be noted that the peak values of forces, F and F' are not identical. The plot of F' against deflection must be recognized to constitute a virtual force deflection, since the force and the deflection do not exist at the same time. It can be shown analytically that the ratio of the peak forces in Figure A1 is defined by the following:

$$\left(F/F'\right)_{PEAK\ \text{VALUES}} = \sqrt{\frac{K_1}{B}} \qquad (7)$$

The unit-width crush stiffness for increasing loads can also be established from the following relationship, which is independent of the extent of restored energy and, thereby, of the restitution coefficient:

$$K_1 = \frac{1}{B}\left\{\frac{A}{(\delta_m)_1} + B\left(\delta_f/\delta_m\right)_1\right\}^2 \quad \text{LB/IN}^2 \qquad (8)$$

Sample applications of equation (8) are presented in Fig. A3.

Further analytical relationships include the following:

$$\delta_f = \delta_m\sqrt{\frac{K_1}{B}} - \frac{A}{B} \quad \text{INCHES} \qquad (9)$$

$$\delta_f/\delta_m = \frac{\sqrt{K_1/B}}{\left(1+\frac{A}{B\,\delta_f}\right)} \qquad (10)$$

$$\rho = \sqrt{\frac{K_2}{K_1}}\left(1-\sqrt{\frac{K_1}{B}}\right) \qquad (11)$$

$$\varepsilon = \frac{\frac{A}{B}\sqrt{\frac{K_2}{B}}}{\left(\delta_f+\frac{A}{B}\right)} + \rho \qquad (12)$$

The relationship of the four fitted constants to A and B may be further defined:

$$B = \frac{K_1}{\left(1-\rho\sqrt{\frac{K_1}{K_2}}\right)^2} \quad \text{LB / IN}^2 \qquad (13)$$

$$A = B\Gamma\sqrt{\frac{K_1}{K_2}} \quad \text{LB / IN} \qquad (14)$$

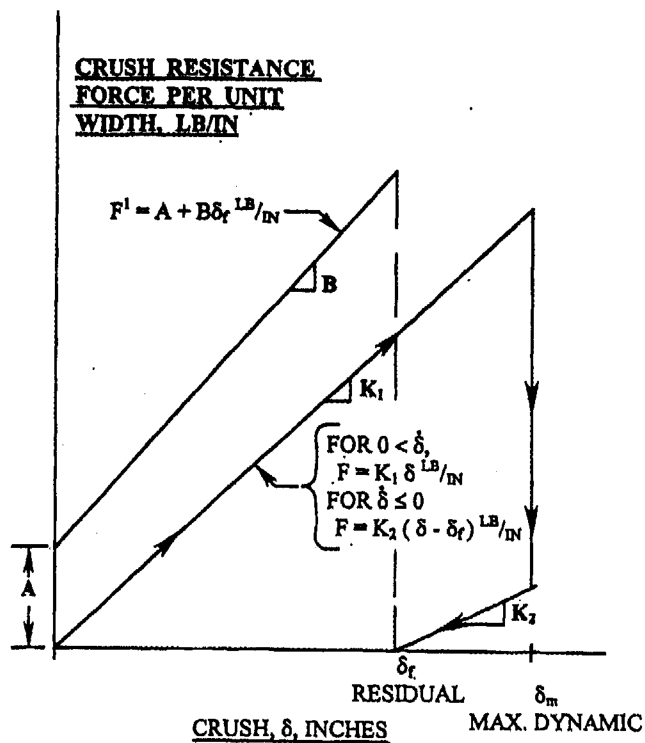


Figure A1: Crush Resistance Force Per Unit Width, Lb/In

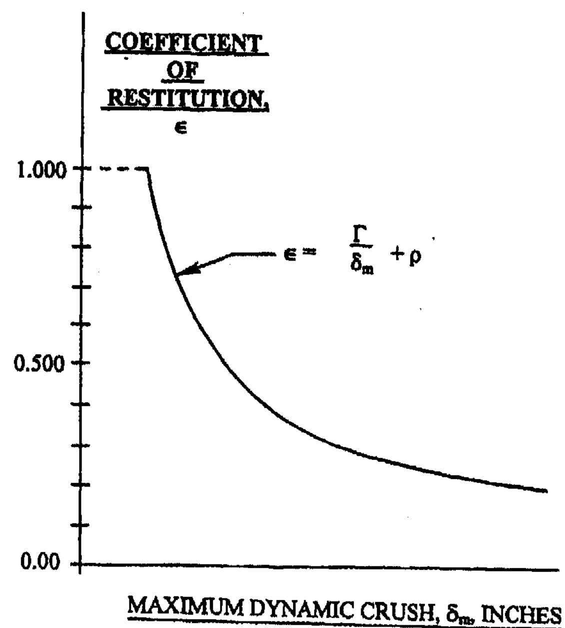
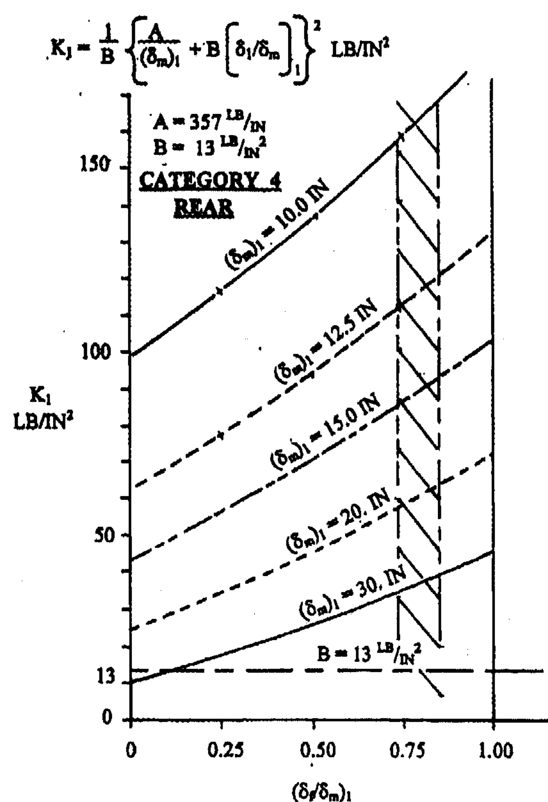


Figure A2: Coefficient of Restitution, ε



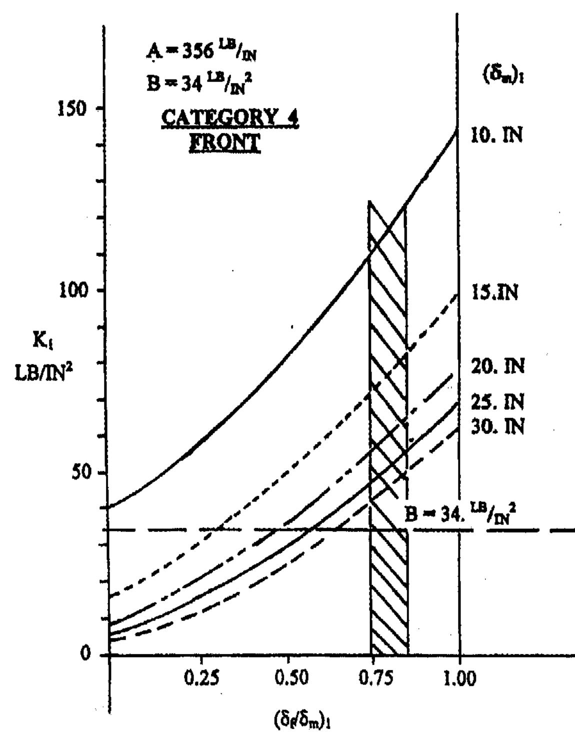


Figure A3: Loading Stiffness vs. Residual/Max Crush Ratio

APPENDIX 2: Derivation of Equations

From Figure A1, the ratio of returned to absorbed energy may be expressed:

$$\frac{E_R}{E_A} = \frac{K_2(\delta_m - \delta_f)^2}{K_1\delta_m^{\,2}} = \frac{K_2}{K_1}\left(1 - \frac{\delta_f}{\delta_m}\right)^2 \quad (15)$$

Since the coefficient of restitution is equal to the square root of the energy ratio,

$$\varepsilon = \sqrt{\frac{E_R}{E_A}} = \left(1 - \frac{\delta_f}{\delta_m}\right)\sqrt{\frac{K_2}{K_1}} \quad (16)$$

Solving (16) for δ_f,

$$\delta_f = \delta_m\left(1 - \varepsilon\sqrt{\frac{K_1}{K_2}}\right) \quad (17)$$

In order for ΔV_c at a given value of residual crush, δ_f, to be equal in the CRASH and SMAC forms of analysis (see Figure A1), the relationship between δ_f and δ_m must be defined by the following two equations:

$$\Delta V_c = A\sqrt{\frac{L}{BM}} + \delta_f\sqrt{\frac{BL}{M}} \quad (18)$$

$$\Delta V_c = \delta_m\sqrt{\frac{K_1 L}{M}} \quad (19)$$

From (18) and (19),

$$\delta_f = \delta_m\sqrt{\frac{K_1}{B}} - \frac{A}{B} \quad (20)$$

$$\frac{d\delta_f}{d\delta_m} = \sqrt{\frac{K_1}{B}} \quad (21)$$

From equation (17),

$$\frac{d\delta_f}{d\delta_m} = \frac{\delta_f}{\delta_m} - \delta_m\sqrt{\frac{K_1}{K_2}}\frac{d\varepsilon}{d\delta_m} \quad (22)$$

$$\frac{d^2\delta_f}{d\delta_m^2} = -\sqrt{\frac{K_1}{K_2}}\left(\delta_m\frac{d^2\varepsilon}{d\delta_m^2} + 2\frac{d\varepsilon}{d\delta_m}\right) \quad (23)$$

Since $\frac{d\delta_f}{d\delta_m}$ = constant, $\frac{d^2\delta_f}{d\delta_m^2} = 0$.

Solution of (23), for $\frac{d^2\delta_f}{d\delta_m^2} = 0$, yields:

$$\varepsilon = \frac{\Gamma}{\delta_m} + \rho \quad (24)$$

From (24), (22), and (21),

$$\frac{\delta_f}{\delta_m} + \frac{\Gamma}{\delta_m}\sqrt{\frac{K_1}{K_2}} = \sqrt{\frac{K_1}{B}} \quad (25)$$

From (20)

$$\frac{\delta_f}{\delta_m} = \sqrt{\frac{K_1}{B}} - \frac{A}{B\delta_m} \quad (26)$$

Substitution of (26) into (25) yields:

$$\Gamma = \frac{A}{B}\sqrt{\frac{K_2}{K_1}} \quad (27)$$

The ratio K_2/K_1 is defined on the basis of equation (16):

$$\frac{K_2}{K_1} = \left(\frac{\varepsilon_1}{1 - (\delta_f/\delta_m)_1}\right)^2 \quad (28)$$

where $(\varepsilon)_1$ and $(\delta_f/\delta_m)_1$ are obtained from crash test data for the specific vehicle for which crush properties are being defined.

From equation (24),

$$\rho = (\varepsilon)_1 - \frac{\Gamma}{(\delta_m)_1} \quad (29)$$

where $(\varepsilon)_1$ and $(\delta_m)_1$ are obtained from crash test data for the specific vehicle for which crush properties are being defined.

From equations (25), (17) and (24):

$$K_1 = B\left\{1 - \rho\sqrt{\frac{K_1}{K_2}}\right\}^2 \quad (30)$$

APPENDIX 3: Clarifications

The review process has revealed several points that may be confusing to some readers. The following is aimed at clarification of those aspects of the paper:

1. The existence of significant errors related to restitution that are present in both the original CRASH (EDCRASH) and original SMAC (EDSMAC) computer programs must be acknowledged. It is common knowledge that the original CRASH (EDCRASH) program underestimates the ΔV in barrier crashes by approximately 10 to 20% at 30 MPH and by a greater amount at lower speeds as a result of the fact that restitution is completely ignored. The original SMAC (EDSMAC) form of simulation of restitution is crude, with rarely changed inputs, and it cannot rationally be expected to produce reliable and accurate combinations of dimensional recovery and partial return of absorbed energy for all vehicles under all collision conditions. In fact, its under-prediction of structural recovery produces a similar range of underestimates of ΔV. Thus, the status quo regarding restitution effects in existing computer programs is very difficult to defend on a logical basis.
2. The ranges of errors (underestimates) indicated in the paper for damage-based ΔV values from the original CRASH and SMAC programs, that are produced either by a total neglect of restitution (CRASH) or by under-prediction of structural recovery (SMAC) are intended to provide the reader with approximate measures of the practical significance of the effects of restitution.
3. On the basis of Figures 2, 6A, and A2 as well as the closely related SAE 861894 it should be clear that ε is set to 1.000 whenever the calculated value exceeds 1.000. Therefore any concern about an "infinite limit" for ε is unfounded.
4. The modeled value of ε, at zero residual crush, is not necessarily equal to 1.000. Rather, it is defined by equation (16):

$$\varepsilon = \sqrt{\frac{E_R}{E_A}} = \left(1 - \frac{\delta_f}{\delta_m}\right)\sqrt{\frac{K_2}{K_1}}$$

   Results of applications of equation (16) to test data where the limiting value of ε was substantially less than 1.00 are shown in Figures 13 and 14. In Figure 13, the restitution coefficient for the Escort never exceeds 0.314. In calculations related to Figure 13 performed by one reviewer the value of ε was incorrectly set equal to 1.00.
5. Energy is absorbed whether or not a structure is elastic. It is the extent of return of the absorbed energy that distinguishes elastic from inelastic behavior.
6. The term "damage" in this paper is used to refer to the generally accepted residual crush, as opposed to any cosmetic disfigurement of "fragile body parts".
7. The general form of the A,B crush coefficients used in the CRASH(EDCRASH) program implies an effective elastic deformation range, in terms of full dimensional recovery, equal to A/B.
8. The lack of exact repeatability of individual measurements in crash tests that are performed under identical test conditions acts to produce ranges of measured responses rather than single values. For this reason, rigorous measures of reconstruction accuracy must be based on the mean experimental measurements for a given set of test conditions.
9. Available experimental measurements of restitution behavior over ranges of impact speed (e.g., Figures 12, 13, and 14) include only single measured values at the individual test conditions. Clearly, progress toward a rigorous and complete validation study is data-limited at the present time.

SAE #970960
Reviewer Discussion
by Raymond M. Brach, University of Notre Dame
Effects of Restitution in the Application of Crush Coefficients
Raymond G. McHenry, Brian G. McHenry, Authors

Many methods used by accident reconstructionists to estimate collision speeds assume that vehicle structures are perfectly plastic with insignificant rebound. For some collision configurations and for low speed collisions, this is not true. The authors present an approach to model the way in which restitution takes place within a vehicle structure during a collision. Restitution is viewed as the process of restoration of kinetic energy from strain energy following the maximum normal deformation. The model takes into account the variation of restitution with the severity of a collision by modeling it as a function of the maximum crush, using an equation with constants determined by crash test data. Restitution varies from location to location over a vehicle's periphery and varies from vehicle to vehicle. These variations are taken into account through suitable selection of test conditions. The model averages the restitution over the lateral extent of crush and ultimately yields refined values of crush coefficients for SMAC and CRASH.

The approach is interesting and forms a workable model of an extremely complex problem. Before being adapted for widespread use, the accuracy of the method needs to be verified more extensively than done in the paper, especially for low speed collisions. Additional issues may need to be investigated such as differences in restitution between car-to-barrier and car-to-car collisions. Another is the (potentially large) amount of experimental data needed to completely characterize a single vehicle.

Readers should note that the coefficient of restitution used in this paper is not the same as the one used in the impulse and momentum analysis of vehicle collisions. In this paper it is used to model the local behavior of the *structure*; in collision analysis it is used to represent the total energy loss of a *collision.*

SAE #970960
Reviewer's Discussion
by James A. Neptune, J_2 Engineering, Inc.
Effects of Restitution in the Application of Crush Coefficients
Raymond R. McHenry, Author

The authors' revised model for restitution takes a step in the right direction. The revised model, however, is not entirely realistic. As a result, it may not correctly represent a vehicle's crush response characteristics.

The revised model for restitution in figure 6A is a completely elastic crush response (ε=1.0) through the first several inches of dynamic crush. Completely elastic dynamic crush is a theoretical condition. In the real world, restitution may approach the theoretical magnitude of 1.0 but will never equal 1.0. Restitution is a mathematical "limit." The limit of the magnitude of restitution is 1.0 as dynamic crush approaches zero. Therefore, it is not realistic for the first several inches of dynamic crush to be completely elastic. In addition, the revised equation for restitution (appendix 1, equation 1) is an "infinite limit," or asymptote, where restitution (ε) becomes infinite as dynamic crush (δ_m) approaches zero. This equation is not realistic. If this model is implemented in a computer program, hopefully the program will provide the necessary limit not present in equation (1).

The crush response of the high restitution hypothetical vehicle in figure 2 is completely elastic through the first 6.2 inches of crush. A completely elastic crush response (ε=1.0) through the first 6.2 inches of crush is not realistic. Clearly front bumpers, with rare exception, do not extend 6.2 inches beyond easily damaged vehicle components such as the fenders and hood (note: any residual damage means $\varepsilon \neq 1.0$). Also "energy absorbing" bumpers that are present on vehicles "absorb energy" and, therefore, are not completely elastic. In addition, the vehicle in figure 2 would begin to develop residual crush (ε<1.0) after the first 5.6 inches of dynamic crush. This means that, between 5.6 inches and 6.2 inches of dynamic crush, the vehicle's crush response is partially plastic (ε<1.0) yet at the same time is completely elastic (ε=1.0). This is impossible.

Considering the above, the conclusions and graphs that are drawn from analyzing the hypothetical vehicles are subject to question (Analytical Approach section, figures 4 & 5). If the revised restitution model can produce a vehicle that could not exist in the real-world, can one be certain that the revised coefficients will correctly represent a vehicle's crush response characteristics when applied to a real-world vehicle?

Figure 2 Calculations:

Use equation (1), set ε=1.0 and solve for δ_m. $$\delta_m = \frac{6.2897}{1.0 + 0.0097} = 6.2 \text{ inches.}$$

Use $K_1 = \frac{4102}{72} = 57$ lb/in^2 & equation (9), set δ_f =0 and solve for δ_m. $$\delta_m = \frac{317}{56}\sqrt{\frac{56}{57}} = 5.6 \text{ inches.}$$

Or set equation (9)=equation (16), set δ_f =0 and solve for δ_m. $$\delta_m = \frac{6.2897}{\sqrt{\frac{5065}{4102} + 0.0097}} = 5.6 \text{ inches.}$$

The revised coefficients for 1979-82 Ford LTD's (figure 12) indicate that the first 5.7 inches of dynamic crush is completely elastic with no energy being absorbed by the vehicle's structure [use equation (1), set ε=1.0 and solve for δ_m]. Does this correctly represent this vehicle's crush response characteristics?

In figure 13 the revised coefficients for 1981-85 Ford Escort's would indicate the first 2.5 inches of dynamic crush is completely elastic [use equation (1), set ε=1.0 and solve for δ_m] and that a dynamic crush of 7.4 inches would occur before the onset of residual crush [use equation (9), set δ_f =0 and solve for δ_m]. Can this vehicle's front end structure flex rearward through 7.4 inches with no residual damage? Wouldn't the easily damaged vehicle parts such as the fenders and hood sustain significant damaged before 7.4 inches of dynamic crush occurs? Does this correctly represent this vehicle's crush response characteristics? [Note: Equations (1) & (16) can be used to evaluate the crush-restitution relationship in the range of crush that occurs after the onset of residual crush (δ_f >0). Equation (16), however, can not be used for the initial elastic-only portion of dynamic crush (before the onset of residual crush, δ_f) since the ratio δ_f/δ_m is zero. Equation (1), therefore, must be used to evaluate the crush-restitution relationship for the dynamic crush that occurs before the onset of residual crush.]

MacInnis Engineering (SAE Paper 940916) has tested a 1981 Ford Escort. Full stroke of the front bumper isolators is 57 mm (2.25 inches). Residual crush began at a dynamic crush of 55 mm (2.17 inches). Maximum restitution of 0.61 occurred at 0.5 mm of isolator stroking (dynamic crush). The MacInnis data contradicts the revised restitution model. Restitution is not equal to a constant value of 1.0 over the first 2.5 inches of dynamic crush. In addition, the onset of residual crush occurs at a dynamic crush of 2.17 inches rather than 7.4 inches. Analysis of the revised coefficients for the 1975-79 VW Rabbit (figure 14) when compared to data from SAE Paper 940916 produce similar contradictions.

The original CRASH/SMAC programs and the revised CRASH/SMAC programs were used to estimate ΔV for a hypothetical vehicle. The authors characterize the differences (10 to 30%) between the estimated ΔV's of the original versus revised CRASH/SMAC programs as "errors." Differences between the estimated ΔV's would be more accurately characterized as "potential improvements in accuracy" rather than "errors." A characterization of "error" should be limited to comparisons between estimated ΔV's and the actual ΔV's measured during the thorough testing of a real-world vehicle.

In the Introduction section of the paper, the authors state that the original SMAC program inaccurately estimates ΔV approximately 10 to 30% for the case of direct central barrier collisions. There is no reference to the source of this information/conclusion. Hopefully this conclusion is not based upon the above "error" analysis involving hypothetical vehicles. Also in appendix 3, the authors indicate that it is "common knowledge" that the original CRASH program "underestimates the ΔV in barrier crashes by approximately 10% to 20% at 30 MPH and by greater amounts at lower speeds." Stating this is "common knowledge" should not relieve the authors of the responsibility of providing the supporting evidence for this conclusion. The lack of supporting evidence raises questions regarding the accuracy of these numbers.

The authors' revised model for restitution, although a step in the right direction, is not completely realistic. The restitution model needs to be reconfigured such that restitution is less than, or equal to, the theoretical limit of 1.0 at zero dynamic crush ($\varepsilon_0 \leq 1.0$) and decreases from that point as dynamic crush increases.

970961

RICSAC-97 - A Reevaluation of the Reference Set of Full Scale Crash Tests

Brian G. McHenry and Raymond R. McHenry
McHenry Consultants, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Research performed in the 1970's revealed significant limitations in the available documentation of vehicle crush information and trajectory spinout information. As a result a series of full-scale crash tests were performed which became known as the Research Input for Computer Simulation of Automobile Collisions (RICSAC) crash tests.

Previous research using the RICSAC test results, particularly in relation to the validation of accident reconstruction computer programs, has varied widely in acceptance, interpretation and presentation of the RICSAC test results.

This paper presents a detailed review and decipherment in useable form of the original 12 crash tests that were performed within the RICSAC program. A new method of analyzing accelerometer data from arbitrary sensor positions, on the basis of discrete measures of the vehicle responses rather than complete time-histories, is defined. A discussion of previous research which included reference to the RICSAC test results as a measure of the validity of reconstruction computer programs is included.

INTRODUCTION

Research performed in the 1970's to "locate, review, decipher and place in useable form available experimental data on the structural crush properties of automobiles, and on the spinout trajectories produced by measured collision conditions" [1][1] revealed significant limitations in the available documentation of vehicle crush information and trajectory spinout information. As a result, an initial test matrix of 12 full-scale crash tests was performed in 1978 which became known as the Research Input for Computer Simulation of Automobile Collisions (RICSAC) crash tests [2].

For each of the RICSAC tests, a minimum of 13 accelerometers were mounted on-board each vehicle to record acceleration components at six to seven stations. At three locations triaxial (XYZ) packages were installed ("hard mounted") to provide coverage between the front and rear of the vehicle. The front steer angles were measured on each vehicle by a linear stroke potentiometer attached to the vehicle steering linkage. The time history of the change in vehicle yaw, pitch and roll angles and yaw rate were recorded by two-degree-of-freedom, free gyroscopes and a rate gyro.

For vehicle trajectory measurement, high-speed motion picture cameras were located overhead with supporting cameras at eye level height. Also, vehicle trajectory marker systems were designed and fabricated by Calspan to spray high pressure colored water trails (using very small amounts of liquid to avoid any effects of the water on the roadway coefficient) from several locations on each of the vehicles.

Plots of the full-scale test acceleration time-histories were generated by a computer program which also integrated the acceleration time histories (using a combination of Simpson's and Newton's 3/8 methods) in order to obtain velocity time history along each of the vehicle's three axes. For most of the reported results, the ΔV values reported for the vehicle CG were based on data collected by a Firewall mounted triaxial (XYZ) accelerometer.

The time of approximate separation reported in [3] was based on observation of the time histories of the acceleration. The time of separation was defined as "the point at which both of the involved vehicles' accelerations reapproached 0 g's". The authors of [3] noted that "the separation point was not always obvious in collisions in which the two vehicles spun out together".

The components of the vehicle's change in velocity (ΔV) were computed in [3] "by subtracting the initial velocity at impact from the velocity at the time of separation". The procedure was done for the X and Y velocity components separately.

The authors of [3] acknowledged that the "value for the separation velocity was contaminated by the effects of rotation of the vehicles between impact and separation". The problem was attributed to the post-processing integration software used for the tests which was primarily set up to integrate acceleration data for frontal collisions with barriers

1 numbers in brackets [] indicate references at end of paper

and therefore did not account for the rotation of the vehicle (the data reduction software assumed a constant direction cosine matrix). Errors which may occur in data collected from an accelerometer not at the center of gravity include the effects of rotation which are dependent on the magnitude of the offset of the accelerometer from the CG and the magnitudes of the vehicle's angular velocity and acceleration.

Table 1 contains best estimates of the actual accelerometer locations for the 12 RICSAC tests based on values reported by Jones [3]. Some of the locations have been corrected on the basis of a detailed review of the RICSAC test reports [2].

Table 2 contains a list of factors as reported in [3] which influenced the magnitude of effects of rotation on the RICSAC accelerometer data. Note that any differences in the accelerometer locations as listed in Tables 1 and 2 are due to the refinement in **Table 1** of some ambiguous dimensions reported in [2] and/or [3].

Table 1 **Firewall and Rear Deck Accelerometer Locations for RICSAC Tests (Note:Test 4(S&N 6), Veh 2 Front Deck)**

RICSAC [2,3] Test No.	S&N [4] Test No.	Vehicle No. 1 Firewall x (in)	Vehicle No. 1 Firewall y (in)	Vehicle No. 1 Rear Deck x (in)	Vehicle No. 1 Rear Deck y (in)	Vehicle No. 2 Firewall x (in)	Vehicle No. 2 Firewall y (in)	Vehicle No. 2 Rear Deck x (in)	Vehicle No. 2 Rear Deck y (in)
1	1	6.9	-9	-88	0	13.5	-7	-73.5	5
2	2	8.5	-9	-86.6	0	14.6	-7	-72.5	5
3	5	15.7	0	-81.3	0	5.2	-7	-62.8	4
4	6	15.9	0	-81.1	0	3.2	-7	33.7	18
5	7	15.6	0	-81.4	0	13.1	4	-40	5
6	3	11.5	-9	-83.5	0	29	-7	-58	5
7	4	11.5	-9	-83.5	0	28.3	-7	-58.7	5
8	10	11.5	-9	-92.5	0	11.5	-10	-79.5	5
9	11	10.2	-9	-50.2	0	14.1	-7	-59.7	0
10	12	9.4	-9	-69.6	0	14.9	-7	-83.1	0
11	8	15.5	11.5	-44.4	0	14.4	1.5	-63.6	0
12	9	13.4	11.5	-48	0	15.6	1.5	-62.4	0

Table 2 **From Jones [3] Table 1-3 "Factors Influencing Magnitude of Effects of Rotation" on Accelerometer data. (Note: Contact Duration from Table 1-1 of [3])**

RICSAC [2,3] Test No.	S&N [4] Test No.	Veh No.	Heading angle Change (deg)	Angular Velocity Max (deg/sec)	Angular Velocity @Separation (deg/sec)	Distance cg to Firewall Accelerometer r_x (in)	Distance cg to Firewall Accelerometer r_y (in)	Contact Duration* (sec)
1	1	1	15	120	90	6.8	-6.9	0.225
		2	0	0	0	13.4	-6.2	
2	2	1	18	150	150	8.4	-9.6	0.225
		2	-9	-120	90	15.2	-5.5	
3	5	1	1	20	15	15.7	-0.1	0.2
		2	0	0	0	5.9	-6.8	
4	6	1	lost data	45	37	15.7	-0.1	0.275
		2	lost data	30	30	4.2	-6.2	
5	7	1	5	20	12	-15.8	-0.4	0.25
		2	lost scale factor	90	70	10.6	4.0	
6	3	1	5	45	30	6.5	-8.8	0.2
		2	20	210	180	30.4	-4.4	
7	4	1	12	65	30	6.5	-9.2	0.2
		2	22	210	192	29.8	-4.6	
8	10	1	15	135	114	11.4	-9.0	0.2
		2	0	18	18	11.2	-6.8	
9	11	1	27	210	180	10.6	-8.9	0.2
		2	-10	-45	-45	14.5	-6.7	
10	12	1	55	300	300	9.4	-8.5	0.2
		2	-12	-90	-72	14.9	-5.9	
11	8	1	-5	-45	-30	19.3	12.2	0.225
		2	0	0	0	14.4	2.8	
12	9	1	-10	-90	-90	12.4	12.1	0.225
		2	-2	-60	-60	15.6	1.9	

BACKGROUND

In 1980, Smith and Noga [4] created a summary of data associated with 16 full-scale crash tests (three more tests were performed as a part of RICSAC and reported in [5] and the other was from [6]). The authors "attempted to achieve corrections to acceleration and velocity data by approximating vehicle rotational velocity and rotation of the reference frame from it's orientation at impact, by making reference to multiple accelerometer outputs at different locations on the vehicles and by analysis of high speed photography." They also included an adjustment procedure for the ΔV's to insure that the momentum changes associated with the collision partners were equal and opposite (including consideration of tire forces). The adjustments were approximations to account for the effects of external tire-forces. They found that "cases number 10, 12 and 15 are believed to be in error" due to poor accelerometer data for one of the two involved vehicles. The determination of the separation velocity was not refined or presented.

In 1983, Brach [7] presented a table of refinements to the reported RICSAC separation velocities on the basis that: "Accelerometers used to record data on each vehicle were not located at the vehicle's center of gravity. Consequently, the final velocity data had to be corrected by the angular velocity at separation and the distance of the accelerometer from the mass center." The report does not include the specific equations and the values for distances of the accelerometers from the CG and other variables used in the adjustment process (The values from Jones reported in **Table 2** were the values probably used).

In 1985, Wooley, as a part of an attack on the CRASH3 analytical procedures [8] and as part of a presentation and testing of the "IMPAC" computer program [9], included a review and analysis of the RICSAC test results as reported by Smith and Noga [4]. In addition to agreeing with Smith & Noga that there were problems with tests 10, 12 and 15, tests 4 and 6 were identified as having either "a typing error or reporting error." A "Data Integrity Ranking" was calculated for the tests based on Newton's 3rd Law for equal and opposite forces and the law of conservation of linear momentum. The tests utilized the post-impact velocities (the separation velocities) which were the values as reported by Jones [3]. The testing procedures of [8] and [9] cited many instances where the "Data Integrity Ranking" calculations revealed cases with erroneous momentum *gains* and/or instances where the forces acting on the collision partners where not equal and opposite.

Also included in the reported results of [4] was a "velocity change" angle. The "velocity change" angle was calculated by finding the angle with tangent of the Δv (Y component) velocity change divided by the Δu (X component) velocity change. In [8], the "velocity change" angle reported in [4] was referred to as the Principal Direction of Force (PDOF).

IN 1989, as a part of a presentation of a comparison of EDCRASH program results with the RICSAC test results Day [10] concludes that "No general observations could be made using the RICSAC data analyzing the accuracy of the {computed} ΔV by any program because of measurement errors in the RICSAC test data." The report also states that "Previous program evaluations which used the RICSAC data as a means of validation for ΔV should be viewed as suspect."

Also in 1989, Cheng and Guenther [11] discussed the localized variations in speed-change (ΔV) due to rotational effects as applied to the speed-change experienced by the vehicle occupants. The paper included a discussion of the general differences in measurements produced by the various accelerometer locations in the RICSAC tests.

IN 1990, Day [12] as a part of a presentation of a comparison of the EDSMAC program results with the RICSAC test results included a Table of ΔV results for RICSAC which were obtained from [1-3] and "from personal correspondence from Dr. Russell A. Smith.". Some of the reported values are different from Smith and Noga [4] and may reflect refinements by Smith and Noga reported in [13]. The report conclusions include the observation: "The measured ΔVs were not of acceptable accuracy for use in a validation study because the motion transducers were placed at the vehicle's firewall, rather than the CG. This problem might be improved or eliminated by re-analyzing the original data with software which included a transformation matrix."

In 1996, as part of a validation effort of the PC-CRASH computer program, Cliff [14] used the RICSAC test results for the post-impact separation speeds as reported by Brach [7] for comparison "validation" purposes. They noted that "there appeared to be some error in a few of the reported post-impact speeds in [the Brach] paper when compared to the reported pre-impact speeds and vehicle weights."

Also in 1996, Bundorf [15] presented techniques for analysis of the CG ΔV from accelerometer data at locations other than the CG. He observed that "the potential for error is always present if yaw velocity is high and accelerometers are distant from the center of gravity." As an example, he presented a detailed analysis of RICSAC Test no. 9 (S&N Case 11).

RESEARCH APPROACH

The RICSAC tests [1-3] contain the most comprehensive collection of full-scale test results available to date. The test reports include objective information on the impact speeds, vehicle weights, dimensions, weight distributions, spinout trajectory and positions of rest. Some interpretation of the reported results is required, for example, to obtain speed-change (ΔV) and separation velocities from the accelerometer data. Also, some evaluations are required for the approximate extent of wheel drag and steer angles. The primary purpose of the research associated with this paper has been to develop and apply techniques for

interpretation of the crash phase of RICSAC and other full-scale tests. Of particular interest are the correct values for the impact speed-changes and separation velocities effective at the centers of gravity, which can be used as an obstacle course to evaluate and/or "validate" accident reconstruction techniques.

During full-scale testing, a hard-mounted accelerometer at the CG will record the correct acceleration *including* the effects of rotation. The time integral of the accelerometer data for the X and Y directions will produce the correct vehicle-fixed X and Y components of speed-changes. However, in the RICSAC tests and in many other full-scale tests, the ΔV calculations are complicated by accelerometer data for locations other than the vehicle center of gravity. The first task of interpretation was to develop generalized analytical techniques to transform the speed-change information from arbitrary accelerometer locations to the center of gravity. A secondary task was to use the calculated CG speed-change information to calculate the vehicle separation velocities. To correctly use the CG speed-change information to calculate the separation velocities, provisions must be included for the changes in the vehicle orientation.

The technique used by Jones [3] for calculation of the separation speeds for the RICSAC tests is in error, primarily due to the direct subtraction of the speed-change determined by integration of accelerometer data from the initial impact speed. The u and v speed-changes computed by integration of the accelerometer X and Y data cannot be directly subtracted from the initial vehicle velocities except for the case where no rotation occurs. A secondary source of error in the calculation of separation velocities reported in [3] is the assumption that the acceleration and velocity data from the firewall location are equivalent to those at the vehicle center of gravity.

CG-TRANSFORM

To provide a simple approximation of the RICSAC ΔV and separation velocities at the vehicle centers of gravity, an analytical procedure was developed for transformation of accelerometer data from accelerometers not located at the vehicle center of gravity on the basis of discrete measures of the responses rather than complete time-histories.

Bundorf [15] presented a procedure for determination of the speed-change (ΔV) at the vehicle center of gravity (CG) using accelerometer time histories from accelerometers not located at the vehicle CG combined with rate gyro data. The technique requires either the availability of the FM data recorder tapes or a manual digitization of the accelerometer and yaw velocity traces to determine the integral of yaw velocity squared and, thereby, the speed-change (ΔV). A drawback of the procedure is that it can be time consuming for the case where there are not FM data tapes and/or processing software available. Also, possible sources of errors can occur from integrating and synchronizing randomly sampled data (the accelerations were recorded with triaxial accelerometers, the angles and yaw rates were recorded with a pair of two-degree-of-freedom, free gyroscopes and a rate gyro).

The ***CG-Transform*** analytical procedure has input requirements listed in Table 3 (for a detailed definition of ***CG-Transform*** please see **Appendix 1**).

Table 3 Information required for *CG-Transform* analytical procedure

1. Accelerometer Location on vehicle, X, Y.
2. Impact speed and heading angle.
3. Integrated speed-change from accelerometer data.
4. Duration of crash pulse.
5. Separation angular (yaw) velocity and angle.

To permit initial testing and refinement of the ***CG-Transform***, the SMAC computer program was utilized to create a set of mathematically correct "test case" reconstructions. Extensions to the SMAC program developed for the present research include a provision to monitor the acceleration for any arbitrary position on the vehicle (e.g., **Figure 1**, **Figure 2**). ***CG-Transform*** was then used to transform the data from an arbitrary accelerometer location back to the CG. Since the CG data is a direct output of the SMAC program the procedure provided a detailed check of ***CG-Transform.***

The SMAC reconstructions used for this purpose were generalized reconstructions of the RICSAC test runs. The SMAC runs were based on preliminary refinements of the original RICSAC SMAC reconstructions presented by Jones [3] and more recently Day [16] in validation studies of the SMAC and EDSMAC computer programs. The refinements of some of the inputs were required due to the questionable nature of some of the SMAC inputs used in the "validation" reconstructions (e.g., some torque and steer values were arbitrarily varied). A more detailed discussion of the rationale for the changes to the SMAC inputs and suggestions for refinement of the inputs are to be presented as part of a separate research project [17].

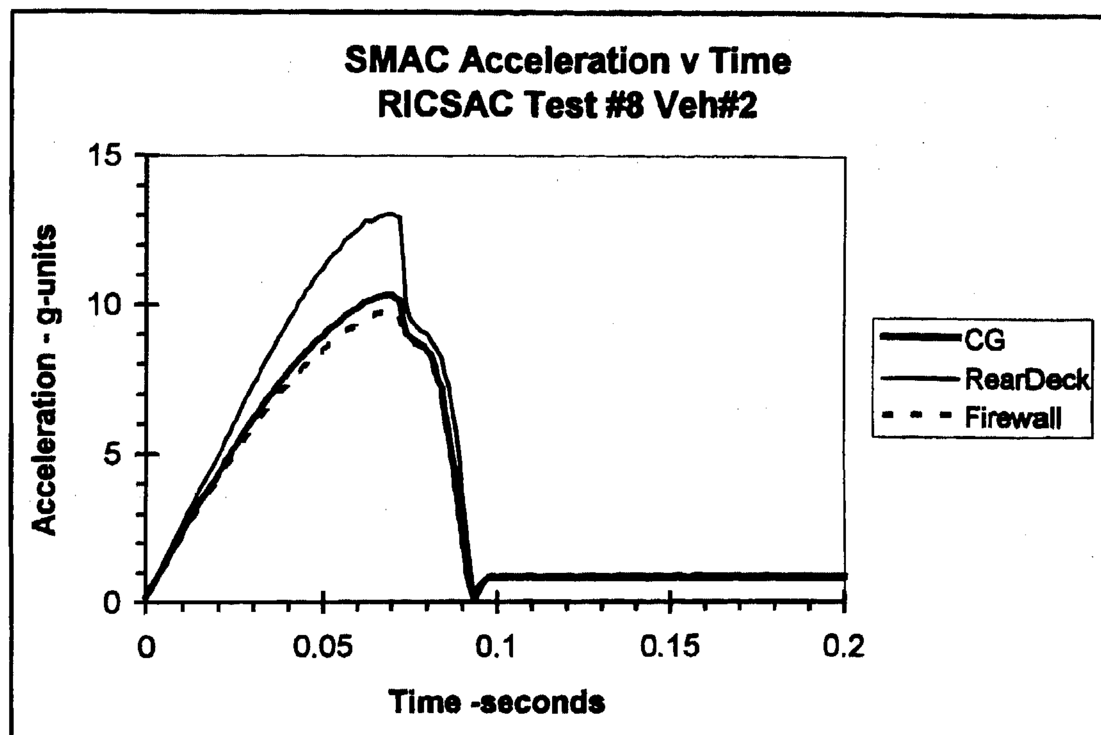


Figure 1 SMAC resultant acceleration time history at CG, Rear Deck and Firewall for RICSAC test No. 8, Vehicle No. 2

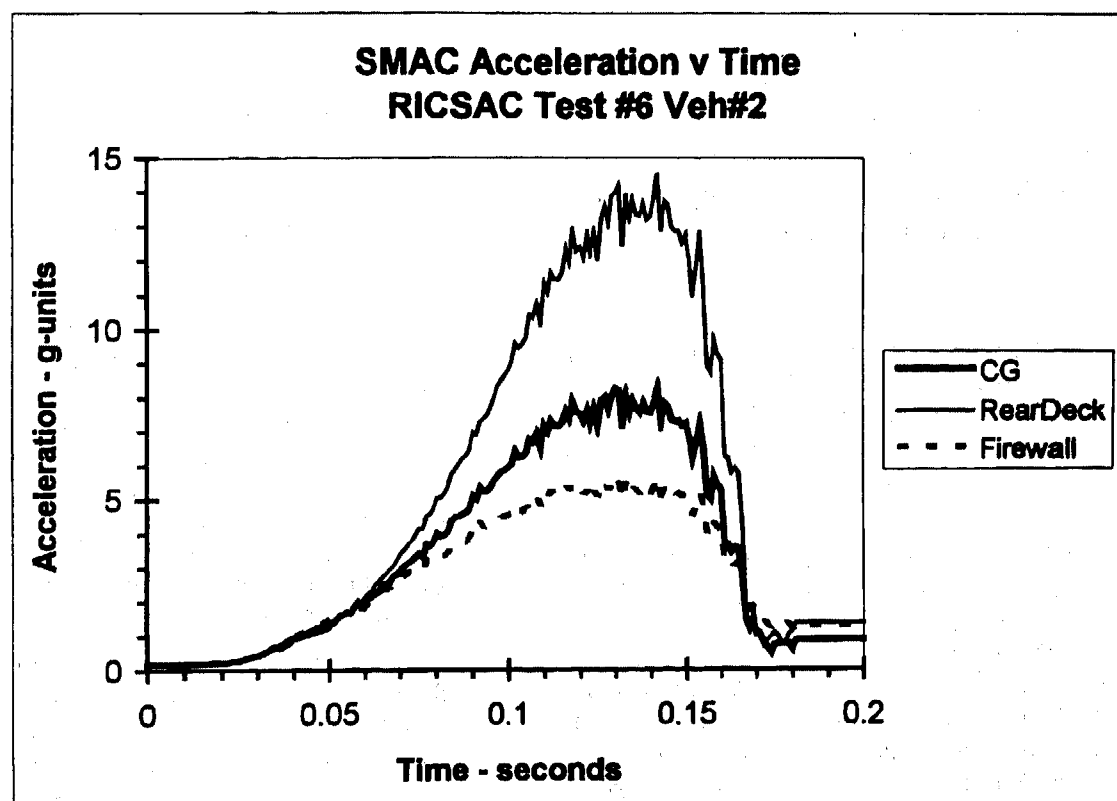


Figure 2 SMAC resultant acceleration time-history at CG, Firewall and RearDeck for RICSAC test No. 6, Vehicle No. 2

For each of the refined RICSAC SMAC runs, the acceleration was monitored for the Firewall (Cowl) and RearDeck locations as specified in the RICSAC reports. The installation locations for the Firewall and RearDeck accelerometers are included in **Table 1**. The acceleration at any given location, X_1, Y_1, can be expressed for a given time, t, as:

$$a_{xx} = \dot{u} - v * \dot{\psi} \quad (1)$$

$$a_{yy} = \dot{v} + u * \dot{\psi} \quad (2)$$

$$a_{xx1} = \dot{u} - v * \dot{\psi} - x_{a1} * \dot{\psi}^2 - y_{a1} * \ddot{\psi}^2 \quad (3)$$

$$a_{yy1} = \dot{v} + u * \dot{\psi} + x_{a1} * \ddot{\psi} - y_{a1} * \dot{\psi}^2 \quad (4)$$

$$u_{x1} = u - y_{a1} * \dot{\psi} \quad (5)$$

$$v_{x1} = v + x_{a1} * \dot{\psi} \quad (6)$$

Where, for a given time interval dt:

u, v = forward and lateral velocity componenets at cg
$\dot{u}, \dot{v}$ = change in forward and lateral velocity components at cg
a_{xx}, a_{yy} = acceleration components at cg
ψ = vehicle heading angle
$\dot{\psi}$ = yaw velocity
$\ddot{\psi}$ = yaw acceleration
x_{a1}, y_{a1} = location of accelerometer relative to cg
u_{x1}, v_{x1} = forward and lateral velocity components at x_{a1}, y_{a1}
a_{xx1}, a_{yy1} = x, y accelerations at location x_{a1}, y_{a1}

The SMAC generated acceleration time histories at the CG, Firewall and RearDeck were then integrated to determine the instantaneous velocities and resulting speed-changes. Two techniques were used to determine the impact speed-changes of the individual tests: ***Test Equivalent*** and ***SMAC Equivalent***.

TEST EQUIVALENT: For the reported RICSAC full-scale tests **[1-3]** the analysis of the accelerometer data to determine the vehicle speed-change consisted primarily of the integration of the vehicle a_x & a_y accelerometer traces to determine the time histories of velocities. The velocity time-history plots were then analyzed to determine the magnitude of the changes in the individual velocity components, ΔV_x and ΔV_y. The resultant speed-change and speed-change angle were determined by the following:

$$\Delta v_r = \sqrt{\Delta v_x^2 + \Delta v_y^2} \quad (7)$$

Where:

Δv_r = resultant speed change
Δv_x = speed change in the vehicle x direction
Δv_y = speed change in the vehicle y direction

$$\Delta V_{angle} = \text{atan2}^{-1}\left[\frac{\Delta v_y}{\Delta v_x}\right] \quad (8)$$

Where:

ΔV_{angle} = speed change angle
Δv_x = speed change in the vehicle x direction
Δv_y = speed change in the vehicle y direction

SMAC EQUIVALENT:In the original form of the SMAC computer program, the speed-change for a given acceleration exposure was determined by the integration of the *resultant* acceleration. For each time increment, a resultant acceleration was determined by the vector sum of the x and y components of the vehicle accelerations:

$$a_c = \sqrt{a_x^2 + a_y^2} \quad (9)$$

Where:

a_c = resultant acceleration, g-units

a_x = acceleration in the vehicle x direction, g-units

a_y = acceleration in the vehicle y direction, g-units

The resultant speed-change was determined by integration of the *resultant* acceleration from initial contact to separation. The initial contact is defined as the first instance where the resultant acceleration goes above 1 g-unit. The point of separation is defined as the first time that the acceleration again drops back below 1 g-unit. It was assumed that in general handling maneuvers and vehicle spinouts that the resultant acceleration would normally be below 1 g-unit (given that the nominal friction coefficient of roadways is normally less than 1 g-unit).

The Direction of Principle Force (DOPF) was determined for a given acceleration exposure in the SMAC program as the direction of the acceleration at the instance of peak *resultant* acceleration by:

$$DOPF = \text{atan2}^{-1}\left[\frac{a_y}{a_x}\right] \quad (10)$$

Where:

DOPF = direction of principal force

a_x = acceleration in the vehicle x direction

a_y = acceleration in the vehicle y direction

A summary of the steps required to create SMAC "full-scale test results" to be used for testing purposes of the ***CG-Transform*** analytical procedure was as follows:

1. A SMAC simulation was performed for each RICSAC test
2. Acceleration time histories were created for the CG and the reported Firewall and Rear Deck accelerometer locations for each vehicle of each test. **Table 1** is a list of the reported Firewall and Rear Deck accelerometer locations used in the SMAC tests.
3. The SMAC vehicle fixed X & Y acceleration time-histories were integrated for the CG, the firewall and the Rear Deck as follows:

    3.1. The individual speed-changes in the vehicle X and Y directions were determined by integration of the X and Y acceleration time-histories.

    3.2. A ***Test equivalent*** resultant speed-change was calculated.

    3.3. The ***SMAC equivalent*** resultant speed-change was also calculated by integration of the time-history of the resultant acceleration.

For some of the RICSAC tests the values for the ***Test equivalent*** speed-change was found to be less than the ***SMAC equivalent*** speed-change (ΔV) by as much as 10%-15%. The difference is due to the following:

When we use integration to find the velocity change produced by an acceleration curve, we are finding the equivalent velocity that is produced by the variation of the acceleration over a fixed interval of time. In full scale tests, the accelerometers measure the acceleration for the X and Y directions separately. Therefore, the integration of these acceleration time-histories produce separate velocities (areas) for the X and Y directions. The *resultant* speed-change velocity from full-scale tests is determined by finding the square root of the sum of the squares for the X and Y velocities. In the SMAC program, in addition to the separate X and Y acceleration time-histories, we also have available the resultant acceleration. Therefore, to determine the *resultant* speed-change we can integrate directly the resultant acceleration vs. time curve. In a motor vehicle full-scale test, the vehicle and occupants experience the resultant acceleration, not individually the X and Y components. The possibility of differences in the resultant speed-change by these two methods reveals a need for accident reconstruction simulation programs to individually calculate the speed-changes in the X and Y directions for comparisons with full-scale tests, while also calculating and reporting the resultant speed-change determined by integration of the resultant acceleration over time.

For testing and comparison purposes of the ***CG-Transform*** analytical technique, reconstructions were performed of the collision phase of the 12 RICSAC tests with the SMAC computer program. **Table 4** contains a summary of the RICSAC SMAC impact speed-change (ΔV) for the CG, Firewall and Rear Deck, **Table 5** contains the impact conditions, and **Table 6** contains the separation conditions.

The results of the SMAC reconstructions of the RICSAC tests were used in an evaluation of the ***CG-Transform*** analytical technique. The integrated speed-change for the Firewall and Rear Deck accelerometer locations were used with the other items per **Table 3** in an application of the ***CG-Transform*** analytical procedure to determine the speed-change at the center of gravity and the separation velocity. The results for the speed-change (ΔV) and separation velocity for each of the 12 RICSAC tests were calculated. For all cases the speed-change and separation velocity calculated for the CG based on the Firewall and Rear Deck accelerometer data was less then 1 mph maximum error, with an average error of ±0.1 MPH and a standard deviation of 0.30 MPH. The comparison gave a good indication that the ***CG-Transform*** analytical procedure properly approximates the CG speed-change and separation speed from speed-change data at arbitrary accelerometer locations combined with discrete measures of the yaw responses.

A dramatic demonstration of the positional differences in accelerations and therefore the corresponding differences in the integrated speed-change calculations is illustrated in **Figure 3** and **Table 7**. **Figure 3** is the resultant acceleration time-history for the CG, Firewall and RearDeck for the SMAC reconstruction for vehicle number 1 of RICSAC Test No. 10. The full scale test and the SMAC simulation of RICSAC 10 each contained a "sideslap" impact of the vehicles subsequent to the primary impact. In **Figure 3** there are two peaks in the acceleration time histories, particularly at the RearDeck location. **Table 7** contains the integrated speed-changes calculated for the primary impact, the side-slap and for the total duration. The speed-change calculated by components (***Test equivalent***) and based on the resultant acceleration (***SMAC equivalent***) for vehicle No. 1 illustrates a problem which can occur when integrating individual acceleration components. **Figure 4** is the X-Y acceleration components for vehicle number 1 of RICSAC Test No. 10. The X-acceleration for the RearDeck location in **Figure 4** reverses sign after the primary impact which acts to reduce and *reverse* the calculated X component of speed-change for the ***Test equivalent*** integration technique (The reported time-histories of the velocity components produced

by integration of the RICSAC accelerations contain this anomaly, e.g., p 16-41,[2]). Determining the speed-change based on the resultant acceleration per the ***SMAC equivalent*** technique eliminates the directional sensitivity.

The side-slap acceleration of RICSAC Test 10 also reveals a problem of determining which speed-change and/or separation velocity should be reported when a sideslap occurs. Note that the ***SMAC Equivalent*** resultant speed-change for the Rear Deck calculated from the initial impact through the end of the sideslap (from 0.50 to 0.862 sec.) is dramatically higher (63 mph) than the sum of the individual speed-changes for the primary impact (25.5 mph) and sideslap impact (17.5 mph). The difference is due to the integration of the acceleration (> 5 g's) which occurs at the RearDeck (due to it's offset from the center of gravity) during the rapid rotation of the vehicle between the primary and sideslap impact (approximately 0.20 seconds).

It is herein proposed that all test reports include provisions to indicate a "side-slap" impact. The post-processing analysis of the acceleration data should also integrate and report the speed-change for the primary and secondary impacts separately.

Large accelerations that can occur at locations other than the center of gravity, as demonstrated at the RearDeck for RICSAC Test 10 vehicle 1, should not be included in the speed-change calculations for the individual vehicles since the vehicles are not in contact during this time period. However, the additional accelerations due to offset from the center of gravity and rotational velocity may have dramatic effects on the magnitude of occupant exposure and consideration should be given to localized ΔV calculations as suggested in [11].

Table 4 **SMAC RICSAC reconstructions predicted ΔV's for CG, Firewall and Rear Deck accelerometer locations**

SMAC RICSAC Reconstruction Results for Speed-change at CG, Firewall and Rear Deck

RICSAC Test No.	S&N Test No.	Location	Vehicle1 Delta-V Components X (mph)	Vehicle1 Delta-V Components Y (mph)	Vehicle1 *Test Equiv* Resultant (mph)	Vehicle1 *SMAC Equiv* Resultant (mph)	Vehicle2 Delta-V Components X (mph)	Vehicle2 Delta-V Components Y (mph)	Vehicle2 *Test Equiv* Resultant (mph)	Vehicle2 *SMAC Equiv* Resultant (mph)
1	1	cg	-13.2	3	13.5	13.6	-12.7	-12.8	18.1	18.2
		firewall	-12.3	3.7	12.9	13	-12.5	-12.4	17.6	17.8
		reardeck	-12.5	-6	13.8	14.1	-12.9	-14.9	19.7	20.1
2	2	cg	-19.9	5	20.5	21.3	-20.7	-20.8	29.4	30.9
		firewall	-18.7	6.7	19.8	21	-20.3	-19.8	28.4	29.9
		reardeck	-17.5	-8.8	19.6	21.2	-20.9	-25.7	33.1	34.6
3	5	cg	-9.7	-1.1	9.8	9.9	15.4	0.2	15.4	15.5
		firewall	-9.7	-1.2	9.8	9.9	15.5	0.3	15.5	15.6
		reardeck	-9.7	-1	9.8	10	15.4	-1.1	15.4	16.4
4	6	cg	-16.3	-2.3	16.5	16.7	24.7	0.3	24.7	25.4
		firewall	-16.3	-2.1	16.4	16.7	25	0.5	25	25.6
		reardeck	-16.3	-2.8	16.5	17	23.6	2	23.7	23.7
5	7	cg	-14.8	-2.2	15	15.2	26.6	-0.5	26.6	27.2
		firewall	-14.8	-2.1	15	15.3	26.3	0.4	26.3	26.5
		reardeck	-14.8	-2.8	15.1	15.4	26.5	-3.3	26.7	28.3
6	3	cg	-10	1.9	10.2	10.2	-11.8	-9.8	15.3	15.4
		firewall	-9.4	2.8	9.8	9.8	-11.1	-4.7	12	12.2
		reardeck	-9.6	-4.6	10.7	10.8	-11.7	-19.8	23	23.1
7	4	cg	-12.9	0.8	12.9	12.9	-12.2	-15.9	20.7	20.2
		firewall	-12.3	1.5	12.4	12.4	-11.5	-11.2	16	16.2
		reardeck	-12.7	-4.4	13.4	13.5	-12.3	-25.6	28.4	28.8
8	10	cg	-13.1	5.2	14.1	15	-5	-10.8	11.9	13
		firewall	-12.3	6.4	13.9	15.4	-4.5	-10.1	11.1	12
		reardeck	-12.1	-4	12.7	14.1	-5.1	-15.4	16.2	16.9
9	11	cg	-17.2	9.8	19.8	20.1	-4.5	-6.3	7.8	8.2
		firewall	-16	11.9	20	20.5	-4.9	-7	8.5	8.9
		reardeck	-15.3	0.8	15.3	16.2	-4.4	-3.5	5.6	5.9
10*	12	cg	-24.1	16.4	29.2	30.4	-7.3	-11.3	13.5	14.1
		firewall	-22.1	20.6	30.3	31.4	-8.1	-12.6	15	15.7
		reardeck	-16.4	-7.2	17.9	25.5	-6.3	-4	7.5	7.9
11	8	cg	-27.8	2.2	27.9	28.1	-16.9	3.4	17.3	17.4
		firewall	-27.4	1.6	27.4	27.4	-16.9	3.5	17.3	17.5
		reardeck	-27.7	3.8	27.9	28	-16.9	3	17.2	17.5
12	9	cg	-38.7	2.1	38.7	39.5	-26.4	5.5	27	27.4
		firewall	-38.6	1	38.6	39.7	-26.4	5.2	26.8	27.5
		reardeck	-38.4	6.2	38.9	39.2	-26.4	7	27.3	27.5

***Note: Reported test #10 results are for primary collision only. Please see discussion in text.**

Table 5 SMAC RICSAC reconstructions Impact Conditions

SMAC Impact Conditions

RICSAC Test No.	S&N Test No.	Veh No.	x (ft)	y (ft)	psi (deg)	u (mph)
1	1	1	-11	1.2	-30	19.8
		2	0	6.3	90	19.8
2	2	1	-11.1	7.8	-30	31.8
		2	0	-0.5	90	31.8
3	5	1	-0.1	0	0	21.2
		2	16.2	3.6	10	0
4	6	1	0.1	0	0	38.7
		2	16.4	3.5	10	0
5	7	1	-14.9	0	0	39.3
		2	0	3.1	10	0
6	3	1	4.5	0.6	0	21.5
		2	16.4	1.4	120	21.4
7	4	1	5.4	0	0	29.1
		2	17.5	0.4	120	29.1
8	10	1	-11.1	1.3	0	20.3
		2	0	1	90	20.5
9	11	1	0	0	0	21
		2	8.5	-5.4	90	21
10	12	1	0	0	0	33.3
		2	8.4	-6.7	90	33.3
11	8	1	15.7	-4.1	170	20.4
		2	0	0	0	20.4
12	9	1	15.6	-4.2	170	31.5
		2	0	0	0	31.5

Table 6 SMAC RICSAC reconstructions Separation Conditions

SMAC Separation Conditions

RICSAC Test No.	S&N Test No.	Veh No.	Contact Duration (sec)	x (ft)	y (ft)	psi (deg)	Psi (deg)	u (mph)	v (mph)	PsiD (deg/sec)
1	1	1	0.168	-7.6	-0.47	-23	-23.4	6.62	1.66	102
		2	0.168	1	-2.5	90	90.2	6.86	-12.8	30
2	2	1	0.135	-6.7	5.7	-19	-19	11.94	1.54	164
		2	0.135	1.3	4.1	92	92.5	10	-21.5	69
3	5	1	0.119	2.9	0	1	0.5	11.5	-1.3	-1
		2	0.119	17.3	3.7	12	12	15.4	-0.22	22
4	6	1	0.13	6	-0.2	1	1.2	22.3	-2.9	6.9
		2	0.13	18.6	3.8	14	14.4	24.8	-1	51
5	7	1	0.119	-9.2	-0.2	1	1	24.4	-2.8	5.5
		2	0.119	2	3.4	15	14.9	26.6	-1.8	64.5
6	3	1	0.126	8.6	0.7	5	4.7	11.3	0.7	79.5
		2	0.126	15.1	5.2	128	128	8.3	-11.7	176.5
7	4	1	0.125	9.8	0	3	3.4	16	-0.4	63.5
		2	0.125	16.3	4.6	126	125.9	15.1	-16.1	172
8	10	1	0.091	-9.1	1.8	7	6.5	7.6	3.8	102
		2	0.091	0.6	3.3	93	92.8	15.1	-11.6	58.1
9	11	1	0.118	2	0.9	14	14.4	5.1	7.9	191
		2	0.118	9	-2.2	88	88	16.6	-5.3	-42
10*	12	1	0.134	3.7	1.4	24	24	12.5	10.5	350
		2	0.134	9.7	-0.9	81	81.5	26.9	-7.2	-91.1
11	8	1	0.125	14.1	-4	166	166	-7.4	2	-45
		2	0.125	2.5	0.2	-0.5	-0.4	3.4	3.3	2.2
12	9	1	0.105	13.3	-3.9	165	165	-7.6	2.9	-77.2
		2	0.105	3.1	0.3	-1.4	-1.4	4.9	6.1	-20
10**	12	1	0.059	5.7	6.4	88	87.8	16.9	-1.2	88.5
		2	0.059	13.2	7.4	68	67.5	25.8	-1.9	20.1

Test#10: * Primary Impact, ** Secondary impact

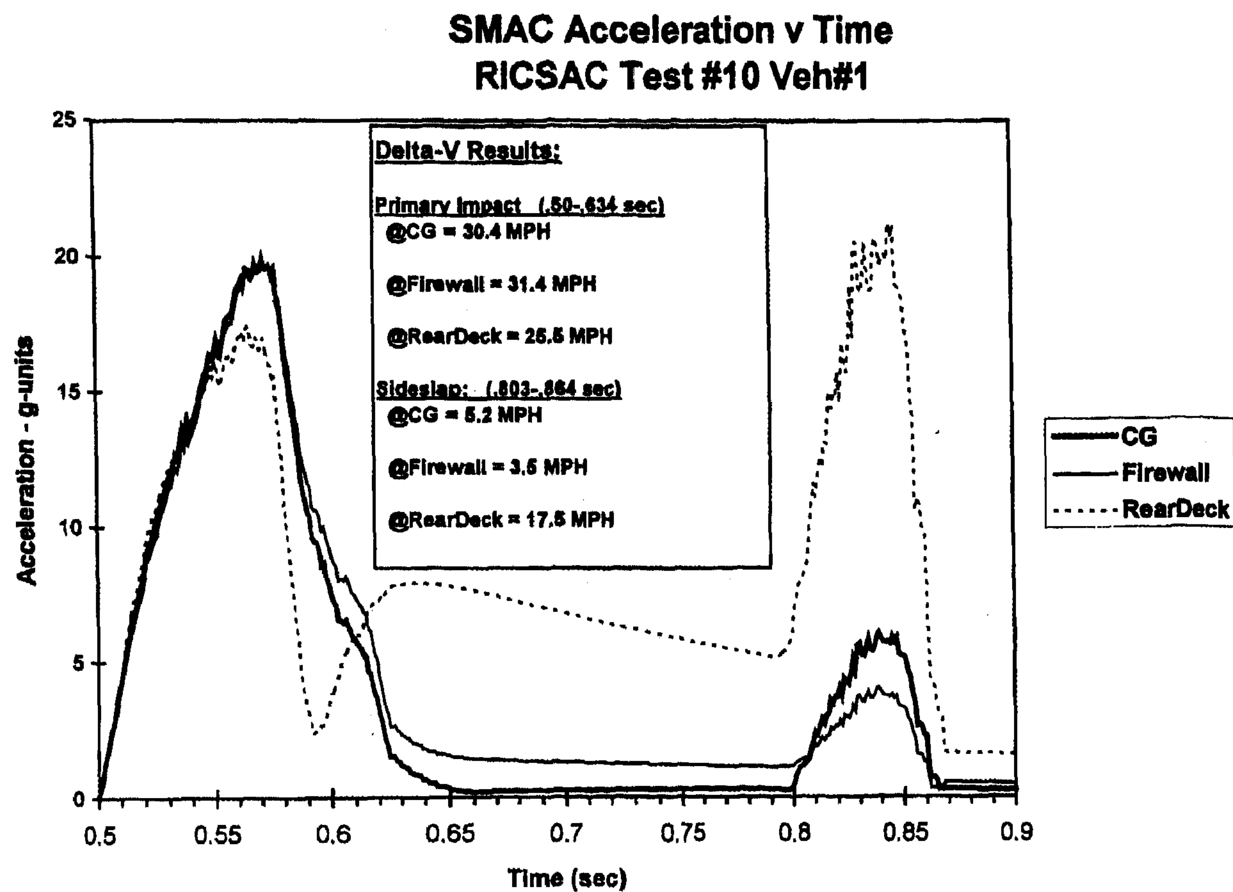


Figure 3 Resultant Acceleration time histories for RICSAC Test No. 10, Veh#1 for CG, Firewall and RearDeck

Table 7 Comparison of Calculated Speed-changes for RICSAC test No.10 for CG, Firewall and RearDeck.

Location	Vehicle1 Delta-V Components X (mph)	Vehicle1 Delta-V Components Y (mph)	Vehicle1 Test Equiv Resultant (mph)	Vehicle1 SMAC Equiv Resultant (mph)	Vehicle2 Delta-V Components X (mph)	Vehicle2 Delta-V Components Y (mph)	Vehicle2 Test Equiv Resultant (mph)	Vehicle2 SMAC Equiv Resultant (mph)
Delta-V's for primary event: (T=0.50 to 0.634):								
cg	-24.1	16.4	29.2	30.4	-7.3	-11.3	13.5	14.1
firewall	-22.1	20.6	30.3	31.4	-8.1	-12.6	15	15.7
reardeck	-16.4	-7.2	17.9	25.5	-6.3	-4	7.5	7.9
Delta-V's for sideslap:(T=0.803-862 sec):								
cg	2.1	4.7	5.1	5.2	-0.8	-2.1	2.2	2.2
firewall	0.1	3.4	3.4	3.5	-0.3	-1	1	1.1
reardeck	4.7	16.8	17.4	17.5	-0.7	-8.1	8.1	8.1
Delta-V's for total event:(T=0.50-862 sec):								
cg	-22.9	21.1	31.2	36.7	-9.2	-12.5	15.5	17.8
firewall	-26.2	26	36.9	39.7	-9.6	-12	15.4	18.7
reardeck	7.6	14.9	16.7	63.1	-6.7	-14.1	15.6	18.2

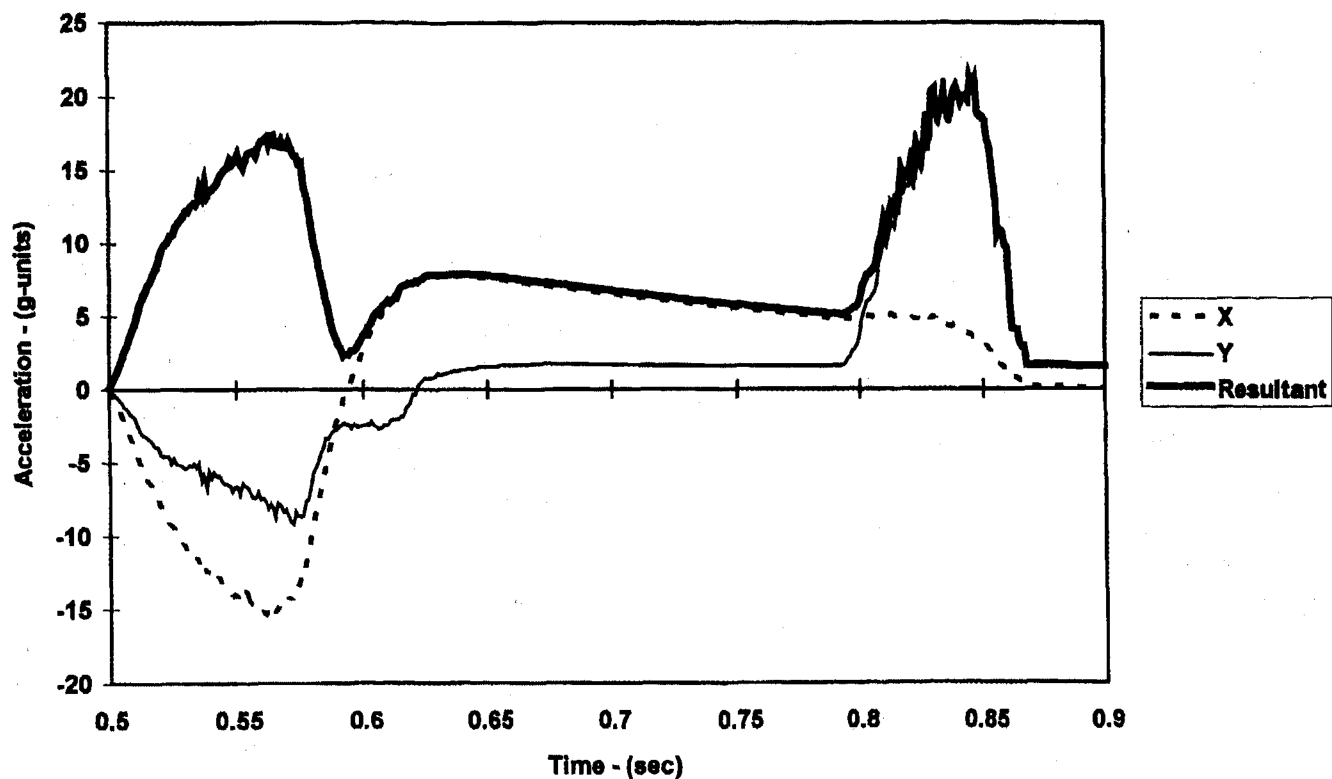


Figure 4 Acceleration components at RearDeck Location for RICSAC Test 10, Vehicle No. 1

Another check of the *CG-Transform* was to perform a comparison of the calculated results with results reported by Bundorf in [15]. Bundorf included a procedure for transformation of accelerometer information from locations other than the vehicle center of gravity to the CG. The procedure presented by Bundorf includes digitizing and analyzing the acceleration and angular velocity time-history data. The reported results by Bundorf were compared with results calculated with the *CG-Transform* procedure. The results of the comparison are contained in **Table 8**. The results of the *CG-Transform* procedure correlate well with the results of the Bundorf analysis procedure. The average result and standard deviation are nearly identical. Some small differences did occur between the different accelerometer locations.

The results of application of *CG-Transform* analytical procedure to the various locations tested demonstrate the ability of the *CG-Transform* procedure to accurately calculate the speed-change at the center of gravity for any arbitrary accelerometer location.

Table 8 Comparison of Bundorf [15] with *CG-Transform* results for RICSAC Test 9

Delta-V (MPH) for RICSAC Test 9 (S&N 11), Vehicle 1

Description	Location X (in)	Location Y (in)	Measured @Location X (mph)	Measured @Location Y (mph)	Measured @Location Res (mph)	Calculated for CG per Bundorf X (mph)	Calculated for CG per Bundorf Y (mph)	Calculated for CG per Bundorf Res (mph)	Calculated for CG with CG-Transform X (mph)	Calculated for CG with CG-Transform Y (mph)	Calculated for CG with CG-Transform Res (mph)
LF	-3.3	-21	-9.8	11.0	14.7	-14.6	10.9	18.2	-14.6	10.9	18.3
RSide	-41.8	21	-17.8	0.4	17.8	-14.7	10.9	18.3	-14.7	10.8	18.3
Cowl	10.2	-9	-16.4	15.0	22.2	-17.8	11.3	21.1	-18.0	12.3	21.8
Rear Deck	-64.3	0	-12.2	-3.4	12.7	-15.2	11.4	19.0	-14.8	11.3	18.6
			Average Result			-15.6	11.1	19.2	-15.5	11.3	19.3
			Std Deviation			1.5	0.3	1.3	1.7	0.7	1.7

EVALUATION OF THE "RAW" RICSAC DATA

The task of evaluating and transforming the accelerometer data for the RICSAC full scale tests was undertaken to determine the CG speed-change and separation velocity for each test. A few possible problem areas which might affect the correlation are:

Source of Raw Data.

Acceleration Time-history Oscillations.

Excessive Collision Duration.

SOURCE OF RAW DATA: The raw ΔV values as reported in [3] were used for the comparison. A brief visual check of the time-history plots vs. the reported values was performed to check for any major differences. The reason the other reported refinements of the values could not be used was that the other reported values contained adjustments and corrections to transform the results from the Firewall accelerometer locations to the CG, the very task our ***CG-Transform*** is to perform.

ACCELERATION TIME-HISTORY OSCILLATIONS: Some of the acceleration time traces contained in the RICSAC reports [2] appear to have excessive oscillations (e.g., RICSAC Tests.Veh: 2.1,2.2,5.1,5.2,8.2,11) and some have resultant accelerations exceeding 40 g's (e.g., RICSAC Test.Veh: 10.1,12.1,7.2). These results may either reflect a possible problem with the accelerometers or with the choice in cutoff frequency filter. Sample oscillograph data from Mayor and Naab **[18]** contained in [19] illustrates the differences between unfiltered and filtered acceleration data for a 20 mph lateral impact. Peak accelerations vary from approximately 100 g's for unfiltered data to 40 g's for 50 Hz cutoff frequency to 28 g's for 25 Hz cutoff frequency. Also a dramatic reduction in the amount of oscillation occurs with increased filtering.

EXCESSIVE COLLISION DURATION: The collision contact duration reported by Jones (see **Table 2**) appears to be excessive. For example, Bundorf **[15]** determined the pulse duration for RICSAC Test 9 to be approximately 0.125 seconds whereas Jones reported the duration to be 0.200 seconds. Generally for car-to-car collisions, a normal range of pulse duration is 0.075 to 0.125. A review of the time-history accelerometer traces and the integrated velocity traces contained in the original RICSAC test reports also appears to indicate that the durations of the pulses reported by Jones are excessive. A comparison of the SMAC reconstruction pulse durations indicates that an average adjustment of the duration of approximately 0.090 seconds would be appropriate. For the purposes of this reported research, an approximate adjustment of 0.0825 was used (the average of .075 + .09).

RESULTS

The inputs for the reported raw RICSAC data [2, 3] which were used with the *CG-Transform* analytical technique are contained in ***Table 9***. A comparison of the *CG-Transform* calculated results with the speed-change at the CG as predicted by the SMAC program is contained in Table 10. A comparison of the *CG-Transform* calculated separation velocities with the separation velocities predicted by the SMAC program is contained in **Table 11**.

Given the possible variations and problems which can occur in full-scale test measurements the correlation of the ***CG-Transform*** calculated results and the SMAC results for most of the RICSAC tests appears very good. In all cases the cited correlation of ***CG-Transform*** with the mathematically correct SMAC results is within ±10% for ΔV and within ±13% for the separation velocity for at least one of the vehicles. This gives a good indication that properly interpreted results of all the tests are in complete agreement with Newton's Laws.

Some areas of difference can be explained as follows:

- For the ΔV values from the full-scale tests, differences in the individual components may be attributable to possible accelerometer calibration errors. Errors also can be caused by cross-coupling problems with gyros which can occur due to vehicle roll and pitch angles. Larger errors may be produced by problems with the individual accelerometers either in the crush zone or on components possibly affected by the impact configuration.
- For the separation velocity, differences in the individual components and resultant are produced by any of the following possibilities:
  - possible errors in determination of the time of separation in the full-scale tests
  - possible errors in the reported orientation of the vehicle at the time of separation
  - possible synchronization problems between the triaxle accelerometers and the gyros

Table 9 *CG-Transform* inputs for the 12 RICSAC full-scale tests raw data

RICSAC Test No.	S&N Test No.	Veh No.	Impact Speed (mph)	Impact Heading Angle (deg)	Impact Speed Change u (mph)	Impact Speed Change v (mph)	Impact Speed Change Res (mph)	Reported Pulse Width (sec)	Adj Pulse Width (sec)	Sep Angular Velocity (deg/sec)	Sep Heading Change (deg)	Sep Heading Angle (deg)
1	1	1	19.7	-30.0	-10.6	6.0	12.1	0.225	0.143	90	15	-15
		2	19.7	90.0	-12.1	-9.8	15.5	0.225	0.143	0	0	90
2	2	1	31.3	-30.0	-16.5	10.5	19.6	0.225	0.143	150	18	-12
		2	31.3	90.0	-20.1	-19.2	27.8	0.225	0.143	90	2.5	93
3	5	1	21.1	0.0	-9.5	-0.4	9.5	0.2	0.118	15	1	1
		2	0.0	10.0	15.8	-0.2	15.8	0.2	0.118	0	0	10
4	6	1	38.5	1.0	-18.7	0.4	18.7	0.275	0.193	37	0	1
		2	0.0	10.0	22.2	-2.8	22.4	0.275	0.193	30	0	10
5	7	1	39.5	0.0	-16.3	0.2	16.3	0.25	0.168	12	5	5
		2	0.0	10.0	25	-1.8	25.1	0.25	0.168	70	0	10
6	3	1	21.4	0.0	-8.5	3.0	9.0	0.2	0.118	30	5	5
		2	21.4	120.0	-11.5	-3.2	11.9	0.2	0.118	180	20	140
7	4	1	29.0	0.0	-11.5	3.5	12.0	0.2	0.118	30	12	12
		2	29.0	120.0	-14.1	-8.5	16.5	0.2	0.118	192	22	142
8	10	1	20.6	0.0	-12.7	8.6	15.3	0.2	0.118	114	15	15
		2	20.6	90.0	-7.2	-8.0	10.8	0.2	0.118	18	0	90
9	11	1	21.1	0.0	-17.7	12.0	21.4	0.2	0.118	180	27	27
		2	21.1	90.0	-5.0	-7.4	8.9	0.2	0.118	-45	-10	80
10	12	1	33.1	0.0	-27.3	22.0	35.1	0.2	0.118	300	25	25
		2	33.1	90.0	-8.8	-11.0	14.1	0.2	0.118	-72	-12	78
11	8	1	20.3	171.0	-24.0	0.8	24.0	0.225	0.143	-30	-5	166
		2	20.3	0.0	-15.6	2.0	15.7	0.225	0.143	0	0	0
12	9	1	31.3	171.0	-40.0	-2.2	40.1	0.225	0.143	-90	-10	161
		2	31.3	0.0	-26.0	4.8	26.4	0.225	0.143	-60	-2	-2

Table 10 Comparison of ΔV at the CG based on *CG-Transform* calculated results and SMAC reconstructions for the 12 RICSAC full-scale tests

RICSAC Test No.	S&N Test No.	Veh No.	*CG-Transform* Results Delta-V at CG X (mph)	Y (mph)	Res (mph)	Difference from SMAC Results Components X (mph)	Components Y (mph)	Resultant (mph)	Resultant (%)	Direction diff (deg)
1	1	1	-11.3	4.8	12.3	1.9	1.8	-1.2	-9	-10
		2	-12.6	-11.2	16.9	0.1	1.6	-1.2	-7	-4
2	2	1	-17.4	8.8	19.5	2.5	3.8	-1.0	-5	-13
		2	-18.8	-17.7	25.8	1.9	3.1	-3.5	-12	-2
3	5	1	-9.5	-0.7	9.5	0.2	0.4	-0.2	-3	-2
		2	15.8	-0.2	15.8	0.4	-0.3	0.4	3	-1
4	6	1	-18.7	-0.3	18.7	-2.4	2.0	2.2	14	-7
		2	22.0	-2.9	22.2	-2.7	-3.1	-2.5	-10	-8
5	7	1	-16.2	-0.1	16.2	-1.4	2.1	1.3	9	-8
		2	25.4	-3.0	25.6	-1.2	-2.5	-1.0	-4	-6
6	3	1	-8.8	2.4	9.2	1.2	0.5	-1.0	-10	-5
		2	-10.5	-9.8	14.4	1.3	0.0	-1.0	-6	3
7	4	1	-11.7	2.5	11.9	1.2	1.7	-1.0	-8	-8
		2	-12.8	-15.1	19.8	-0.6	0.8	-0.3	-1	-3
8	10	1	-13.4	6.6	14.9	-0.3	1.4	0.8	6	-5
		2	-7.4	-8.2	11.0	-2.4	2.6	-0.9	-7	-17
9	11	1	-18.0	8.5	20.0	-0.8	-0.7	0.5	2	3
		2	-4.4	-6.9	8.2	-0.2	-0.6	0.6	8	1
10	12	1	-28.8	18.1	34.0	-4.7	1.7	4.9	17	2
		2	-7.7	-9.9	12.5	-0.4	1.4	-0.9	-7	-6
11	8	1	-24.4	1.6	24.5	3.4	-0.6	-3.4	-12	1
		2	-15.6	2.0	15.7	1.3	-1.4	-1.5	-9	4
12	9	1	-40.8	-0.8	40.8	-2.1	-2.9	2.1	5	4
		2	-26.1	5.7	26.7	0.3	0.2	-0.3	-1	-1

Table 11 Comparison of Separation velocities based on *CG-Transform* calculated results and SMAC reconstructions for the 12 RICSAC full-scale tests

RICSAC Test No.	S&N Test No.	Veh No.	*CG-Transform* RESULTS Separation Velocity X (mph)	Y (mph)	Res (mph)	Difference from SMAC Results Components X (mph)	Y (mph)	Resultant (mph)	(%)	Direction diff (deg)
1	1	1	8.5	1.6	8.6	1.8	-0.1	1.7	25	-4
		2	7.1	-10.5	12.7	0.2	2.3	-1.9	-13	6
2	2	1	14.2	2.2	14.4	2.2	0.7	2.3	19	2
		2	11.4	-20.6	23.5	1.4	0.9	-0.2	-1	4
3	5	1	11.6	-0.9	11.6	0.1	0.4	0.0	0	2
		2	15.8	-0.2	15.8	0.4	0.0	0.4	3	0
4	6	1	19.8	-0.2	19.8	-2.5	2.7	-2.7	-12	7
		2	21.9	-3.2	22.2	-2.9	-2.2	-2.6	-11	-6
5	7	1	23.2	-2.4	23.3	-1.2	0.4	-1.3	-5	1
		2	25.3	-3.8	25.5	-1.3	-2.0	-1.1	-4	-5
6	3	1	12.6	1.2	12.6	1.3	0.5	1.3	12	2
		2	9.1	-14.6	17.2	0.8	-2.9	2.8	20	-3
7	4	1	17.2	-1.5	17.3	1.2	-1.1	1.3	8	-4
		2	13.2	-22.4	26.0	-1.8	-3.9	2.2	9	-8
8	10	1	7.5	3.3	8.2	-0.1	-0.5	-0.3	-3	-3
		2	13.2	-8.2	15.6	-1.9	3.5	-3.5	-19	8
9	11	1	3.4	2.9	4.5	-1.7	-5.0	-4.9	-52	-17
		2	16.8	-4.1	17.3	0.2	1.2	-0.1	0	4
10	12	1	6.0	10.1	11.8	-6.5	-0.5	-4.6	-28	19
		2	25.7	-4.8	26.1	-1.1	2.4	-1.6	-6	4
11	8	1	-4.2	2.3	4.8	3.2	0.3	-2.9	-37	-13
		2	4.7	2.0	5.1	1.3	-1.3	0.3	7	-21
12	9	1	-9.6	1.2	9.6	-2.0	-1.7	1.6	19	14
		2	5.1	6.2	8.0	0.2	0.1	0.2	2	-1

CONCLUSIONS

1. The RICSAC data, when interpreted with the *CG-Transform* procedure, are reasonably accurate and are suitable for their intended purpose of testing the validity of reconstruction techniques.
2. Previously reported findings of gross errors and violations of Newton's laws in the reported RICSAC data are erroneous.
3. The SMAC program has been demonstrated to correlate well with properly analyzed full-scale test results

RECOMMENDATIONS

1. Future "validation" comparisons with the RICSAC full-scale tests should also give consideration to utilizing the mathematically correct SMAC results presented herein. Use of the SMAC program results for comparisons avoids the variability that normally occurs in full-scale tests. Without direct measures of the variability of experimental responses, by means of repeated runs of tests, the accuracy of analytical predictions cannot be properly evaluated.
2. Future full-scale tests should utilize a time-history simulation program like SMAC or a three-dimensional equivalent to permit a check of results and conversion procedures for data from accelerometers not located at the CG.
3. Any integration analysis of accelerometer data should include consideration for possible directional changes of individual components. This can be accomplished by direct analysis of the resultant acceleration to avoid the possibility of a reduction in the reported values for individual component velocities due to acceleration directional reversals.
4. Full-scale tests should include provisions for separately reporting and analyzing the occurrence of "side-slap" types of collisions
5. Full-scale test reporting and subsequent analysis should be in a coordinate system in conformity with standard SAE J670e

REFERENCES

1. McHenry, R. R., Lynch, J. P., Segal, D. J., "Research Input for Computer Simulation of Automobile Collisions", Calspan Report ZQ-6057-V-1, Contract DOT-HS-7-01511, Jun 1977
2. Shoemaker, N. E., "Research Input for Computer Simulation of Automobile Collisions - Staged Collisions, Vol. II & Vol. III", Calspan Report ZQ-6057-V-4 & V-5, Contract DOT-HS-7-01511, December 1978
3. Jones, I. S., Baum, A. S., "Research Input for Computer Simulation of Automobile Collisions - Volume IV - Staged Collision Reconstructions", Calspan Report ZQ-6057-V-6, Contract DOT-HS-7-01511, December 1978
4. Smith, R. A., Noga, J. T., "Examples of Staged Collisions in Accident Reconstruction", presented at ASME Winter Meeting, Nov 1980 and contained in ASME publication "Highway Collision Reconstruction"

5. James, M. E., Ross, H. E., Whittington, C., "Improvement of Accident Simulation Model and Improvement of Narrow Object Accident Reconstruction", Texas Transportation Institute, DOT-HS-803-620, DOT-HS-5-01262 & DOT-HS-7-01656, April 1978
6. Shaw, L. M., "Countermeasures for Side Impact", Contract DOT-HS-9-02177
7. Brach, R. M., "Impact Analysis of Two-Vehicle Collisions", SAE Paper 83-0468
8. Wooley, R.L., Warner, C.Y., Tagg, M. D.,"Inaccuracies in the CRASH3 Program", SAE paper 85-0255
9. Wooley, R.L, "The "IMPAC" Computer Program for Accident Reconstruction", SAE paper 85-0254
10. Day, T.D., Hargens, R.L.,"Further Validation of the EDCRASH Using the RICSAC Staged Collisions", SAE paper 89-0740
11. Cheng, P.A., Guenther, D. A., "Effects of Change in Angular Velocity of a Vehicle on the Change in Velocity Experienced by an Occupant during a Crash Environment and the Localized Delta-V Concept", SAE Paper 89-0636
12. Day., T. D., Hargens, R. L., "Further Validation of EDSMAC Using the RICSAC Staged Collisions", SAE Paper No. 90-102
13. Smith, R.A., Noga, J. T., "The Accuracy and Sensitivity of Crash", DOT-HS-806-152, March 1982
14. Cliff, E. C., Montgomery, D. T., "Validation of PC-Crash - A Momentum-Based Accident Reconstruction Program", SAE Paper 96-0885
15. Bundorf, R.T., "Analysis and Calculation of Delta-V from Crash Test Data", SAE Paper 96-0899
16. Day., T.D., Hargens, R.L.,"Further Validation of EDSMAC Using the RICSAC Staged Collisions", SAE Paper No. 90-0102
17. A review of the trajectory aspects of the RICSAC tests are expected to be completed in early 1997 and will then be published.
18. Mayor, R. P., Naab, K. N., "Basic Research in Automobile Crashworthiness - Testing and Evaluation of Modifications for Side Impact", Cornell Aeronautical Laboratory, CAL Report No. YB-2684-V-4, October 1969)
19. McHenry R. R., Miller, P. M., "Automobile Structural Crashworthiness", SAE Paper 70-0412

CONTACT INFORMATION

Questions or comments on the paper are welcomed and can be addressed to the authors by:

e-mail: **mchenry@interpath.com**

Postal Service Mail: **103 Brady Court, Suite 200**

Cary, NC 27511 USA

WWW: **http://www.mchenrysoftware.com**

APPENDIX 1: *CG-TRANSFORM* PROCEDURE

In the following, a procedure is defined whereby accelerometer measurements taken at locations away from the CG are transformed into impact speed-changes at the center of gravity and into separation velocities on the basis of discrete measures of the vehicle responses rather than complete time-histories. The procedure, entitled ***CG-Transform***, is validated by means of applications to SMAC runs in which the outputs include acceleration components at selected locations on the simulated vehicles.

DEFINITION OF SYMBOLS

OXY =	Vehicle-fixed coordinate system with origin at the center of gravity.
O'X'Y'=	Space-fixed coordinate system.
ψ =	Heading angle of vehicle relative to space-fixed system.
u, v =	Components of velocity in the directions of the vehicle-fixed X and Y coordinates respectively, inches/second.
u', v' =	Components of velocity in the directions of the space-fixed X' and Y' coordinates respectively, inches/second.
x1, y1 =	Coordinates of accelerometer location in vehicle-fixed system, inches.
ax1, ay1 =	Components of acceleration at point 1 relative to the vehicle-fixed system, inches/sec^2.
a_x, a_y =	Components of acceleration at the center of gravity relative to the vehicle-fixed system, inches/sec^2.
Δu, Δv =	Changes in velocity in the directions of the vehicle-fixed X and Y coordinates respectively, inches/sec.

In the following pages, a dot over a symbol is used to indicate a time derivative.

The subletter **1** indicates the value of a variable at point **1**

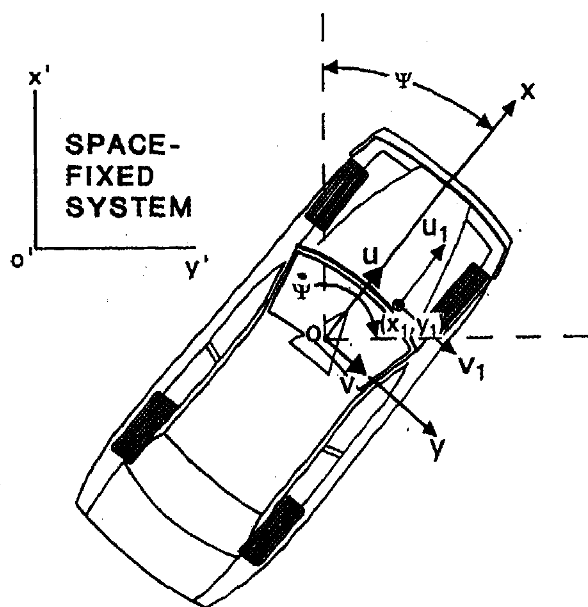


Figure 5 Velocity components of a Point on a vehicle

In ***Figure 5***, velocity components of point 1 relative to the vehicle-fixed coordinate system are defined by:

$$u_1 = u - y_1\dot{\psi} \qquad (11)$$

$$v_1 = v + x_1\dot{\psi} \qquad (12)$$

The velocity components of point P relative to the space-fixed coordinate system may be expressed as:

$$u_1' = u_1 \cos\psi - v_1 \sin\psi \qquad (13)$$

$$v_1' = u_1 \sin\psi + v_1 \cos\psi \qquad (14)$$

Substitution of (11) and (12) into (13) and (14) and differentiation of the resulting equations yields the following components of acceleration relative to the vehicle-fixed system:

$$a_{x1} = \dot{u} - v\dot{\psi} - y_1\ddot{\psi} - x_1\dot{\psi}^2 \qquad (15)$$

$$a_{y1} = \dot{v} + u\dot{\psi} + x_1\ddot{\psi} - y_1\dot{\psi}^2 \qquad (16)$$

From (15) and (16), the components of acceleration relative to the vehicle-fixed system at the center of gravity of the vehicle (i.e., $x_1 = y_1 = 0$) may be expressed:

$$a_x = \dot{u} - v\dot{\psi} \qquad (17)$$

$$a_y = \dot{v} + u\dot{\psi} \qquad (18)$$

Thus, the ΔV components at the center of gravity relative to the vehicle-fixed system, including rotational effects, may be obtained by integration of (15) and (16):

$$\int_o^t a_x dt = \int_o^t a_{x1} dt + y_1 \int_0^t \ddot{\psi} dt + x_1 \int_o^t \dot{\psi}^2 dt \qquad (19)$$

$$\int_o^t a_y dt = \int_o^t a_{y1} dt - x_1 \int_o^t \ddot{\psi} dt + y_1 \int_o^t \dot{\psi}^2 dt \qquad (20)$$

To obtain the components in the vehicle-fixed system, of the separation velocity (i.e., the velocity at the end of a period of acceleration) it is necessary to integrate (17) and (18) in the following manner:

$$u_f = u_o + \int_o^t \dot{u} dt = u_0 + \int_o^t a_x dt + \int_o^t v\dot{\psi} dt \qquad (21)$$

$$v_f = v_o + \int_o^t \dot{v} dt = v_o + \int_o^t a_y dt - \int_o^t u\dot{\psi} dt \qquad (22)$$

From the preceding, it may be seen that the following integrations are required to correct readings obtained from accelerometers at point 1:

$$\int_o^t \ddot{\psi} dt \qquad (23)$$

$$\int_o^t \dot{\psi}^2 dt \qquad (24)$$

$$\int_o^t v\dot{\psi} dt \qquad (25)$$

$$\int_o^t u\dot{\psi} dt \qquad (26)$$

Reasonable approximations of the values of the integrals listed as equations (23) through (26) can be obtained by the use of analytical functions fitted to test data:

$$\ddot{\psi} = Q_1 \sin \omega t - Q_2 \sin 3\omega t \quad (27)$$

$$u = u_f - (\Delta u) \cos \omega t \quad (28)$$

$$v = v_f - (\Delta v) \cos \omega t \quad (29)$$

Integration of equations (23) through (26) using the functional relationships defined by (27), (28) and (29) and application of the results to equations (19), (20), (21) and (22) yields the following relationships:

$$\omega = \frac{\pi}{2(\Delta t)} \quad (30)$$

$$Q_1 = \frac{3\omega}{4}\left[\frac{\pi}{2}\dot{\psi}_f + \frac{\Delta\dot{\psi}}{3} - \omega\Delta\psi\right] \quad (31)$$

$$Q_2 = 3\left[Q_1 - \omega\Delta\dot{\psi}\right] \quad (32)$$

$$A = \frac{1}{\omega}\left(\frac{\pi}{2}\dot{\psi}_f^{\,2} + \frac{Q_1}{\omega}\left[\frac{Q_1\pi}{4\omega} - 2\dot{\psi}_f\right] + \frac{Q_2}{9\omega}\left[\frac{Q_2\pi}{4\omega} - 2\dot{\psi}_f\right]\right) \quad (33)$$

$$\Delta V_x = \int_o^t a_x dt = \int_o^t a_{x1} dt + y_1\Delta\dot{\psi} + x_1 A \quad (34)$$

$$\Delta V_y = \int_o^t a_y dt = \int_o^t a_{y1} dt - x_1\Delta\dot{\psi} + y_1 A \quad (35)$$

$$B = \frac{1}{\omega}\left(\frac{\pi}{2}\dot{\psi}_f - \frac{Q_1}{\omega} - \frac{Q_2}{9\omega^2}\right) \quad (36)$$

$$C = \frac{1}{\omega}\left(\frac{\pi}{4}\frac{Q_1}{\omega} - \dot{\psi}_f\right) \quad (37)$$

$$D = \int_o^t a_x dt + B v_0 \quad (38)$$

$$E = \int_o^t a_y dt - B u_0 \quad (39)$$

$$delu = \Delta u = \left(\frac{D + (B+C)E}{1 + (B+C)^2}\right) \quad (40)$$

$$delv = \Delta v = \frac{\left(E - (B+C)D\right)}{\left(1 + (B+C)^2\right)} \quad (41)$$

$$u_f = u_0 + \Delta u \quad (42)$$

$$v_f = v_0 + \Delta v \quad (43)$$

SAE #970961
Reviewer's Discussion
by Don C. Stevens, MS, Arndt and Associates, Ltd.
RICSAC-97 - A Reevaluation of the Reference Set of Full Scale Crash Tests
Brian G. McHenry, Raymond R. McHenry, Authors

Although the RICSAC crash tests simulate only a narrow range of real-world impact conditions, they have long served as a tool for validation of vehicle collision simulation software as well as a body of data for the study of automobile impact dynamics. This paper discusses the RICSAC work along with some important limitations in its data.

Speed changes for the RICSAC crash tests were based on analysis of data from accelerometers. The results were affected to varying degrees by the placement of the accelerometers at locations other than the vehicle's CG. The authors present a sound new method for compensation of errors in the RICSAC results.

Velocity change (ΔV) during an instrumented crash test can be calculated through numerical integration of accelerometer data. The authors make insightful observations concerning potentially significant differences between common methods of reducing acceleration data to ΔV.

SAE #970961
Reviewer's Discussion
by Terry D. Day
RICSAC-97 - A Reevaluation of the Reference Set of Full Scale Crash Tests
Brian G. McHenry, Raymond R. McHenry, Authors

This paper provides an important re-evaluation of the test data published in the original RICSAC study. Several previous authors reported the raw RICSAC data were in error because the Delta-Vs reported in the study were based on acceleration vs. time histories from accelerometers not mounted at the CG. This paper provides a solution to this "error" (the authors correctly point out the error is not with the raw data, but with the data *reduction*) by calculating the correct vehicle CG velocities using a simple transformation. In the future, researchers may ignore the Delta-Vs reported in the original RICSAC study, and use the Delta-Vs reported in the paper.

970962

Avoiding the Emerging Pedestrian: A Mathematical Model

Rodney Vaughan
Rodney Vaughan & Associcates Pty. Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A common form of pedestrian accident involves the pedestrian emerging from behind a stopped vehicle, into the path of an oncoming car. A mathematical model ("EMERGE") has been developed for calculating the time available to the driver to see and avoid the pedestrian. It involves calculating in 2 dimensions the equations of motion of the vehicle and the pedestrian, together with the lines of sight of the driver and the pedestrian around the stopped vehicle. The sensitivity of the model to the different variables is demonstrated.

The model allows for deceleration of the car. For any given driver perception-reaction time and car deceleration rate, the model can be used to calculate the maximum speed of the car from which it would have been possible to have avoided the collision by braking to a stop. This can often demonstrate that the driver would only have been able to have avoided the collision if he had driven at an unreasonably low speed.

The model also has application in some vehicle-to-vehicle and bicycle-to-vehicle collisions where vehicles are hidden from each other until just before the collision.

INTRODUCTION

A common form of pedestrian accident involves the pedestrian suddenly emerging from behind a parked vehicle into the path of a passing car: Figure 1 illustrates the general problem. Such collisions very often involve young children and result in severe injuries. For example, in a study of pedestrian accidents in Australia, 388 of 3109 pedestrians killed or injured (or 1 in 8 pedestrian casualties) involved the pedestrian emerging from behind a stationary vehicle. Half (190) of the emerging pedestrians were less than 17 years of age; one quarter of the 758 pedestrian casualties under the age of 17 years suddenly emerged from behind parked vehicles (Road Safety & Traffic Management Directorate 1996).

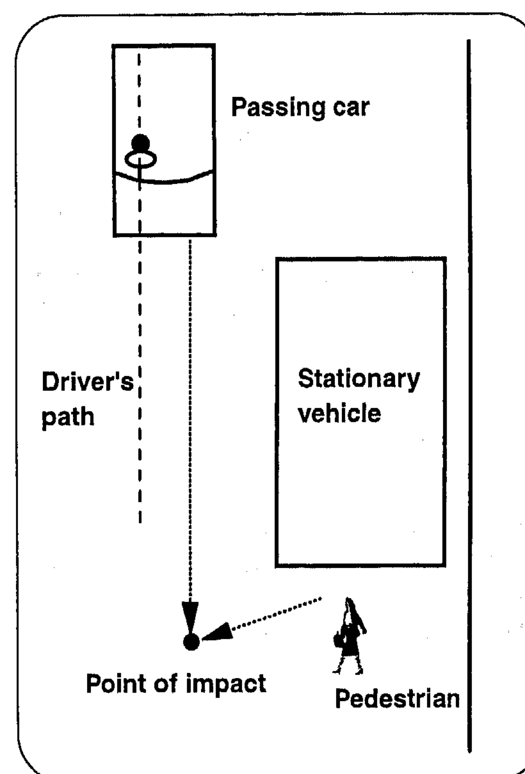


Fig. 1: The general problem of the emerging pedestrian

The time available to the driver of the car from when the pedestrian first came into view, to the point of impact, is an important consideration in apportioning

liability for the collision: could the driver have avoided the collision? The courts often see the driver as having the ultimate responsibility for driving at a speed and in a manner so as to be able to see and avoid pedestrians appearing in their paths. A collision with a pedestrian is very often seen to have at least partially resulted from the driver travelling at a speed which he should have known was too fast for the prevailing traffic conditions.

This paper is about a mathematical model ("EMERGE") which can be used to readily calculate the time available to the driver under various combinations of variables.

Graphical methods have been used to determine the time available to the driver in these types of crashes. The method typically involves establishing the paths followed by the pedestrian and the vehicle to the point of impact, and, starting from the point of impact, working backwards along the paths to establish the line of sight between the two at various times prior to impact. When the line of sight drawn on the diagram intersects the stationary vehicle, the time to impact can then be found. Figure 2 illustrates a typical graphical analysis.

One of the assumptions usually made in such analyses is that the pedestrian and the vehicle move at constant speeds. The assumption can in many (possibly most) cases be justified by statements to the effect *"it happened so fast - I didn't see him before I hit him"* (by the driver) and *"I never saw the car"* (by the pedestrian).

Graphical methods are tedious and prone to error. Where the speed is not constant (for example, where the driver commences braking) the process becomes even more tedious and time consuming as well as increasing the chance of error. If a range of speeds and/or positions/paths followed by the vehicle and pedestrian and positions of impact also have to be considered, the problems of the process are compounded.

For simpler cases, the graphical method is straightforward and easily understood. However, if a range of variables has to be considered and/or speeds are not constant during the crash sequence, the process is much less useful.

EMERGE

In principle, the mathematical model "EMERGE" is not dissimilar to the graphical method. The position of the stationary vehicle or other visual obstruction is established in relation to the paths of the vehicle and of the pedestrian. The point of impact is established in relation to the stationary vehicle and on the car striking the pedestrian. Once again, the analysis commences at the point of impact and works backwards to the time at which the pedestrian would have first come into the field of view of the driver. In the basic EMERGE, the speeds of the pedestrian and the car are assumed to be constant (that is, the pedestrian does not pause and the car driver does not brake before impact).

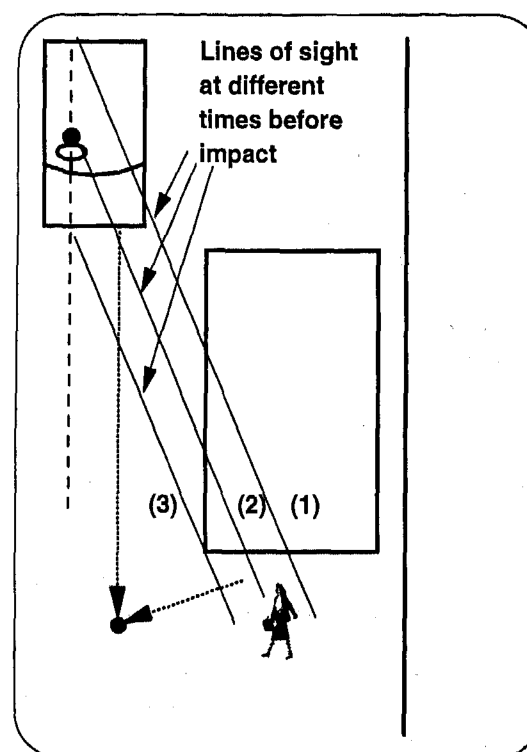


Fig. 2: The graphical method of establishing the theoretical lines of sight between the driver and the pedestrian at different times before impact. The line which first clears the stationary vehicle establishes how long before impact the pedestrian would be in the driver's field of view. For example, lines (1) and (2) are obscured by the stationary vehicle; line (3) establishes when the pedestrian is first in the driver's field of view.

The edge of the stationary vehicle beyond which the pedestrian comes into view is generally a corner of the vehicle and in EMERGE is denoted as the *"critical corner"*. At time t before impact, the lines of sight from the driver to the critical corner and from the pedestrian to the critical corner are calculated. The angle of each line of sight at time t before impact can thus be readily calculated: when the angles are

such that the two lines of sight are collinear, the pedestrian is first in the field of view of the car driver (Figure 3). The calculations involved are derived from simple two dimensional geometry.

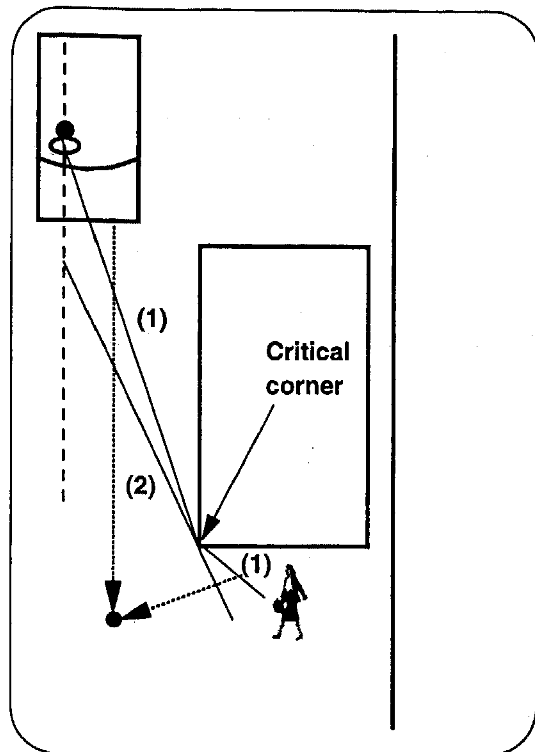


Fig. 3: The principle of EMERGE. The angles of the lines of sight at time t before impact between (a) the driver and the "critical corner" of the stationary vehicle and (b) the pedestrian and the "critical corner" are calculated. The time before impact when the pedestrian would be first in the driver's field of view is established when the two lines of sight are collinear. For example, lines of sight (1) are not collinear and hence the pedestrian is still hidden from the driver. Lines of sight (2) are collinear and the pedestrian is just in the driver's field of view.

Variables used in the basic EMERGE are

u_c speed of the car (assumed to be constant)

u_p speed of the pedestrian (assumed to be constant)

θ the angle between the paths of the pedestrian and the car

s_{lat} the lateral distance from the point of impact on the car to the driver

s_{long} the longitudinal distance from the point of impact on the car to the driver

y_{crit} the lateral distance from the "critical corner" on the stationary vehicle to the point of impact on the road

x_{crit} the longitudinal distance from the "critical corner" on the stationary vehicle to the point of impact on the road

t_{see} the time before impact when the driver can first see the pedestrian

The physical variables are illustrated in Figure 4.

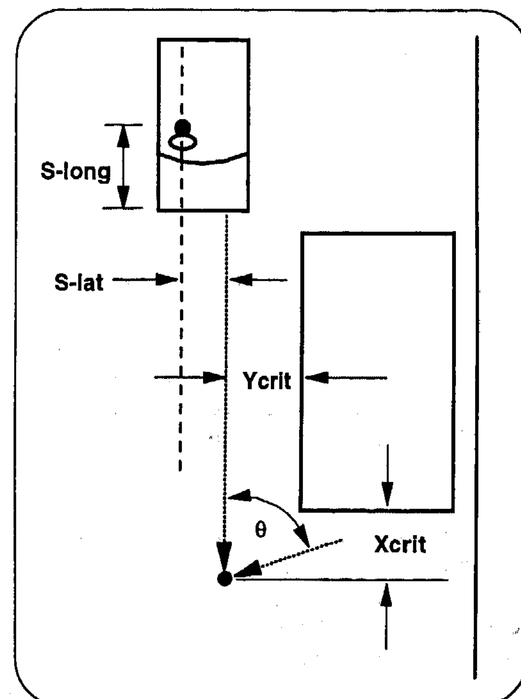


Fig. 4: Variables used in the basic EMERGE.

The analysis is based on Cartesian (or rectangular) coordinates, with the point of impact as the origin (Figure 5). At time t, the pedestrian is at position A and the driver at position E. The coordinates of A are $(u_p*t*\cos\theta,\ u_p*t*\sin\theta)$; the coordinates of E are $(u_c*t + s_{long},\ - s_{lat})$.

Within triangle ABC, angle BAC = α
$= \mathrm{atan}((x_{crit} - u_p*t*\cos\theta) / (u_p*t*\sin\theta - y_{crit}))$ ....(1)

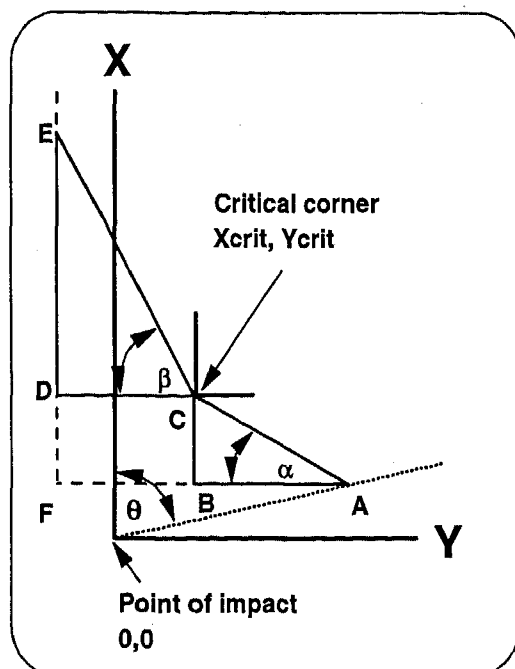


Fig. 5: The geometry of the basic EMERGE: when $\alpha = \beta$, the pedestrian is first in the field of view of the driver.

Within triangle CDE, angle DCE = β
$= \text{atan}((u_c{*}t + s_{long} - x_{crit}) / (s_{lat} + y_{crit}))$ ....(2)

When $\alpha = \beta$, the pedestrian is first in the field of view of the driver: hence the value can be established of the time available to the driver from when the pedestrian first came into his field of view to the point of impact.

Once the time available to the driver from when the pedestrian first came into his field of view to the point of impact (t_{see}) has been established for a given set of conditions (or range of conditions), that time can be compared with perception-reaction time data to assess whether the driver should have been able to have taken evasive action.

At time t_{see}, the line ACE in Figure 5 is straight, forming the hypotenuse of the right angle triangle AEF. The dimension AE is the sight distance when the driver and pedestrian could first see each other. The coordinates of A and E are as indicated above. The coordinates of F are ($u_p{*}t_{see}{*}\cos\theta$, $-s_{lat}$). From Pythagoras' Theorem,

$$AE^2 = EF^2 + AF^2$$
$$= (u_c{*}t_{see} + s_{long} - u_p{*}t_{see}{*}\cos\theta)^2 + (s_{lat} + u_p{*}t_{see}{*}\sin\theta)^2$$

and hence the value of the sight distance can be calculated.

THE INFLUENCE OF DIFFERENT VARIABLES ON THE TIME AVAILABLE TO THE DRIVER

Each of the variables involved has different effects in the calculation of the time before impact when the driver can first see the pedestrian (time t_{see}): this is illustrated by altering individual variables in a basic example.

The basic example involves a Toyota Camry car travelling at 40 km/h (about 25 mph), passing a large stationary vehicle with a clearance of 1.2 m (about 47 inches), and striking a pedestrian who is crossing at a speed of 3 m/s (about 10 fps) at right angles 1 m (about 39 inches) in front of the stationary vehicle. The pedestrian is struck by the nearside (car passenger's side) headlight: the lateral distance from the stationary vehicle to the point of impact is 1.55 m (about 61 inches). The longitudinal and lateral distances from the point of impact on the car to the driver are 2.4 m (about 94 inches) and 0.9 m (about 35 inches) respectively. Solving the equations indicates that the pedestrian would have first come into the driver's field of view about 0.62 seconds before impact. At that time, the sight distance between the driver and the pedestrian would have been about 9.6 m.

The time available to the driver from when the pedestrian first came into his field of view to the point of impact can now be compared with perception-reaction time data to assess whether the driver should have been able to have taken evasive action.

Often the speeds of the car and the pedestrian are only known within ranges. How critical are the speeds? Figure 6 illustrates the effect on the time available to our driver from the basic example, by considering pedestrian speeds of 1.5 m/s (within the range of young adult walking speeds), 3 m/s (within the range of young adult jogging speeds) and 4.5 m/s (within the range of primary or elementary school student running speeds) against car speeds of 10 km/h to 70 km/h (about 6 mph to 43 mph). For a given pedestrian speed, the speed of the car in this example makes very little difference to the time available to the driver: above a car speed of about 20 km/h, the time values in this example rapidly become asymptotic.

To illustrate the effect of a less sudden emergence from behind the stationary vehicle, the analysis shown in Figure 6 is repeated in Figure 7, but in this latter case, the car and the pedestrian have

double the clearance from the stationary vehicle. The car in this case passed the stationary vehicle with a clearance of about 2.4 m and the pedestrian crossed 2 m in front of the stationary vehicle: the impact was with the Camry's nearside headlight, and thus the point of impact was about 2.75 m laterally from the stationary vehicle. While the time available to the driver to see the pedestrian before impact is much greater, the general asymptotic shapes of the curves for different speeds of the pedestrian are similar.

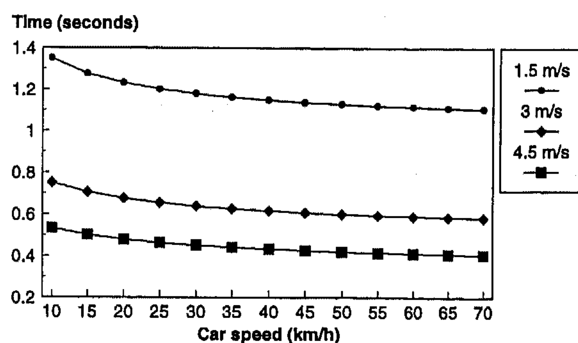


Fig. 6: The effect on time available to the driver of different pedestrian and car speeds, using the variables in the basic example

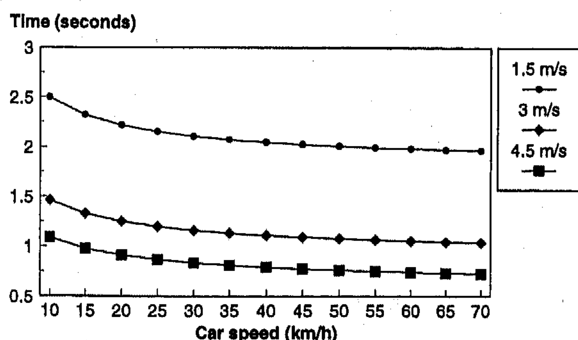


Fig. 7: As for Fig. 6, except that the car and the pedestrian have a greater clearance from the stationary vehicle (see text).

In most real world situations where car speeds are not extremely low, the speed of the pedestrian has a far greater influence than the car speed on the time available to the driver to see and possibly avoid the pedestrian: except at very low car speeds, the speed of the pedestrian essentially determines the time available to the driver (all other variables constant).

As illustrated by comparison of Figures 6 and 7, the clearance of the pedestrian from the visual obstacle is an important variable: increasing the clearance increases the time available to the driver.

Pedestrians do not always cross a road at right angles. For a given set of circumstances, there is an angle at which the time is at a minimum, and where any other angle results in a longer time for the driver. As Figure 8 illustrates, the angle can be an important factor in the time available to the driver: a pedestrian walking slightly away from the oncoming car (angle between the paths being less than 90 degrees) provides the least time, while a pedestrian walking at an angle generally towards the oncoming car (angle between the paths being more than 90 degrees) provides the most time.

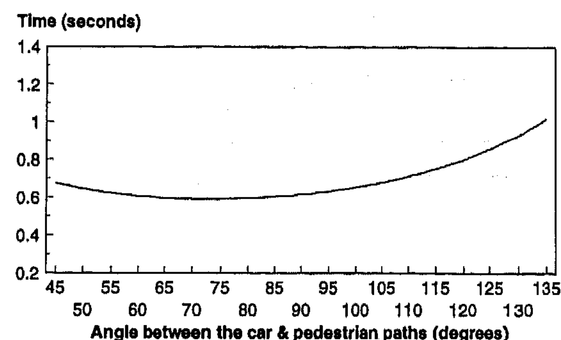


Fig. 8: The time available to the driver to see and avoid the pedestrian, with varying angles between the paths followed by the car and the pedestrian (car speed of 40 km/h, pedestrian speed of 3 m/s).

The dimensions of the impacting car and where on the front of a car the collision occurs has an effect on the time available to the driver: but how much? Figure 9 shows the time for the basic example, varying the position of impact on the front of the car from the extreme nearside front corner of the car (1.4 m laterally from the driver's eye position) to directly in front of the driver, allowing for different dimensions from the driver's eye position longitudinally to the front of the car.

The closer the driver sits to the front of the car, the more time available to the driver to see and possibly avoid the pedestrian, but, as shown in Figure 9, the effect is negligible. An impact on the car nearside corner provides more time to the driver than an impact directly in front of the driver: this

counter-intuitive result is related to the angles of sight lines around the visual obstacle. In any event, the effect on time of where on the front of the car the impact occurs is very small.

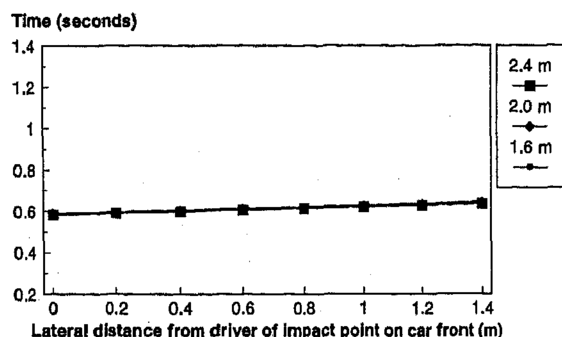


Fig. 9: The time available to the driver to see and avoid the pedestrian in a frontal collision, with different dimensions from the driver's eye position and different impact positions across the front of the car (car speed of 40 km/h, pedestrian speed of 3 m/s). All three lines are essentially superimposed on each other.

A SAFE PASSING SPEED

Courts are sometimes inclined to the view that a driver should have been travelling at a slower speed prior to the collision. And that by travelling at a slower speed, the driver could have taken successfully evasive action. In particular, courts often take the view that drivers should have been able to have braked to a stop before reaching the point of impact.

For any given situation, at what speed should a prudent driver pass a stationary vehicle such that if a pedestrian walks or runs out, the driver can react and stop the car before hitting the pedestrian?

By including in the motion of the car provision for a driver reaction time and a deceleration rate after the driver brakes, EMERGE can calculate the speed at which stopping just prior to the point of impact would have been possible. The basic principle illustrated in Figure 3 still applies, but for a given speed of the car at impact (v_{impact}, usually = 0), the speed of the car as it passes the stationary vehicle (u_{crit}) is a function of

- r the reaction time of driver, and
- a the slowing rate of the car prior to impact

The critical time when the pedestrian and driver are first in each other's field of view, t_{crit}, has the following relationship (utilising the general equation $v = u + a*t$, from Newton's Laws of Motion, where v is final velocity, u is initial velocity, a is uniform acceleration and t is time).

$$t_{crit} = r + (v_{impact} - u_{crit}) / (-a) \quad \text{.........................(3)}$$

where r and a are nominated for the calculation based on test or research data, and v_{impact} is nominated (usually = 0).

The geometry of the sight lines from the driver and the pedestrian is again as shown in Figure 5. The driver and the pedestrian are first within view of each other at t_{crit} before impact, when the angles α and β in triangles ABC and CDE are equal to each other.

At time t_{crit}, the coordinates of A are ($u_p*t_{crit}*\cos\theta$, $u_p*t_{crit}*\sin\theta$).

The Y coordinate for E is $-s_{lat}$. The X coordinate for E is the sum of the distance travelled at the critical speed during the reaction time ($u_{crit}*r$); the distance travelled during the slowing phase($[v_{impact}^2 - u_{crit}^2] / 2*[-a]$), which is found by utilising the general equation $v^2 = u^2 + 2*a*s$ (from Newton's Laws of Motion, where v is final velocity, u is initial velocity, a is uniform acceleration and s is distance over which the acceleration takes place); and the longitudinal distance from the driver to the point of impact on the car (s_{long}). Thus the coordinates of E are ($u_{crit}*r + [v_{impact}^2 - u_{crit}^2] / 2*[-a] + s_{long}$, $- s_{lat}$).

Within triangle ABC, angle BAC = α

$$= \text{atan}((x_{crit} - u_p*t_{crit}*\cos\theta) / (u_p*t_{crit}*\sin\theta - y_{crit})) \quad \text{..........(4)}$$

Within triangle CDE, angle DCE = β

$$= \text{atan}((u_{crit}*r + [v_{impact}^2 - u_{crit}^2] / 2*[-a] + s_{long} - x_{crit}) / (s_{lat} + y_{crit})) \quad \text{..........(5)}$$

The function in equation (3) is then substituted for t_{crit} in equation (4). From equations (4) and (5), the value of u_{crit} can be found which results in $\alpha = \beta$.

The example used earlier of the Toyota Camry passing the stationary vehicle with 1.2 m clearance encountering a pedestrian passing 1 m in front of the stationary vehicle at 3 m/s is used to illustrate the application. Figure 10 shows the effect on the safe passing speed of the relationship between reaction time and car slowing rate. In that particular scenario, the reaction time dominated the safe passing

speed, but even with a fast reaction time, the safe passing speed would be unrealistically low: at best only about 3 km/h. Considering that a young adult's *walking* speed is about 6 km/h, it is obvious that any realistic speed of the Camry would be too high!

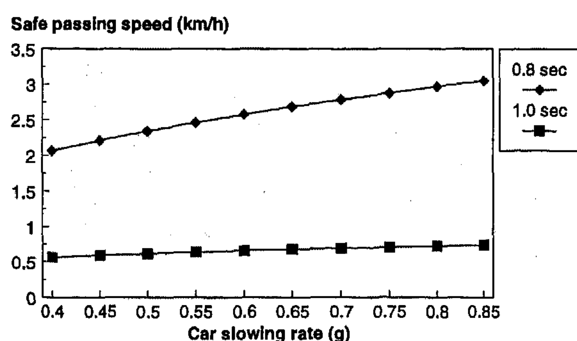


Fig. 10: The safe passing speed where the driver has a reaction time of 0.8 or 1.0 seconds (pedestrian speed of 3 m/s).

DATE FOR EMERGE

The required data for input in EMERGE can be obtained from different sources.

The point of impact on the car can be established by a variety of well documented means, often from evidence of damage on the car. The longitudinal distance from the point of impact on the car to the driver's position is sometimes available from databases (assuming the impact is on the front of the car): if not, measure it! In any event, errors in the distances from the driver to the point of impact on the car have very little influence on the calculations undertaken by the model.

The speed of the car can sometimes be established by witnesses and/or by the analysis of the post impact movement of the pedestrian and/or of the car. Unless the speeds are very low, the precise car speed is relatively unimportant and can often be addressed using speed ranges.

The speed of the pedestrian has an important effect on the outcome of the calculations. It is also usually more difficult to establish than the car speed, but it can at least usually be established as *walking, jogging* or *running* and, sometimes, (for example) *walking briskly* or *running fast.* There are published data based on population information from a variety of sources which can be used in these calculations (for example, Anon 1994; Austroads 1995; Coffin and Morral 1995; Eubanks 1994; Thompson 1991). An example from Eubanks (1994) is that of the running speeds for a sample of 42 eight year old boys: the 15th percentile speed was 4.17 fps (1.27 m/s); the 50th percentile speed was 4.99 fps (1.52 m/s); and the 85th percentile speed was 6.52 fps (1.99 m/s).

Typically a range of pedestrian speeds will be used in the analyses using EMERGE.

The point of impact in relation to the obscuring vehicle or object needs to be established, together with the paths taken by the car and the pedestrian. This information can usually be obtained from debris, objects dropped by the pedestrian and witness statements. Ranges are often used.

One of the most important parameters in considering the time available to the driver to avoid the pedestrian, and in calculation of the safe passing speed, is the driver's reaction time. A detailed study of reaction times is beyond the scope of this paper, but some discussion is warranted in considering the choice of reaction time for analysis.

REACTION TIME

The "reaction time" of a driver is the time taken to *detect* the presence of a person or an object; to *identify* that the person or object represents a potential or actual hazard; to *decide* the action to be taken in response to the hazard; and to *react* or *respond* by starting to take that action (Olson 1989). Reaction time is sometimes known as "perception-reaction" time (Triggs and Harris 1982) or "perception-response time" (Olson 1989).

Reaction time has probably been one of the most studied aspects of driver behaviour (Olson 1989). In essence, the main variables between studies have the degree of warning given to (or expectation by) the subjects of an event; the nature of the stimulus to which subjects were responding; and the complexity of background activities within which the stimulus appears.

The simplest studies involved subjects who were fully aware that they were engaged in an experiment. They were also often told to respond to a light or a sound, or similar discrete stimulus, by pressing a button or other simple action (Olson 1989, Sens et al 1989, Triggs and Harris 1982). In some instances, these experiments were testing anticipation. The net result was that reaction times achieved in such simple experiments were usually

very short, often approaching zero (Triggs and Harris 1982). While monitoring of the driving task overwhelmingly involves visual stimuli (Evans 1991), many of the studies involved responding to auditory stimuli. Whether the results from many of these simple experiments can be applied to driving situations is questionable (for example, Triggs and Harris 1982).

At the other end of the spectrum were experiments which involved monitoring of drivers in actual traffic situations: these subjects were not aware that they were being observed, nor that they were participating in experiments. These types of experiments involved observing or creating hazards on public roads where the behaviour of passing motorists was monitored. The hazards were such that there were several possible actions which motorists might take: that is, the decision making process was more complex than simply deciding whether or not to press a button (Triggs and Harris 1982). Experiments included

- responding to the yellow phase of traffic lights (mean time 1.30 seconds, standard deviation of 0.60 seconds, 85th percentile** 1.8 seconds, but with considerable variation in results from site to site - Wortman & Matthias 1983; mean time 1.14 seconds - Gazis et al 1960)

  ** the 85th percentile is a common criterion in traffic engineering and is the level at which 85 % of the population has responded

- responding to a car suddenly braking (mean time 1.45 seconds - Allen Corporation 1978; mean time 1.38 seconds, 85th percentile 1.95 seconds - Sivak et al 1979; mean time 1.25 seconds, 85th percentile 1.90 seconds - Sivak et al 1981; mean time 0.92 seconds; 85th percentile 1.26 seconds - Triggs and Harris 1982)

- a car door suddenly opening and responding by steering away (mean time 1.5 seconds - Summala 1981a, 1981b)

- amphometers (mean times 1.75 to 2.45 seconds; 85th percentile 3.40 to 3.60 seconds - Triggs and Harris 1982)

- railway crossing lights (mean times 1.16 to 1.77 seconds; 85th percentile 1.50 to 2.53 seconds - Triggs and Harris 1982)

- car under repair (mean time 1.02 seconds; 85th percentile 1.50 seconds - Triggs and Harris 1982)

- parked police vehicle (mean time 2.37 seconds; 85th percentile 2.80 seconds - Triggs and Harris 1982)

There have also been a large number of studies using techniques with conditions which were in between simple tests with alerted subjects on one hand, and the observations of actual traffic responses on the other (Olson 1989, Sens et al 1989, Triggs and Harris 1982). The "in between" studies tended to result in reaction times slower than for the simple tests, but faster than those in actual traffic (Sens et al 1989).

Studies have suggested that reaction time for older drivers is longer than for younger drivers (for example, American Automobile Association 1966). Other studies showed no major difference with age (for example, Olson et al 1984).

Alcohol is known to lengthen reaction times (Fricke 1990; Sens et al 1989). Fatigue is also known to lengthen reaction times to varying extents (for example, Muto & Wienville 1982), and there is also evidence that fatigue exacerbates the effects of alcohol (Corfitsen 1982).

Traffic engineering practice in many countries has been to design for a reaction time of 2.5 seconds, although research in several countries has resulted in proposals for much longer times to be used (Triggs and Harris 1982).

Overall, the published research has shown that reaction times in actual traffic are generally much longer than those measured in test environments which involve at least a degree of artificiality. While there will be considerable variation between individuals and circumstances, a reaction time of 1 second can be considered to be fast to very fast in most cases.

DISCUSSION

EMERGE provides a easy method for calculation of both the time from when the pedestrian came into view to the collision, and of the safe passing speed from which the driver could have braked to a stop. EMERGE can be readily used through a spreadsheet program (such as Lotus 123 or Excel). The speed and ease of use means that an envelope of ranges of variables can be very quickly and accurately compiled.

Some of the variables used in EMERGE are more important than others: in particular, the most important are pedestrian speed and where in relation to the obscuring vehicle or object the collision occurred. The angle between the paths followed by the pedestrian and the car can be important in some cases. Other variables are such that even very wide ranges of values in the input data will have little effect on the results.

The safe passing speed calculation is often a very powerful tool for demonstrating that any driver passing at a reasonable or even low speed could not have avoided the collision. The safe passing speed calculation is sensitive to driver reaction time and car deceleration rate.

The safe passing speed concept can also be applied where a car has braked and skidded before impact, and where the impact speed can be calculated through the pedestrian's post impact kinematics (for example, Aronberg and Snider 1994; Searle and Searle 1983; Searle 1991; Szydlowski and Jenkins 1991) and/or post impact skidding of the car. Consider, for example, where a car had braked and skidded 12 m before impact and continued for another 8 m from impact to rest. The slowing rate when skidding was, say, 0.7 g. From the skidding distances, the impact speed was about 38 km/h and the pre skidding speed was about 60 km/h. By inserting into the model an impact speed of 38 km/h and a deceleration of 0.7 g, the reaction time can be varied until a "safe passing speed" of 60 km/h is found. That will establish the reaction time likely to have been applicable in this instance, as well as the total time and distance to impact. One can then consider whether the reaction time was reasonable, or indicated an inattentive driver.

While EMERGE was developed for collisions involving emerging pedestrians, it has also proved useful for other types of "emerging" crashes, including both bicycle crashes and car crashes.

EMERGE can also be used where, for example, a pedestrian crossing a road "disappears" behind a large vehicle stopped in traffic before reappearing immediately prior to collision. In that case, the process involves two calculations of times before impact: the first in relation to the critical corner which determines when the pedestrian disappears; and the second in relation to the (diagonally opposite) critical corner when the pedestrian reappears. Figure 11 illustrates the application.

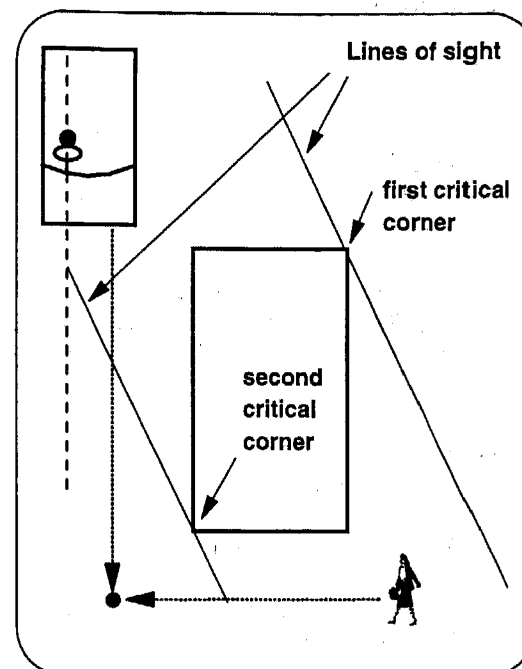


Fig. 11: The two critical corners where a pedestrian "disappears" behind a large vehicle in traffic, before reappearing just before collision.

SUMMARY

A mathematical model, "EMERGE", has been developed which can be used through a spreadsheet program. The model enables the rapid calculation of the time available to a driver from when a pedestrian emerges into his path, until colliding with the pedestrian. It is a tool for assessing whether the driver had sufficient time to take evasive action. It can also be used to calculate the safe passing speed from which the driver could have braked to a stop before colliding with the pedestrian.

EMERGE is mainly sensitive to pedestrian speed, and to the distances of the car and the pedestrian from the obscuring object (usually a stationary vehicle). It is not very sensitive to the car speed, unless the car speed is very low. The safe passing speed calculation is sensitive to the driver's reaction time and to the car slowing rate when braking. Other variables usually have relatively little influence on the outcomes.

It is also a useful tool for collisions involving emerging bicycles and cars.

REFERENCES

Allen Corporation of America (1978) "Field Validation of Taillights - Report on Phase 1: Pilot Testing". Alexandria, Virginia. US Department of Transportation, NHTSA. Contract No DOT-HS-7-01756. In Olson (1989).

American Automobile Association, Traffic Engineering and Safety Department (1966) "Reaction Time in Relation to Age". Report No. 69. In Olson (1989).

Anon (1994) Tabulation of walking and running speeds for pedestrians in the City of San Diego. *Accident Reconstruction Journal*, Vol. 6, No. 2, March/April 1994, p 38.

Aronberg, Ralph, and Snider, Andrew A. (1994) "Reconstruction of Automobile/Pedestrian Accidents Using CATAPULT". SAE paper 940924, in "Accident Reconstruction: technology and Animation IV", publication SP 1030.

Austroads (1995). "Guide to Traffic Engineering Practice, Part 13: Pedestrians". Austroads, Sydney, Australia

Coffin, Ann, and Morral, John (1995). "Walking Speeds of Elderly Pedestrians at Crosswalks". *Transportation Research Record 1487.* Transportation Research Board.

Corfitsen, M. T. (1982) "Increased Viso-Mororic Reaction Time of Young, Tired, Drunk Drivers". *Forensic Science International.* 20(2), 121-125, 1982.

Eubanks, Jerry J (1994). "Pedestrian Accident Reconstruction". Lawyers & Judges Publishing Co., Tucson, Arizona

Evans, L. (1991) "Traffic Safety and the Driver". Van Nostrand Reinhold, New York.

Fricke, Lynn B. (1990) "Traffic Accident Reconstruction. Volume 2 of The Traffic Investigation Manual". Northwestern University Traffic Institute, Evanston, Illinios.

Gazis, D., Herman, R., and Maradudin, A. (1960) "The Problem of the Amber Signal Light on Traffic Flow." *Operations Research*, 1960, 8, 112-132. In Triggs and Harris (1982).

Muto, W. H., and Wienville, W. W. (1982) "The Effect of Repeated Emergency Response Trials on Performance During Extended Duration Simulated Driving". *Human Factors.* 24(6), 693-698.

Olson, Paul L. (1989) "Driver Perception Response Time". SAE paper 89731, in "Motor Vehicle Accident Reconstruction: Review and Update", publication SP 777.

Olson, P. L., Cleveland, D. E., Fancher, P. S., Kostyniuk, L. P., and Schneider, L. W. (1984) "Parameters Affecting Stopping Sight Distances". Washington DC. The National Cooperative Highway Research Program. Report 270. In Olson (1989).

Road Safety & Traffic Management Directorate (1996). Special tabulations of accident data held in the New South Wales accident database. New South Wales Roads & Traffic Authority, Sydney, Australia.

Searle, John A., and Searle, Angela (1983) "The Trajectories of Pedestrians, Motorcycle, Motorcyclists, etc., Following a Road Accident". SAE paper 831622, In "Accident Reconstruction Technologies: Pedestrians & Motorcyclists in Automotive Collisions", publication PT 35.

Searle, John A. (1991) "The Physics of Throw Distance in Accident Reconstruction". SAE paper 930659. In "Accident Reconstruction: Technology and Animation III", publication SP 946.

Sens, Michael J., Cheng, Philip H., Wiechel, John F. and Guenther, Dennis A. (1989) "Perception/Reaction Time Values for Accident Reconstruction". SAE paper 89732, in "Motor Vehicle Accident Reconstruction: Review and Update", publication SP 777.

Sivak, M., Olson, P. L., and Post, D. V. (1979). "Evaluation of Experimental (including High-Mounted) Configurations of Brake Lights in Actual Traffic". Ann Arbor, Michigan. The Highway Safety Research Institute. Report No. UM-HSRI-79-87.

Sivak, M., Olson, P. L., and Farmer, K. M. (1981). "High-Mounted Brake Lights and the Behaviour of Following Drivers". Ann Arbor,

Michigan. The Highway Safety Research Institute. Report No. UM-HSRI-81-31.

Summala, H. (1981a) "Driver's Steering Reaction to a Light Stimulus on a Dark Road". *Ergonomics*, 1981, 24, 125-131.

Summala, H. (1981b) "Driver/Vehicle Steering Response Latencies". *Human Factors*, 1981, 23, 683-692. .

Szydlowski, Wieslaw, and Jenkins, P. E. (1991) "Modeling of a Sliding Phase in Accident Reconstruction". SAE paper 930655. In "Accident Reconstruction: Technology and Animation III", publication SP 946.

Thompson, T (1991) "Pedestrian Walking and Running Velocity Study". *Accident Reconstruction Journal,* Vol. 3 No. 2, March/April 1991, pp 28-9.

Triggs, T. W., and Harris, W. G. (1982) "Reaction Time of Drivers to Road Stimuli". Human Factors Report HFR-12, Department of Psychology, Monash University, Melbourne, Australia

Wortman, R. H. and Matthias, J. S. (1983) "An Evaluation of Driver Behaviour at Signalized Intersections". Final Report. Arizona Transportation and Traffic Institute, Tucson, Arizona. In Olson (1989).

SAE #970962
Reviewer's Discussion
by David Sallmann, Rudny & Sallmann Engineering, Ltd.
Avoiding the Emerging Pedestrian: a Mathematical Model
Rodney G. Vaughan, Author

The author presents a unique mathematical model for analyzing timing in limited sight distance accidents. EMERGE can be used not only with spreadsheet software but also on calculators with equation solving routines. The discussion focuses on the emerging pedestrian case, but the model can be utilized equally well for intersection collisions between vehicles where limited sight distance exists. Driver reaction times can be computed using the model, after impact speeds and speed before braking has been determined.

The case described in this paper involves a right hand drive vehicle with the pedestrian approaching from the left. The model also applies for left hand drive vehicles and for far side impact situations. It should be noted, however, that the lateral distance from the point of impact on the car to the driver (S_{lat}) is a negative value when the point of impact is farther from the emerging vehicle/pedestrian than is the driver's position.

970963

Motorcycle Slide to Stop Tests

Christopher J. Medwell, Joseph R. McCarthy, and Michael T. Shanahan
BTS Consulting Engineers

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper reports the results of testing conducted to determine the post-upset sliding friction factor on asphalt pavement for a sport motorcycle equipped with a full coverage fairing. A literature review revealed limited data regarding this type of motorcycle. Therefore, these tests were conducted and are reported in the interest of adding to the database available to accident investigators.

BACKGROUND

Motorcycle accident reconstruction often involves the determination of the speed of a motorcycle at the beginning of an accident sequence. In virtually all serious motorcycle accidents, the involved motorcycle will, at some point, fall to the ground and slide along the road surface for some distance. Accurate calculations of the motorcycle's speed prior to it falling over depend on evaluating the energy dissipated by the sliding motorcycle. This energy calculation, in turn, depends on the determination of an appropriate friction factor.

A brief review of the published literature reveals some data sources which address the friction factor of motorcycles sliding on their sides. For example, Day and Smith (1) reported on motorcycle friction factors in 1984. Various accident reconstruction handbooks (2,3,4) either refer to Day and Smith's paper or publish similar values from other testing. Typical friction factors for motorcycles sliding on their sides on asphalt pavement range from 0.35 to 0.75. Values for softer surfaces such as grass, gravel, or sand are higher, ranging from 0.65 to 1.1 or more. Refer to the Appendix of this paper for a review of the published data.

Most of the data originate from tests performed in the mid 1980's. Some dependence on speed was noticeable in some testing, with higher speeds generally resulting in lower friction values. Some dependence on motorcycle type was also observed, with lower friction values obtained from tests of motorcycles with broad or rounded surfaces such as crash bars or fairings. Friction factors for racing style sports motorcycles with full coverage fairings were unavailable, since this type of motorcycle was not generally available for street use.

A few more recently published test results for modern motorcycles (5,6,7) indicate that fairing-equipped sport motorcycles do tend to slide on their sides at lower friction values than do conventional motorcycles. Results for these tests range from 0.26 to 0.42 for high speed tests (80 to 100 km/hr). Data for current motorcycle designs, however, continue to be sparse. Therefore, a decision was made to undertake the tests reported herein.

PROCEDURE

The motorcycle utilized in the slide tests was a 750cc 1992 Kawasaki Ninja ZX-7. The motorcycle had approximately 4400 km on its odometer at the time of the testing. The Ninja was a sport-style street motorcycle equipped with triple disc brakes, in-line 4-cylinder engine, aerodynamic full fairing covering both sides of the motorcycle, cast aluminum wheels with relatively wide tires and low mounted handlebars. The fairing presented a large, smooth surface to the roadway, much like a toboggan. The handlebars and foot pegs, made of aluminum, protruded less from the side of the motorcycle than corresponding steel components on other classes of street motorcycles. These factors were expected to produce a lower friction factor than that of a conventional motorcycle. The motorcycle was also equipped with a factory 4-into-1 exhaust with a single large muffler canister on the right side.

This motorcycle had been involved in a single vehicle accident prior to our tests. The accident involved a long slide to a stop on the motorcycle's right side. The motorcycle's left side was undamaged. In order to prepare the right side of the motorcycle for our tests, we undertook a number of repairs prior to the test. These repairs included:

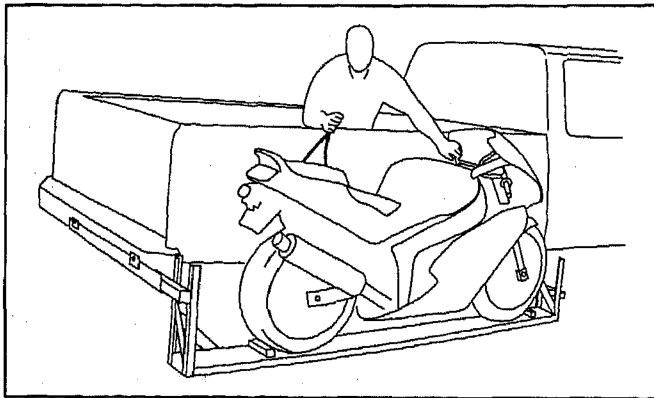
- Rotation of the muffler on the exhaust pipe so an undamaged surface of the muffler was exposed to the road surface during our test.
- Replacement of the right side lower fairing assembly.
- Replacement of the right side driver's foot peg bracket.

- Replacement of the upper fairing support bracket.

- Installation of the handlebar, turn signals, and side reflector from the left side of the motorcycle onto the right side.

With the exception of the right outboard edge of the upper fairing, the right side of the motorcycle was undamaged immediately prior to our testing. The fairing had sustained slight cracking and scuffing during the original accident which we did not deem necessary to repair prior to our tests.

The tests were designed to approximate, as closely as possible, the motion of a motorcycle falling over from an upright position. The motorcycle was positioned upright on a fabricated platform mounted on the right side of a pickup truck. (See Figure 1.) The height of the platform was adjusted so that its underside was as close as possible to the roadway surface. This test setup resulted in the motorcycle tire contact surface being approximately 90mm above the roadway. The motorcycle was held upright by an assistant riding in the bed of the pickup truck. The truck was accelerated to the test speed, then the motorcycle was released and allowed to fall over sideways onto the road surface.


Figure 1. Test Vehicle Platform

The tests took place on a straight, level roadway located on the property of a local airport. The road surface consisted of asphalt pavement approximately 6.5 m in width. The surface was weathered and had not been open to vehicular traffic for some time. The roadway aggregate projected beyond the level of the black binder, giving a light colour and some coarseness to the surface as opposed to a smooth pavement surface. The road surface was bordered on both sides by grass-covered shoulders. The grass was quite long and well-established. The surface underlying the grass was relatively smooth.

Tests were conducted at a speed of 80 km/hr. The test speed was monitored to within 1 km/hr by radar aimed at the pickup truck. The radar device was a Muniquip MDR2 and was calibrated with a 50 km/hr and a 95 km/hr tuning fork just prior to the tests.

The motorcycle was photographed before and after each test and the damage documented. The tests were video taped by cameras located both in the pickup truck and on the side of the roadway.

After the motorcycle slid to a stop, the sliding distance was measured. Measurements were taken from the location of the first evidence of the motorcycle's touchdown on the road surface to the centre of the position of the motorcycle at rest. A total of four tests were conducted, three on the right side of the motorcycle and one on the left.

OBSERVATIONS

RESULTS OF TEST NO. 1 - During test no. 1, the pickup truck was driven parallel to the roadway, and the motorcycle was released at a location near the centre line. Gouge marks created by the right foot peg and front axle, along with green paint transfers from the fairing, were the first evidence of touchdown. Also at touchdown, the front tire and wheel rim of the motorcycle left a deep scrape and a black tire scuff in the asphalt. A radial deformation of the cast aluminum rim of the front wheel was obvious after the test. The motorcycle, with the exception of its front wheel, slid smoothly on the asphalt surface. The front wheel, after the initial touchdown, immediately bounced up and then back down, leaving another black scuff mark at a second location. The tire then came down again and bounced lightly several more times, leaving black scuff marks to the road surface during the contact time. The intermittent black tire scuff marks were visible for the majority of the length of the slide on the paved surface.

The motorcycle slid in a direction about 4 or 5 degrees to the right of the travel path of the pickup truck and test platform. The motorcycle slid for 41 m on the asphalt surface before it left the right edge of the roadway and went onto the grass. During the "on-pavement" portion of the slide, the motorcycle was relatively stable with the front wheel leading. Evidence from the slide along the road surface consisted of intermittent black tire scuff marks from the front tire, as previously described. There were also scrape marks and scuff marks from the various aluminum components of the motorcycle contacting the road surface. Beyond the paint transfers at initial touchdown, there was no other evidence on the asphalt of the fairing contact. After the 41 m on the asphalt, the motorcycle continued on the grass shoulder to a point 56 m from its original touchdown point before coming to rest. While on the grass, it tore up a substantial quantity of turf which became lodged in the fairing. The motorcycle rotated counterclockwise approximately ½ of a revolution about the front wheel during the final 15 to 20 m of the slide. It came to rest with its front wheel pointing in the direction from which it had come.

In addition to the deformation of the front wheel rim, the motorcycle sustained abrasions to the lower right side fairing assembly, the lower edge of the right

fork tube, the right handlebar, the brake lever, the brake lever pivot area, the right side of the upper fairing, the driver's foot peg, the passenger's foot peg, and the outboard casing of the muffler. The right mirror broke from its mount, and the right front turn signal came out of its mounting area. With the exception of the damage to the front wheel, all of these damages were consistent with the expected damage for this kind of slide.

RESULTS OF TEST NO. 2 - Test no. 2 was also conducted at 80 km/hr. The motorcycle's initial sliding trajectory was similar to test no. 1. It came into contact with the grass 43.2 m from the touchdown point. During the slide on the asphalt, the motorcycle rotated slightly in a clockwise direction. As a result, the front tire of the motorcycle came into contact with the grass first. This occurred after the motorcycle rotated approximately 90 degrees to an orientation approximately perpendicular to the roadway, with the tires of the motorcycle leading the slide. As the front tire entered the grassy surface, turf was dug up and the motorcycle began to rotate rapidly clockwise. As it entered the grass, the motorcycle rotated rapidly while traveling approximately parallel to the edge of the road (i.e., having changed direction slightly from its original path towards the right). The motorcycle rotated approximately 540 degrees before coming to rest with its front wheel pointed towards the direction of the start of the slide. The total slide distance was 63 m, 43 m on asphalt and 20 m along the asphalt/grass shoulder interface. At rest, the handlebars and front wheel were located over the asphalt surface with the remainder of the motorcycle on the grass.

The evidence on the road surface from test no. 2 was similar to that from test no. 1 except the initial front wheel contact with the asphalt was less severe. A black scuff from the front tire was visible for the majority of the slide distance on the asphalt. Various scrapes and short duration scratches to the asphalt were noted along the entire distance of the slide. There was sufficient evidence to note the beginning of the slide, the direction of the slide, the location of the interface between the asphalt and the grass, and the distance the motorcycle slid on the edge of the grass.

Additional damage to the same surfaces of the right side of the motorcycle was noted. There was new evidence of additional scuffing to the front tire and wheel rim from this test; however, there was no deformation of the wheel rim.

RESULTS OF TEST NO. 3 - For test no. 3, the pickup truck was driven on a 4 or 5 degree angle to the left at the time of the release so that the motorcycle would travel parallel to the road during its slide. This method produced a slide straight down the roadway a distance of 69.5 m. The motorcycle was initially stable with the front wheel leading, but rotated 180 degrees counterclockwise toward the end of the slide and came to rest with the front wheel pointing back toward the start of the slide. Similar evidence on the road surface was produced during this test. The initial front wheel contact with the asphalt was minimal. The black tire scuff from the front tire was present throughout the length of the slide. Additional damage to the same surfaces of the right side of the motorcycle was noted.

RESULTS OF TEST NO. 4 - The last test was conducted by releasing the motorcycle onto its left side. Since the platform was mounted to the right side of the pickup box, this involved releasing the motorcycle with its rear wheel leading. The absence of a muffler on the left side of the motorcycle exposed more of the plastic fairing to the road surface and allowed the end of the rear axle to contact the asphalt.

During the initial portion of the slide, the motorcycle rotated clockwise 180 degrees to a position with the front tire leading. It was stable in this attitude until near the end of the slide, when it began to rotate counterclockwise slightly, just before it came to rest.

The pickup truck was again driven on an angle so that the total slide distance was on the paved surface as in test no. 3. The test speed was 80 km/hr, and the sliding distance was 86.3 m.

Short gouges were created on the road surface at touchdown by the handlebar end and the rear axle. The remainder of the slide distance exhibited short scratches and scuff marks similar to those observed in the other tests.

The motorcycle exhibited abrasions to a larger area of the lower fairing assembly than had been the case on the right side tests. At the rear, the absence of a muffler canister had allowed abrasions to the rear fairing area just below the seat and to the protruding end of the rear axle.

ANALYSIS

TEST NO. 1 - Based on an 80 km/hr release speed, and the 56 m slide length, the average friction factor for the slide was 0.45. Because of the grass torn and dug up during the portion of the slide on grass, it was obvious that the motorcycle had decelerated more rapidly on the grass than on the asphalt. Also, the digging in of the front wheel rim and tire at the beginning of the slide probably created a temporary additional drag. Drag from the front tire of the motorcycle eventually resulted in the motorcycle's rotation as the sliding velocity decreased.

TEST NO. 2 - Based on the 80 km/hr test speed and the 63 m slide distance, the average friction factor for the total slide was 0.4, a factor slightly lower than in test no. 1. Two reasons were immediately apparent for the difference:

1. The digging in of the front tire at touchdown was much less than during test no. 1.

2.The interaction with the grass at the edge of the road was less, with less material being torn up and the motorcycle never being fully onto the grass.

Other possible factors such as the grinding down of projecting surfaces are less likely, in our opinion, to have had a significant effect.

TEST NO. 3 - The motorcycle slid a distance of 69.5 m on asphalt pavement during test no. 3. Based on the 80 km/hr release speed, the friction factor for this test was 0.36.

TEST NO. 4 - The motorcycle slid a distance of 86.3 m on asphalt pavement during test no. 4. Based on the 80 km/hr release speed, the friction factor for this test was 0.29. We suspect that the lower friction factor was at least partially due to the absence of a left side muffler, which increased the portion of the weight of the motorcycle which was resting on plastic fairing components during this test.

Test Number	Distance (m)		Drag Factor
	Asphalt	Grass	Average
1 (right side)	41	15	0.45
2 (right side)	43	20	0.4
3 (right side)	69.5	-	0.36
4 (left side)	86.3	-	0.29

The tests were designed to obtain values for the friction factor on asphalt. However, due to the 4 or 5 degree angle at which the motorcycle departed from the test platform, the last 15 m of sliding in test no. 1 and the last 20 m of sliding in test no. 2 took place on the grass. The 4 or 5 degree angle stayed constant throughout the four tests, suggesting that it was due to the lateral velocity attained during the fall to the roadway. The angle could affect the test speed by about 0.3 km/hr. The asphalt portions of all tests appeared to be virtually identical in terms of motorcycle dynamics, except for the unusually severe front wheel interaction at the start of test no. 1. Therefore, tests nos. 1 and 2 should have produced results similar to those of tests nos. 3 and 4 if not for the fact that the motorcycle left the asphalt surface.

Based on the similar distances the motorcycle slid on the asphalt in tests nos. 1 and 2, the motorcycle should have been traveling at about the same speed in both tests as it entered the grass. Because the grass portion of test no. 1 involved much more interaction with the turf than was the case in test no. 2, the motorcycle slid a significantly shorter distance on the grass in test no. 1.

To test the hypothesis that the asphalt friction factor was similar for all four tests, we calculated the resultant grass friction factors. Our calculations, assuming an asphalt friction factor equal to the average of the results from tests nos. 3 and 4, suggest that the motorcycle was traveling at about 53 to 55 km/hr when it entered the grass in test nos. 1 and 2. Thereafter, the friction factor on the grass was approximately 0.78 in test no. 1 and 0.55 in test no. 2. The magnitude of these values appeared to be consistent with the observations made of the motorcycle during these two tests.

In analyzing the data from tests nos. 3 and 4, an error factor of plus or minus 1.5 km/hr for the initial speed was included. This error factor accounts for the possible error within the radar unit and for the angle at which the motorcycle slid with respect to the path of the pickup truck. Therefore, the range for the asphalt friction factor for this motorcycle in these tests was between 0.28 and 0.38.

We note that the lower end of the range was obtained from test no. 4, when the motorcycle was sliding on its left side. During this test, more of the weight of the motorcycle was being carried by the fairing, since there was no muffler canister on the left side of the motorcycle. The reduction in friction factor expected from a fairing equipped motorcycle may depend on how much of the weight of the motorcycle is borne by the fairing components as opposed to the metal protrusions such as foot pegs, handlebars, and exhaust components.

The lower end of the range from these tests was consistent with the results reported by Raftery (6). Raftery's method used cables to suspend the motorcycle horizontally an unspecified distance above the roadway before it was released. Raftery reported a result of 0.26 after testing an unspecified make and model motorcycle retrofitted with a fairing at a speed of 85 km/hr.

A similar test method was utilized by Carter et al (7) to test a number of motorcycles, including two sports models with fairings. The motorcycles were suspended with their lowest point approximately 50mm above ground. For the 1987 Yamaha FZ700T, a sports model, tested once at 97 km/hr, the reported result was 0.42. When the same motorcycle was tested at 48 km/hr, the results ranged from 0.40 to 0.48.

Carter et al also tested an earlier sports motorcycle, the 1984 Yamaha FJ1100, and reported results ranging from 0.38 to 0.48 at 97 km/hr and from 0.42 to 0.52 at 48 km/hr. The fairing on the FJ1100 did not extend to fully cover the sides of the motorcycle as it does on most current sport model motorcycles. It may be that the fairing on this motorcycle does not bear the majority of the weight of the motorcycle during sliding tests.

FALL TIME - Analysis of the video tapes showed that the time for the motorcycle to fall from an upright position to roadway contact was about 2/3 of a second. This fall time was relatively constant throughout the testing.

SUMMARY

Based on our test results, the friction factor for fully-faired motorcycles on asphalt ranges from 0.28 to 0.38 when considering a typical accident scenario with a fall; some distinct, short gouges at touchdown; and a final rest position on, or near, the pavement. Where gouge marks are deep or longer than a few inches, or where significant deformation of motorcycle components has occurred due to digging in, higher friction values may be warranted. Where the plastic fairing components bear all or nearly all of the weight of the motorcycle during the slide, the lower values would probably apply.

The values we obtained fall within the lower portion of the range of values previously published. As always, we recommend that accident investigators apply this and any other test data with caution. Motor vehicle accidents are complex events, and the test data reported here does not necessarily apply in every case.

ACKNOWLEDGMENTS

The authors wish to thank the members of the Traffic Unit of the Windsor Police Service for their valuable assistance in the staging of these slide tests.

REFERENCES

1. Day, Terry D., and Smith, Jay R., "Friction Factors for Motorcycles Sliding on Various Surfaces," 840250, Society of Automotive Engineers, 1984

2. Fricke, Lynn B., Traffic Accident Reconstruction, Northwestern University Traffic Institute,1990

3. Limpert, Rudolf, Motor Vehicle Accident Reconstruction and Cause Analysis, Third edition, The Michie Company,1989

4. Baxter, Albert T., Motorcycle Accident Investigation, Institute of Police Technology and Management, University of North Florida

5. Golder, U., and Becke, M., "Skid Deceleration Values for Motorcycles with Full Fairings," Verkehrsunfall und Fahrzeugtechnik,Sept 1990

6. Raftery, Barry, "Determination of the Friction factor of a Fairing Equipped Motorcycle," 950197, Society of Automotive Engineers, 1995

7. Carter, Thomas J.,et al, "Measurement of Motorcycle Slide Coefficients," 961017, Society of Automotive Engineers, 1996

8. Lambourn, Richard F., "The Calculation of Motorcycle Speeds from Sliding Distances," 910125, Society of Automotive Engineers, 1991

9. Donohue, M.D., "Motorcycle Skidding and Sideways Sliding Tests," Accident Reconstruction Journal, Vol. 3,No. 4,1991

10. Daily, John, "Motorcycle Speed Estimates," Accident Reconstruction Journal, Vol. 2, No. 2, 1990

11. Ouellet, James V., "Motorcycles," Personal Injury Scientific Automobile Accident Reconstruction, Vol. 4, 1986

12. Lynch, George F., "Conducting Test Slides - Motorcycles on Asphalt Surfaces (Omaha Police Department)," Law and Order Magazine, November 1984

13. Searle, John A., "The Trajectories of Pedestrians, Motorcycles, Motorcyclists, etc., Following a Road Accident," 831622, Society of Automotive Engineers, 1983

Appendix A: Review of Published Motorcycle Drag Factors

Ref.	Motorcycle Year,Make,Model	Type	Surface Type	Test Method	Test Speed	Friction Factor Low	High	Avg.
(7)	1981 Kawasaki KZ400	Standard	asphalt	drop	48 km/h	0.46	0.53	0.49
	1981 Kawasaki KZ400	Standard	asphalt	drop	97 km/h	0.38	0.41	0.40
	1982 Yamaha XJ650 Maxim	Standard	asphalt	drop	48 km/h	0.62	0.72	0.67
	1982 Yamaha XJ650	Standard	asphalt	drop	97 km/h			0.59
	1982 Yamaha XJ650	Standard	dirt	drop	48 km/h	0.82	1.36	1.02
	1981 Kawasaki KZ440A	Cruiser	asphalt	drop	48 km/h	0.49	0.68	0.61
	1986 Yamaha FZ700T	Sport	asphalt	drop	48 km/h	0.4	0.48	0.43
	1986 Yamaha FZ700T	Sport	asphalt	drop	97 km/h			0.42
	1988 YamahaYX600U Radian	Standard	asphalt	drop	48 km/h	0.54	0.56	0.55
	1988 YamahaYX600U	Standard	asphalt	drop	97 km/h	0.48	0.6	0.54
	1984 Yamaha FJ1100	Sport	asphalt	drop	48 km/h	0.42	0.52	0.47
	1984 Yamaha FJ1100	Sport	asphalt	drop	97 km/h	0.38	0.48	0.42
	1984 Yamaha XV700LC Virago	Cruiser	asphalt	drop	48 km/h			0.62
	1984 Yamaha XV700LC	Cruiser	asphalt	drop	97 km/h			0.51
	1984 Yamaha XV700LC	Cruiser	gravel	drop	48 km/h			0.83
	1985 Yamaha XVZ12DN Venture Royale	Touring	asphalt	drop	48 km/h			0.38
	1985 Yamaha XVZ12DN	Touring	asphalt	drop	97 km/h			0.38
	1985 Yamaha XVZ12DN	Touring	gravel	drop	48 km/h			0.72
(6)		Sport	asphalt	drop	80 km/h			0.26
		Standard	asphalt	drop	80 km/h			0.33
(8)	Honda CB 750G	Standard	fine textured asphalt	sliding from trailer	34 to 80 km/h	0.25	0.35	0.28
	Honda CB 750G	Standard	fine textured asphalt	dragged	10 km/h	0.25	0.33	0.31
	Honda CX500	Standard	coarse asphaltic concrete	sliding from trailer	46 to 80 km/h	0.33	0.54	0.42
	Honda CX500	Standard	coarse asphaltic concrete	dragged	10 km/h	0.44	0.57	0.51
	Honda C-90	Standard	coarse asphaltic concrete	sliding from trailer	37 to 80 km/h	0.36	0.65	0.50
	Honda C-90	Standard	coarse asphaltic concrete	dragged	10 km/h	0.43	0.66	0.53
	Honda C-90	Standard	coarse asphaltic concrete	towed upright and released	40 to 92 km/h	0.49	0.62	0.54
	Honda CG125 (with rt. side leg shields)	Standard	fine textured asphalt	towed upright released-rt side	41 to 92 km/h	0.32	0.42	0.36
	Honda CG125 (no left side leg shields)	Standard	fine textured asphalt	towed upright released-lt side	41 to 91 km/h	0.33	0.42	0.37
	Honda CG125	Standard	fine textured asphalt	dragged	10 km/h	0.32	0.37	0.35
	Suzuki 125ER (side stand-left side)	Dirt	fine textured asphalt	towed upright released-lt side	40 to 90 km/h	0.45	0.53	0.49
	Suzuki 125ER (side stand-left side)	Dirt	fine textured asphalt	dragged on left side	10 km/h	0.42	0.51	0.45
	Suzuki 125ER (no side-stand)	Dirt	fine textured asphalt	towed upright released-rt side	41 to 92 km/h	0.34	0.4	0.37

Ref.	Motorcycle Year,Make,Model	Type	Surface Type	Test Method	Test Speed	Friction Factor		
						Low	High	Avg.
(8)	Suzuki 125ER (no side-stand)	Dirt	fine textured asphalt	dragged on right side	10 km/h	0.34	0.37	0.35
	Yamaha TZR125 (right side fairing)	Sport	coarse asphaltic concrete	towed upright released-rt side	39 to 95 km/h	0.39	0.54	0.43
	Yamaha TZR125 (right side fairing)	Sport	coarse asphaltic concrete	dragged on right side	10 km/h	0.33	0.47	0.39
	Yamaha TZR125 (left side- no fairing)	Sport	coarse asphaltic concrete	towed upright released-lt side	40 to 78 km/h	0.47	0.79	0.62
	Yamaha TZR125	Sport	coarse asphaltic concrete	dragged on left side	10 km/h	0.48	0.61	0.54
(9)	1982 Kawasaki KZ1000 Police Special		sharp asphaltic concrete	towed upright and released	51 to 69 km/h	0.38	0.5	0.44
(2)	unspecified		asphalt or portland cement	unspecified	unspecified	0.4	0.75	
	unspecified		gravel	unspecified	unspecified	0.65	1.05	
(10)	unspecified		Std street (light pavement scratching)	unspecified	unspecified	0.3	0.35	
	Crash bars			unspecified	unspecified			0.2
	unspecified		Std street (1/4 inch deep scratching)	unspecified	unspecified	0.5	0.6	
(3)	unspecified		unspecified	unspecified	48 km/h	0.5	0.55	
	unspecified		unspecified	unspecified	80 km/h	0.4	0.5	
	unspecified		unspecified	unspecified	105 km/h	0.35	0.4	
(11)	Standard	Standard	ordinary pavement	unspecified	unspecified	0.35	0.45	
	front & rear crash bars		ordinary pavement	unspecified	unspecified			0.2
	objects sticking out of side (pegs)		ordinary pavement	unspecified	unspecified			0.5
			oil on roadway	unspecified	unspecified			0.2
			soft dirt, tumbling	unspecified	unspecified			0.7
(1)	1967 Honda CB305	Standard	moderately worn asphalt	drag	40 km/h	0.45	0.54	0.51
	1967 Honda CB305	Standard	gravel	drag	40 km/h	0.68	0.79	0.74
	1967 Honda CB305	Standard	hard-packed,dry grassy earth	drag	32 km/h			0.7
	1973 Yamaha 500 Special	Standard	moderately worn asphalt	drag	40 km/h	0.49	0.58	0.54
	1973 Yamaha 500	Standard	gravel	drag	40 km/h			0.74
(12)	unspecified		asphalt	unspecified	unspecified			0.57
	Dual Purpose		asphalt	unspecified	unspecified			0.38
(13)	unspecified		dry asphalt	unspecified	unspecified	0.35	0.5	
	unspecified		wet asphalt	unspecified	unspecified	0.3	0.4	
(4)	unspecified		asphalt	unspecified	unspecified	0.3	0.4	
	unspecified		concrete	unspecified	unspecified	0.43	0.53	
	round & smooth crashbars		asphalt	unspecified	unspecified			0.2
	unspecified		asphalt, oil soaked	unspecified	unspecified			0.2

SAE #970963
Reviewer's Discussion
by Terry D. Day, Engineering Dynamics Corporation
Motorcycle "Slide to a Stop" Test
Christopher J. Medwell, Joseph R. McCarthy, Michael T. Shanahan, Authors

This paper extends the current knowledge of drag factors for motorcycles sliding after an upset (either caused by impact or simple loss of control). The tests are well-documented, and the additional data for the new generation of fairing-equipped sport cycles is valuable. In addition to the quantified results, an important contribution of this paper is the qualitative description of the cycle's motion during each test, allowing one to conclude there is a variability in the resulting drag that is dependent on how the cycle interacts with the surface. Viewing the test videos might be as useful (perhaps educational is a better word) as the actual drag factors.

SAE #970963
Reviewer's Discussion
by John Hunter, Washington State Patrol
Motorcycle "Slide to a Stop" Test
Christopher J. Medwell, Joseph R. McCarthy, Michael T. Shanahan, Authors

The title "Motorcycle Slide to Stop Tests" is very appropriate. The authors were unable to locate adequate references and decided to test a motorcycle. This paper discusses a series of four tests completed by the authors and is very specific on how the tests were conducted. The material is presented in a straight forward fashion allowing the reader to understand the method of testing and how the findings can be related to collisions involving a specific type of motorcycle.

The information is limited to one series of tests. The limitation does not reflect on the value of the material presented. The authors willingness in sharing their test information becomes a valuable resource and adds to work completed by other investigators.

An added benefit of the paper is the compilation of information provided by the Appendix. The authors took the time to research other papers and articles concerning deceleration factors for sliding motorcycles. This appendix has been organized into a very useful resource for the reconstructionist or investigator.

970969

On-Demand Four Wheel-Drive Transfer Case Modeling

J. Asgari and D. Hrovat
Ford Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper describes the development of a Matrixx Model of an electronically controlled on-demand 4 Wheel-Drive (4WD) Transfer Case. The model was partially validated with respect to available vehicle test data and subsequently used for control system design and evaluation.

1. INTRODUCTION

As an aid in conceptual development of sophisticated control strategies in ground vehicles, efficient and validated physically based system models are necessary. The model should capture the basic features of the system, and represent the most essential system behavior. It should be readable and should preserve the physical meaning of system variables and parameters. The models must be sufficiently complete to include the important dynamics, but not so complex that insight is obscured.

The purpose of this project was to develop a Matrixx-based computer model of on-demand 4 Wheel-Drive (4WD). This model is complete enough to predict a correct trend of behavior in vehicle simulation and simple enough to study the influence of the main design factors on the dynamic behavior of the vehicle.

The transfer case is an independent unit in the 4WD system, mounted at almost the mid-section of the vehicle. Its primary function is to split power automatically between the rear and front axle, when predetermined slip of the rear wheels is exceeded.

Bond graph techniques [1] were used for model development. Bond graphs are a pictorial representation of different energy domain systems, and allow a priori determination of the computability of a proposed model through assignment of causality. It further permits the modeling of vehicle subsystems which are later assembled into overall system models. These subsystem models can even be tested independently and as long as their causality remains invariant, the subsystem model can straight-forwardly be integrated into system models. This paper describes the application of such an approach, described in general terms in [1], where additional bond graph applications are given.

2. SYSTEM CONCEPT

Figure 1 shows a simplified concept cross section of the transfer case. When the on-demand 4WD is engaged, torque from the transmission goes to the rear prop-shaft, and also through the clutch assembly to the front prop-shaft. A more complete drawing of the transfer case is illustrated in Fig. 2.

The transfer case transfers torque using four main components: control module, solenoid, ramp and roller assembly, and clutch assembly.

The control module measures the rotational speed of front and rear prop-shafts, and when this difference is large, it increases the duty cycle of PWM input signal to the solenoid. The solenoid average current follows the PWM command and generates an electromagnetic force which attracts the solenoid plate and collar (see Fig. 1). A change in PWM will translate in a change in the pressure and (primary) torque between the solenoid plate and collar. The solenoid plate is connected to the front prop-shaft by a spline coupling, and has the same angular speed. It is free to move axially.

The apply plate is connected to the rear prop-shaft by a spline coupling, which can move axially while rotating at the speed of rear prop-shaft. The collar is connected to the apply plate by a ramp and roller assembly (Fig. 3). When the 4WD is disabled, the collar rotates with the speed of the rear prop-shaft (Fig. 3a) and no torque is transferred to the front prop-shaft.

When the 4WD is actuated, the collar speed reduces due to the (primary) friction torque between the collar and solenoid plate (see Fig. 1). As the result of the difference between the speeds of the collar and the apply plate, the rollers roll up the ramp (Fig. 3) causing axial motion of the apply plate. This axial movement of the apply plate causes the engagement of the front and rear clutch plates which transfer the torque from the rear prop-shaft to the front prop-shaft (secondary torque).

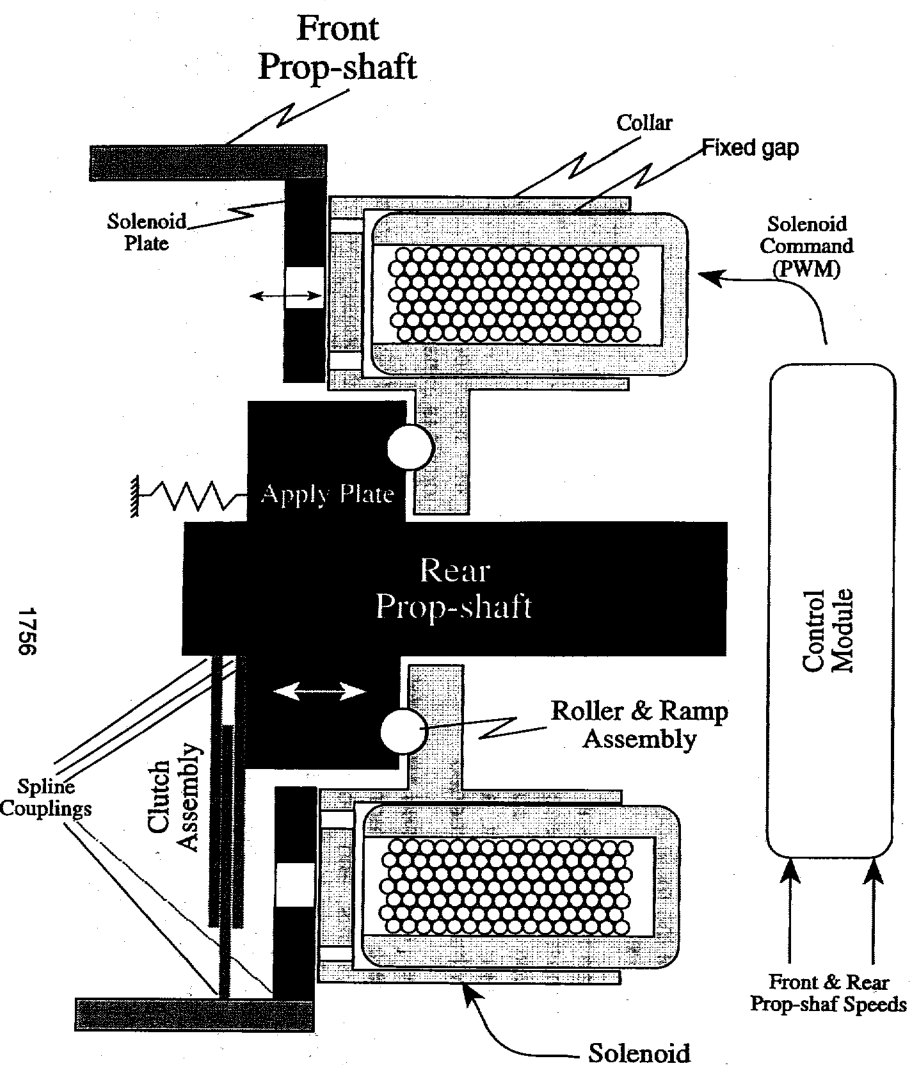


Fig. 1. Simplified Concept Cross Section of Transfer Case

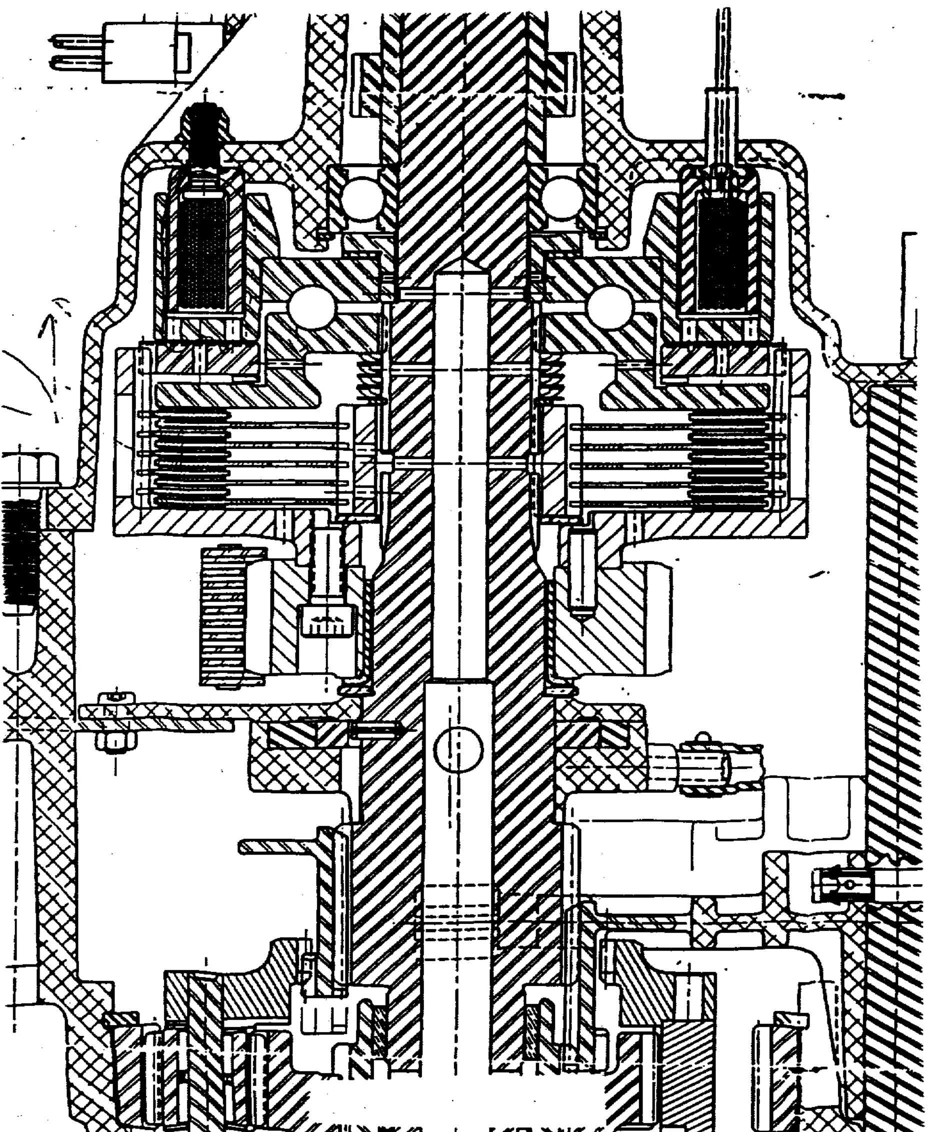


Fig. 2. Transfer Case Drawing

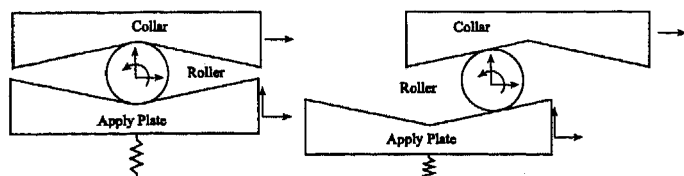


Figure 3. Ramp and Roller Assembly.

3. MODEL DEVELOPMENT

As indicated earlier, the bond graph modeling technique has been used for the model development. Figure 4 shows the word bond graph representation of the system. The word bond graph represents the major subsystems to be considered. The main submodels consist of:

1) transfer case
2) drive train
3) differential and axles
4) tires
5) vehicle model

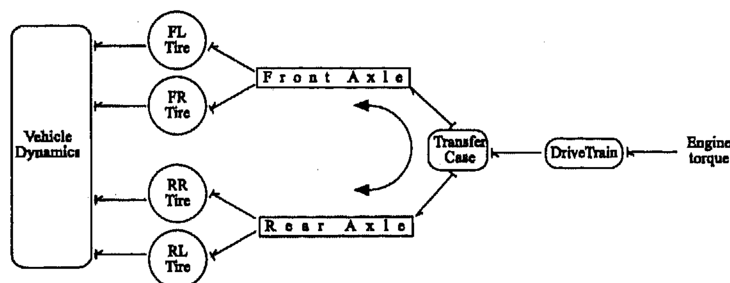


Figure 4. Word Bond Graph Representation of the Model

These submodels are discussed below.

3.1) TRANSFER CASE

As was previously discussed, the transfer case consists of four main sub components: solenoid, ramp and roller, clutch assembly, and control module. Figure 5 shows the way these parts are connected together. These subcomponents will be discussed in detail below.

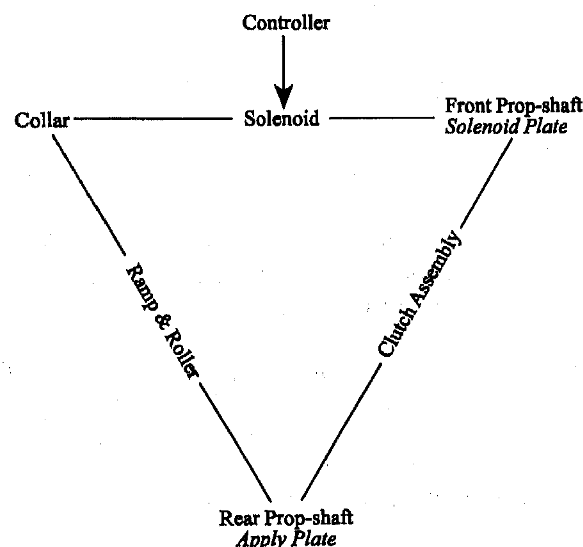


Figure 5. Transfer Case Connectivity

3.1a) Solenoid

The solenoid is capable of transforming an electrical signal into mechanical movement. The response time of the solenoid is of primary importance. The system (Fig. 1) consists of a coil of wire, a fixed piece of soft iron, and a moveable core (solenoid plate). When the current increases in the coil, the solenoid plate gets closer to the stationary core.

The corresponding model is divided into three areas: (I) electrical model (II) the electromagnetic model which describes the interaction of the iron magnetic flux path with the solenoid coil winding; and (III) the magnetic to mechanical interaction model which describes the force due to magnetic flux.

I) Electrical model

The electrical model follows the rise of current in the coil. The effective voltage in this system is reduced by the resistance in the coil as shown by equation 1.

$$e = V_{in} - R_E i \qquad (1)$$

II)Electromagnetic model

The electromagnetic model follows the rise of magnetic flux in the iron path, gaps, moving and stationary parts. By using the Faraday's law applied to the coil, we have

$$e = \dot{\lambda} = N\dot{\phi} \qquad (2)$$

where λ is flux linkage, N is the number of turns of coil, and ϕ is the flux. The driving force which sets up the flux in the core is the magnetomotive force, M.

$$\sum M = M_g + M_c + M_{g_a} = Ni \qquad (3)$$

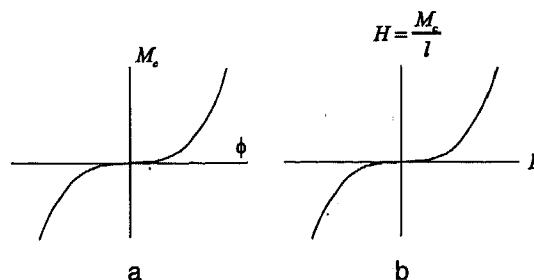


Figure 6. Solenoid Housing Charactristics

In the iron the magnetomotive force, M_c, and flux are related by a nonlinear relation (Fig 6a) which is function of material and geometry. To characterize a core material [1], it is convenient to use a curve independent of core configuration. For this purpose usually the B-H curve is used (Fig. 6b) where

B ≡ Magnetic flux density (weber/ square meter)
H ≡ Field strength (Amper-Turn/meter)

$$M_c = f(\phi) \qquad (4)$$

In the gap, the relation between the magnetomotive force and flux is defined by

$$M_g = \frac{\phi_g}{C_g} \quad (5)$$

$$C_g = \frac{\mu_o A_g}{4x_g} \quad (6)$$

where x is the gap length, A is the gap area and μ_o is permeability of free space $\equiv 4\pi * 10^{-7}\ Tm/A$. (assumed air (free space), space may contain some transmission fluid).

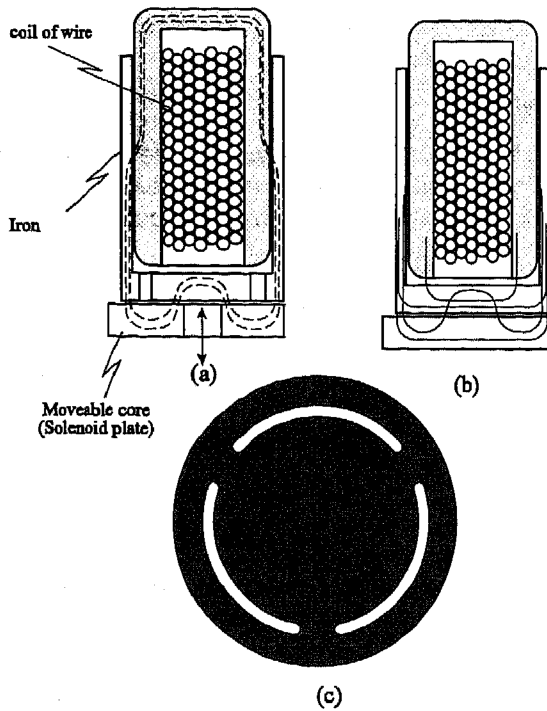


Figure 7. Flux Path.

For a fixed gap,

$$M_{g_a} = \frac{\phi_g}{C_{g_a}} \quad (7)$$

$$C_{g_a} = \frac{\mu_o A_{g_a}}{2L_{g_a}} \quad (8)$$

Ideally it is desired to have flux follow the path shown in Fig. 7a, but since some paths through the solenoid plate (Fig. 7c) and the stationary part are solid, and since there is some leakage in the system, the flux follows a combination of the paths shown in Fig 7b. This results in a significant flux leakage in the system.

III) The magnetic to mechanical interaction model

In [1] it has been shown that the force due to flux is equal to:

$$F = \frac{\phi_g^2}{2\mu_o A}. \quad (9)$$

Figure 8 shows the complete bond graph representation of the solenoid. Using Fig. 8 the overall equations of the solenoid can be written as follows:

$$i = (M_g + M_{g_a} + M_c)/N \quad (10)$$

$$\dot{\phi} = \frac{V_{in} - R_E i}{N} \quad (11)$$

$$F = \frac{\phi_g^2}{2\mu_0 A_g} \quad (12)$$

$$\ddot{x}_g = \frac{-F - Kx_g - Rv_g}{m} \quad (13)$$

Where *K* and *R* are the effective stiffness and damping applied at the solenoid plate.

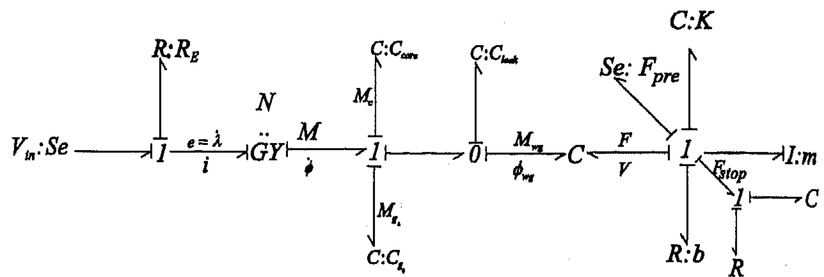


Figure 8. Bond Graph Representation of Solenoid

3.1b) Ramp & roller

Figure 3 shows a free body diagram of the ramp and roller. Figure 9 shows the bond graph representation of the system which includes the dynamics of the roller (3 DOF) and the apply plate. The inertias on bonds 1, 2, and 3 (in the right hand side of Fig. 9) represent the dynamics of the roller in x-direction, y-direction, and rotation of the roller (see Fig. 3). The compliances on the bonds 4 and 5 represent the stiffness of the roller with respect to the apply plate and the collar, respectively. The resistance on the bonds 6 and 7 represent the friction force between the collar and roller. Since there are two ramp surfaces in the apply plate and collar, there is another set of compliance- and resistance-like bonds with α replaced with -α in the appropriate transformers. The inertia on bond 8 represents the dynamics of the apply plate in the axial direction. The compliances on bonds 9 and 10 represent the stiffness of the return spring and clutch assembly respectively. The force in the bond 10 is used to generate the secondary torque. Bond 11 will be connected by a transformer to a 1-junction representing the collar rotational rate, and bond 12 will be connected to a 1 junction representing the rear prop-shaft rotational rate. This model can be substantially simplified by (approximately) replacing the above dynamics with a simple spring and damper.

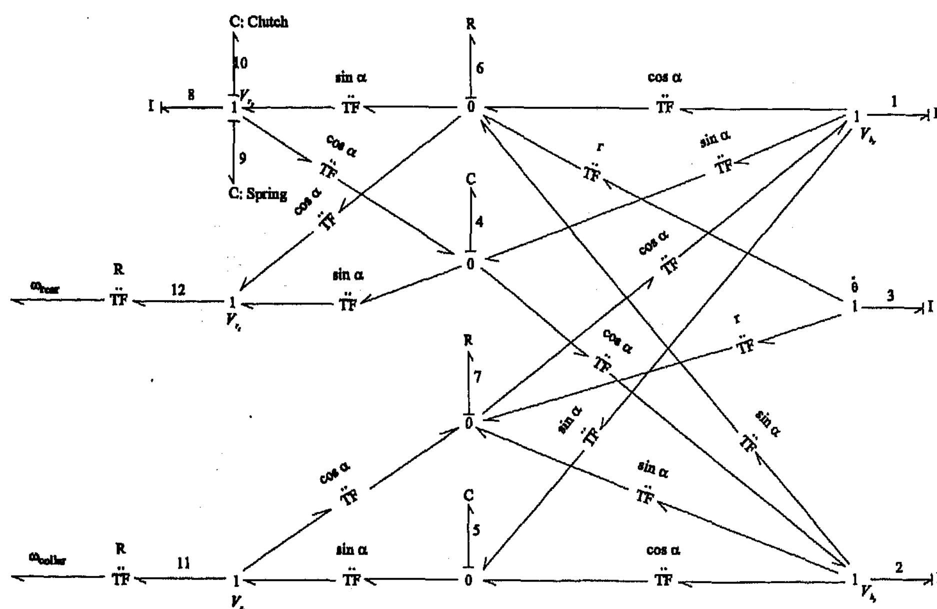


Figure 9. Bond Graph Representation of Ramp and Roller

3.1c) Clutch Assembly

The friction torque in the clutch was calculated by [2]:

$$if\ \Delta\omega < \Delta\omega_{sat}\ then$$

$$\tau = NRF_{n_c}\mu(1 + 0.2e^{-0.2\Delta\omega_{sat}})\frac{\Delta\omega}{\Delta\omega_{sat}}$$

$$else$$

$$\tau = NRF_{n_c}\mu(1 + 0.2e^{-0.2\Delta\omega})$$

where

$N \equiv$ the number of clutch plates,

$R \equiv$ effective radius,

$F_{nc} \equiv$ Normal load (calculated in the ramp and roller),

$\mu \equiv$ friction coefficient,

$\Delta\omega \equiv$ the difference between rotational rate of front and rear prop-shaft,

$\tau \equiv$ the friction torque in the clutch (secondary torque),

$\Delta\omega_{sat} \equiv$ the saturation point.

This equation approximates the stick/slip characteristic of the clutch.

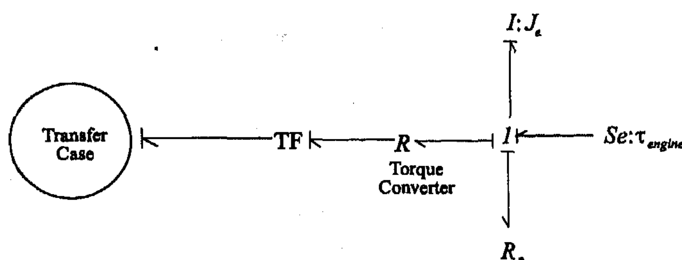


Figure 10. Bond Graph Representation of Drive Train.

3.2) Drive Train

The bond graph representation of the drive train model is shown in Fig. 10. The model includes the engine torque source, engine rotational inertia, engine loss, torque converter, and the transmission gear ratio. The downstream rotational inertia has been combined with the rear prop-shaft inertia.

3.3) Differential and Axle

The function of the differential is to provide driving torque to both left and right tires, while allowing for unequal left and right angular speeds. The axle and differential are modeled [3, 4] by two transformers followed by a three-port 0-junction (Fig. 11). The first transformer represents the final drive speed ratio, and the second transformer accounts for the torque split to the two wheels. The half shaft compliance and damping are also included. The drive shaft compliance could be ignored or alternatively included in half shaft compliance.

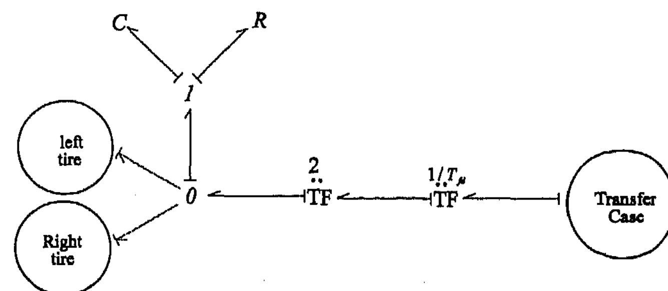


Figure 11. Bond Graph Representation of Axle and Differential

3.4) Tire

Except for aerodynamic forces and gravity, all the forces which affect vehicle handling are determined by tires. Tire forces provide the primary external influence, and because of their highly nonlinear behavior, cause the biggest variation in vehicle handling properties throughout the longitudinal and lateral acceleration maneuvering range. As a result, it is important to use a realistic nonlinear tire model, especially when investigating large control inputs that result in response near the limits of the maneuvering capability of the car. In such situations, the lateral and longitudinal motions of the vehicle are strongly coupled through the tire forces and large values of longitudinal slip and slip angle occur simultaneously.

The many existing tire models are predominantly "semi-empirical" in nature. The model structure is determined through analytical considerations, but the key parameters still depend on tire data measurements. They range from extremely simple (where lateral forces

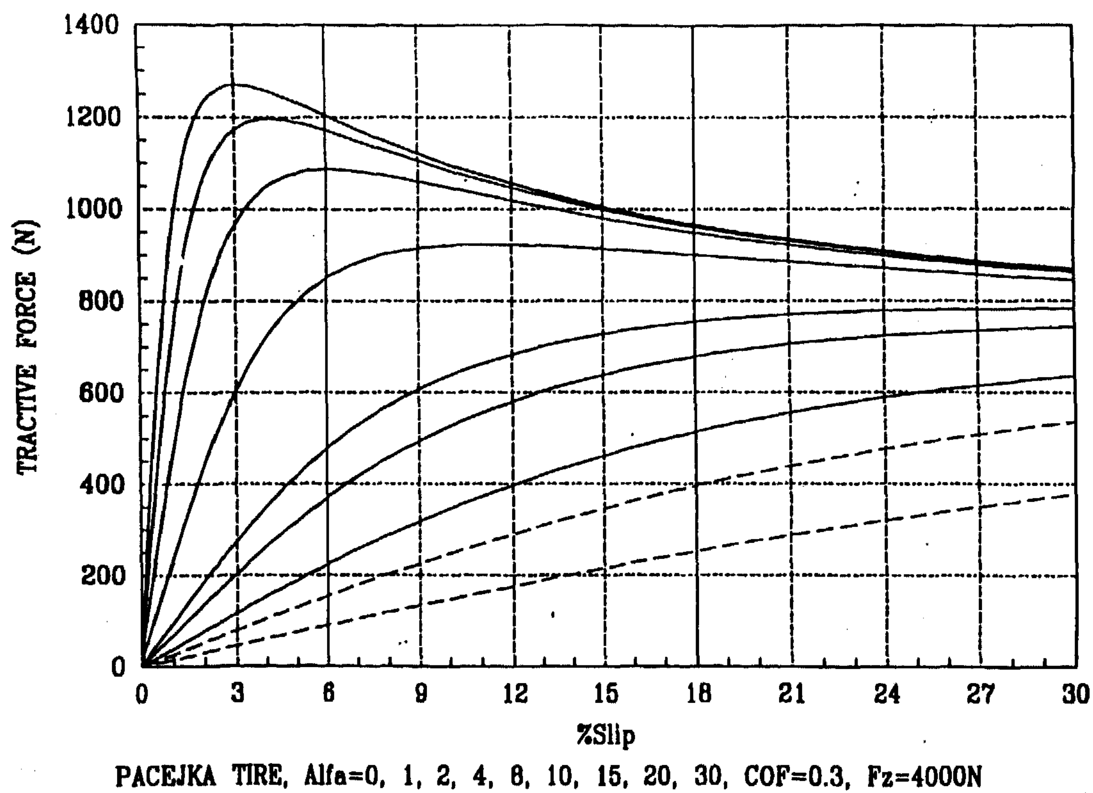


Fig. 12a. Tractive Force as a Function of % Slip during Braking and Cornering.

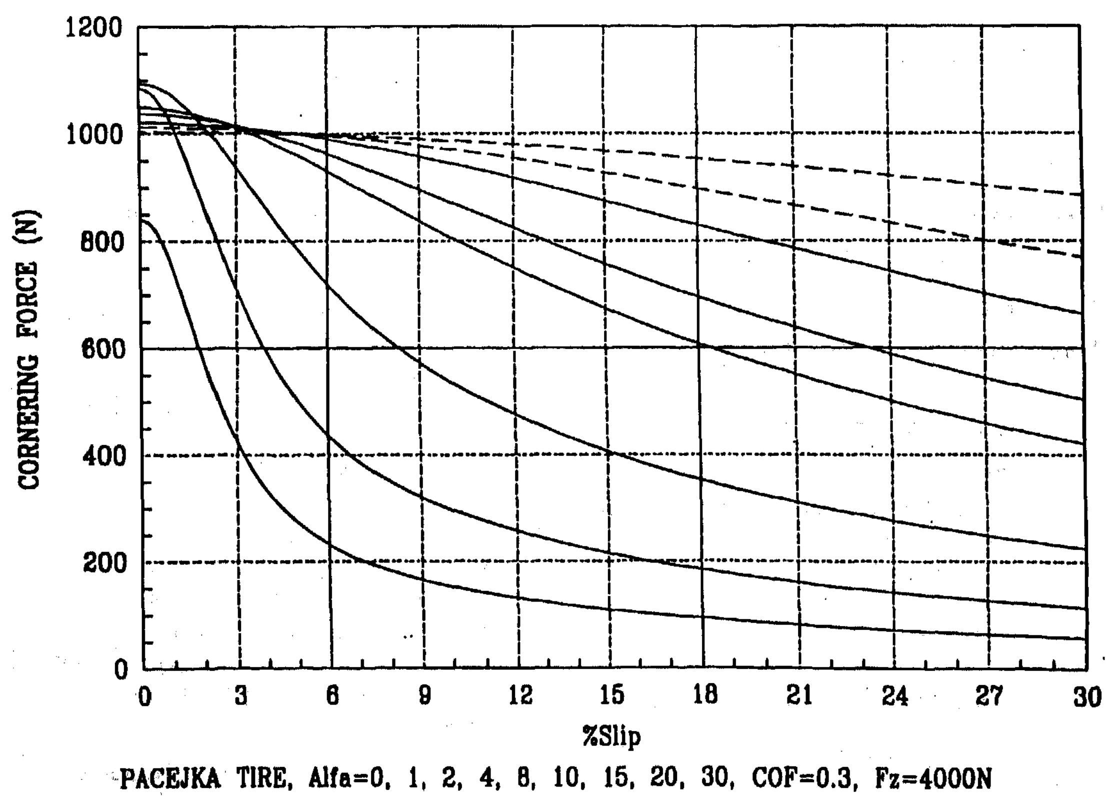


Fig. 12a. Cornering Force as a Function of % Slip during Braking and Cornering.

are computed as a function of slip angle, given only a measured slope at α=0 and a measured value of the maximum lateral force) to relatively complex expressions, which use tire data measured at several loads and slip angles.

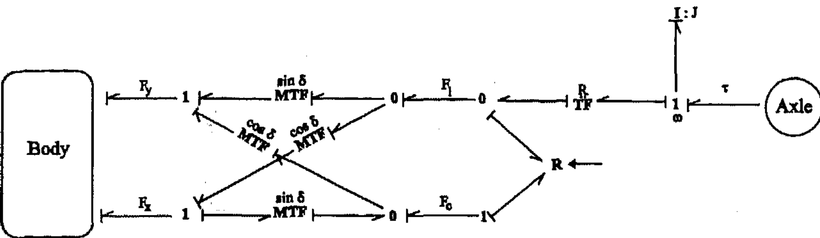


Figure 13. Bond Graph Representation of Tire.

The model for tire tractive and cornering forces used here (Pacejka's tire model [5]) is a complex semi-empirical model of a tire that takes into consideration the interaction between the tractive force and the cornering force in combined braking and steering. The longitudinal and cornering forces are assumed to depend on the normal force, slip angle, and longitudinal slip. Sample plots of modeled longitudinal and lateral forces versus longitudinal slip and slip angle are shown in Figs. 12a and 12b. In this model, the dynamic effect of tires while negotiating sudden changes of operating conditions [6] has been ignored, but could be included in the future if needed.

Figure 13 shows the bond graph representation of a tire. It includes: the tire inertia, steering effect, and nonlinear tire cornering and longitudinal force production mechanism modeled as (nonlinear) R-field.

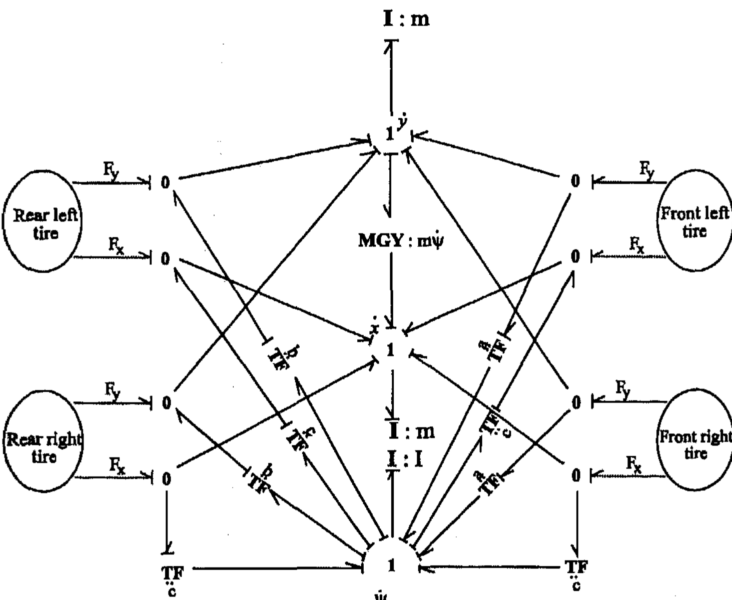


Figure 14. Bond Graph Representation of Body

3.5) Body

This model is intended for simultaneous studies of vehicle fore-aft and handling dynamics [7]. It has 3 degrees of freedom: lateral and longitudinal velocities, $\dot{y}$ and $\dot{x}$, and yaw rate, $\dot{\psi}$. The body moves as a result of forces transmitted from the road through the tires. A bond graph schematic of the model is shown in Fig. 14. Here, m represents the body mass and I represents the body moment of inertia about a vertical axis through the center of gravity.

3.6) Overall Model and Assumptions

Figure 15 shows the overall system model. The main inputs to the model are: engine torque (Fig. 15a), road coefficient of friction (Fig 15b), and steering input (Fig 15b). In this model the transient effect of the tire during sudden changes, the slip/stick characteristic of clutch, engine dynamics, and vehicle heave and pitch dynamics were not included. These submodels are available and can be included in the overall model as necessary. The tire characteristics used are generic and may not necessarily exactly represent the Explorer tire curves.

4. PARTIAL MODEL VALIDATION

To study the accuracy of the model, comparison tests were done between experimental and simulation results for different surface, speed, and load conditions. The experimental data available included: front and rear prop-shaft speeds, front and rear axle torques, and solenoid current.

An example of an experimental test done on packed snow is shown next. The controller strategy requested a medium-to-high level of torque. Figures 16 and 17 show the comparison between the experimental and simulation results. Figure 16a shows the experimental prop-shaft speeds, while Fig. 16b shows the corresponding model results. Figure 17 compares the corresponding solenoid currents. The high frequency oscillation in Fig. 17 is due to PWM duty cycle. Overall, there is a good agreement between the experimental and simulation results especially taking into account the fact that the road (snow) properties are not exactly known.

5. OPERATION OF A LOCKED SYSTEM

It is well-known that the 4WD vehicles should be driven in 4WD-lock position only at low speeds, and preferably on straight sections of slippery/muddy roads. If however, one attempts such a drive on dry roads and higher speeds, then, one could occasionally feel a chatter-like behavior both for cars and trucks. An attempt was done to qualitatively reproduce this result through simulations. Due to the lack of complete data, an exact comparison was not possible. To recreate the observed phenomena, the rear tire radius was increased by about 7% to simulate different front to rear load distribution and tire conditions. Figures 18 to 20 show the corresponding simulation results.

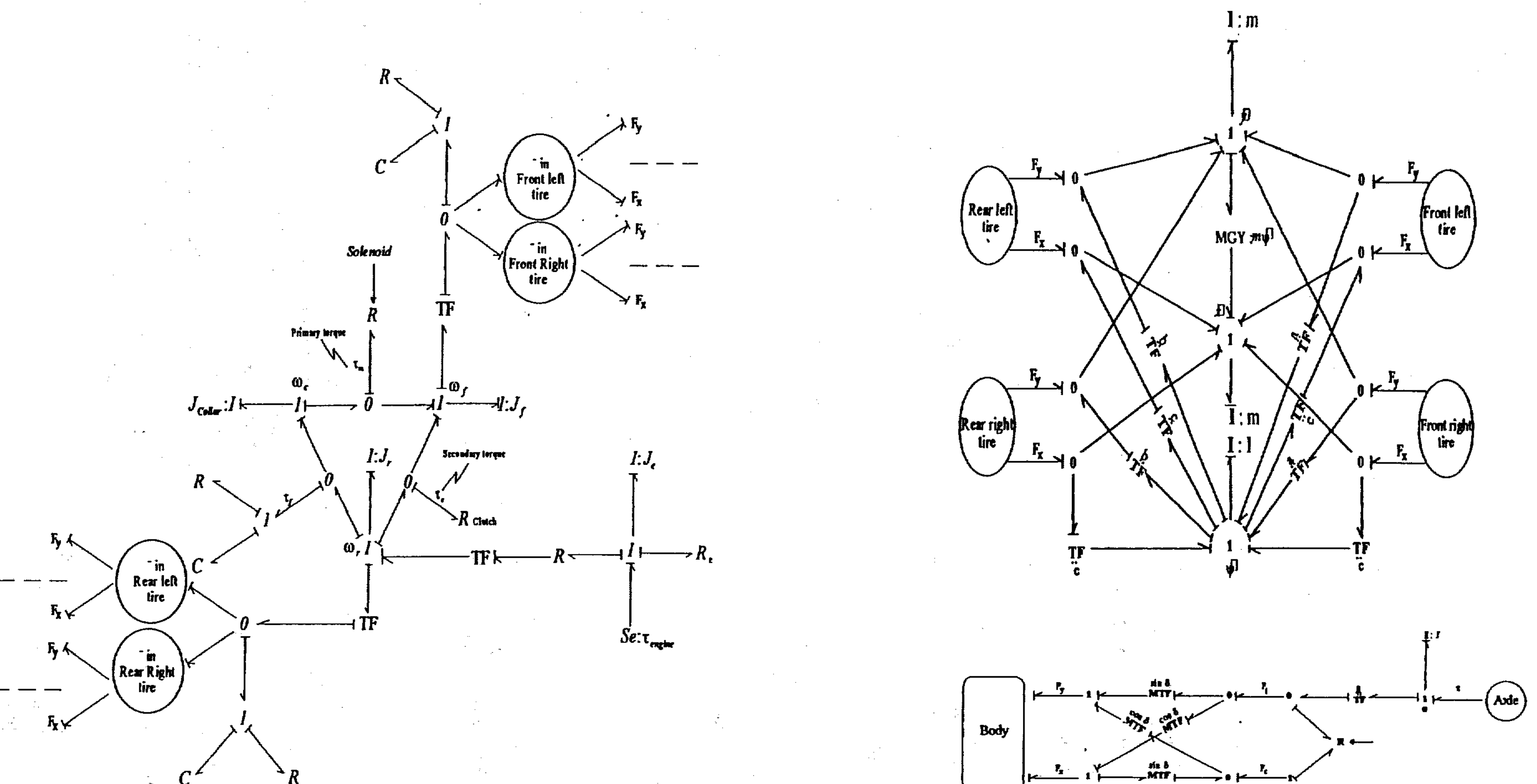


Fig. 15a. Overall Vehicle Model with Simplefied TOD Model

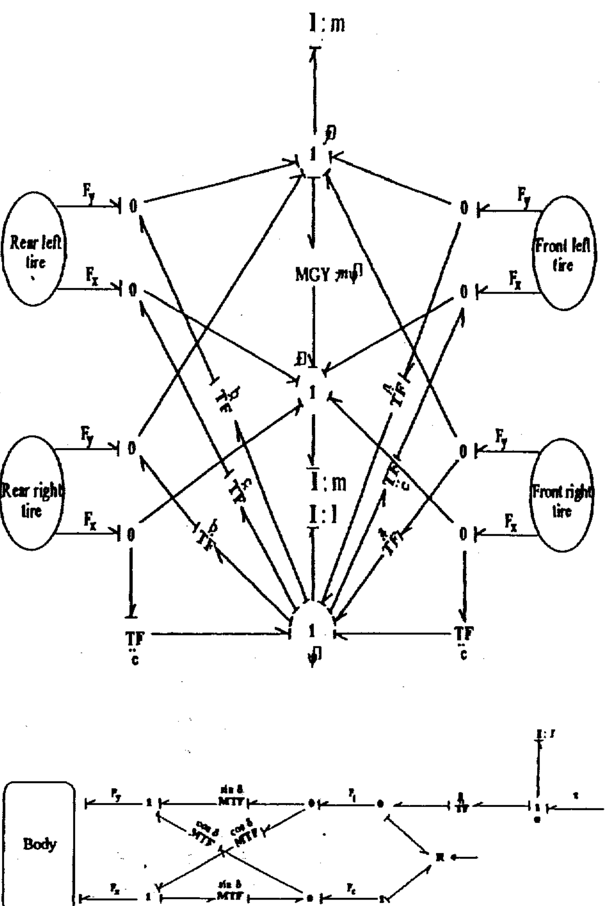


Fig. 15b. Overall Vehicle Model

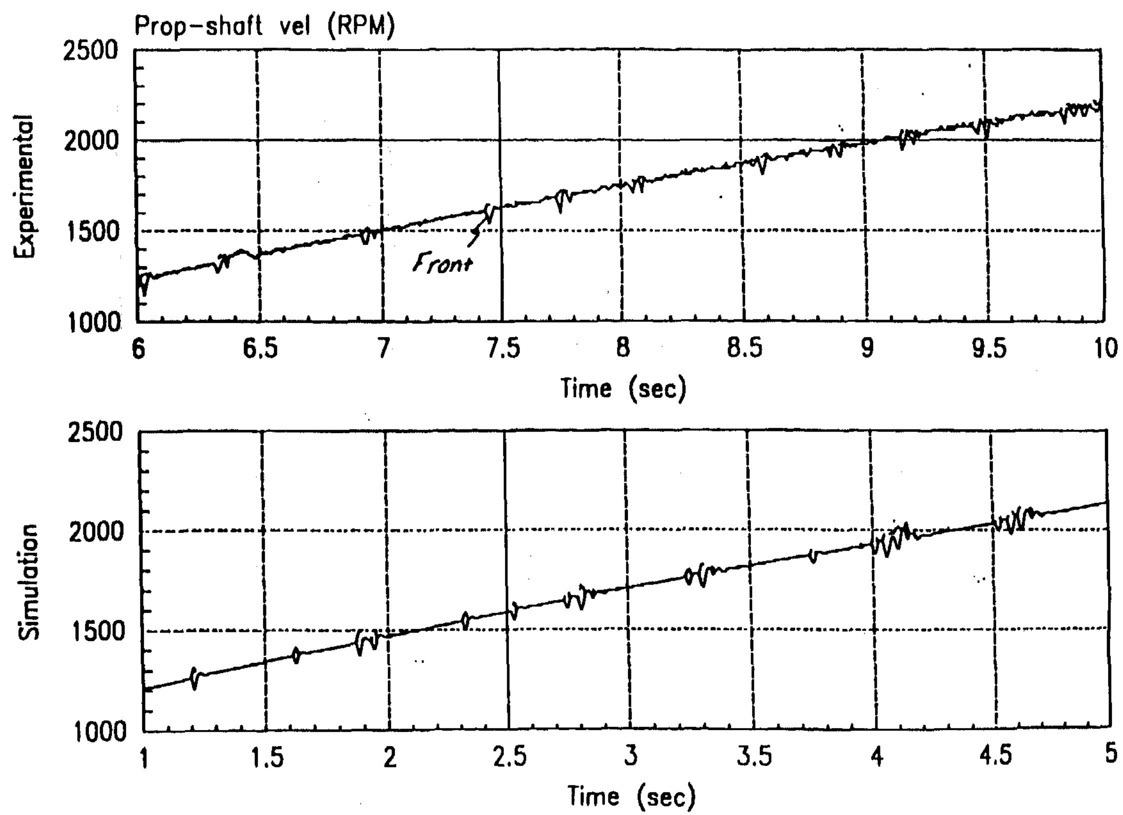


Fig. 16. Vehicle Test and Simulation Prop-Shaft Speeds on Packed Snow at Higher Speeds and Lower Torques.

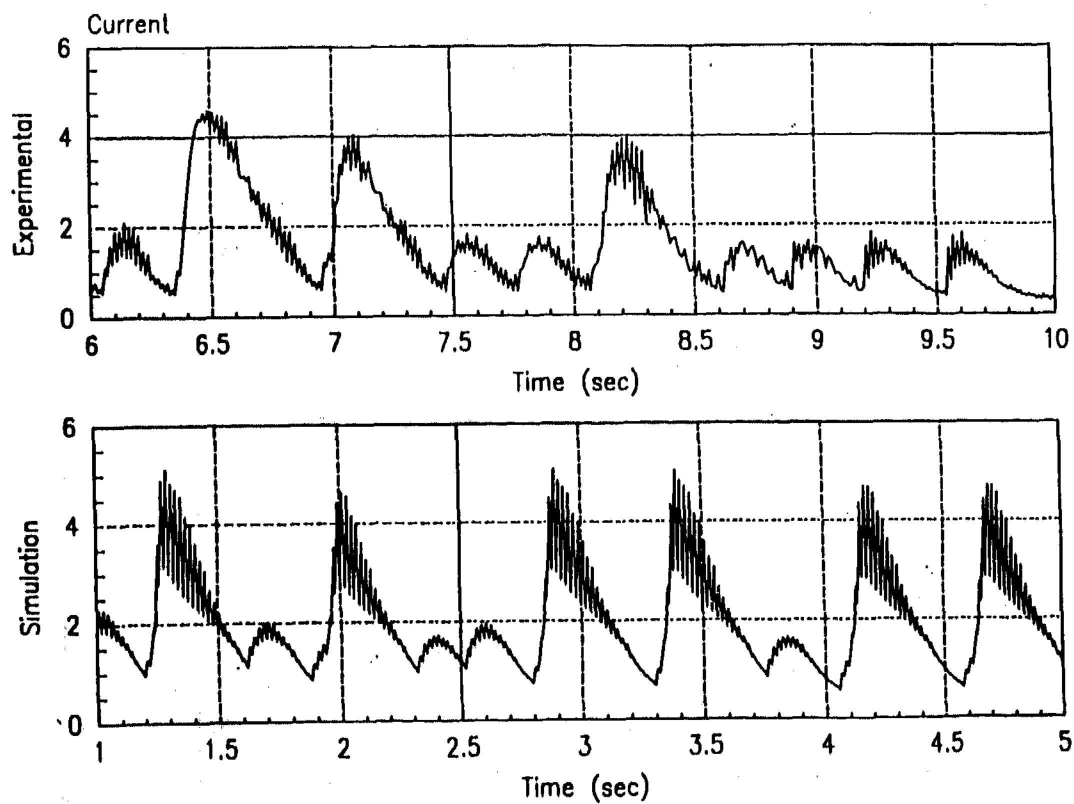


Fig. 17. Vehicle Test and Simulation Solenoid Currents on Packed Snow at Higher Speeds and Lower Torques.

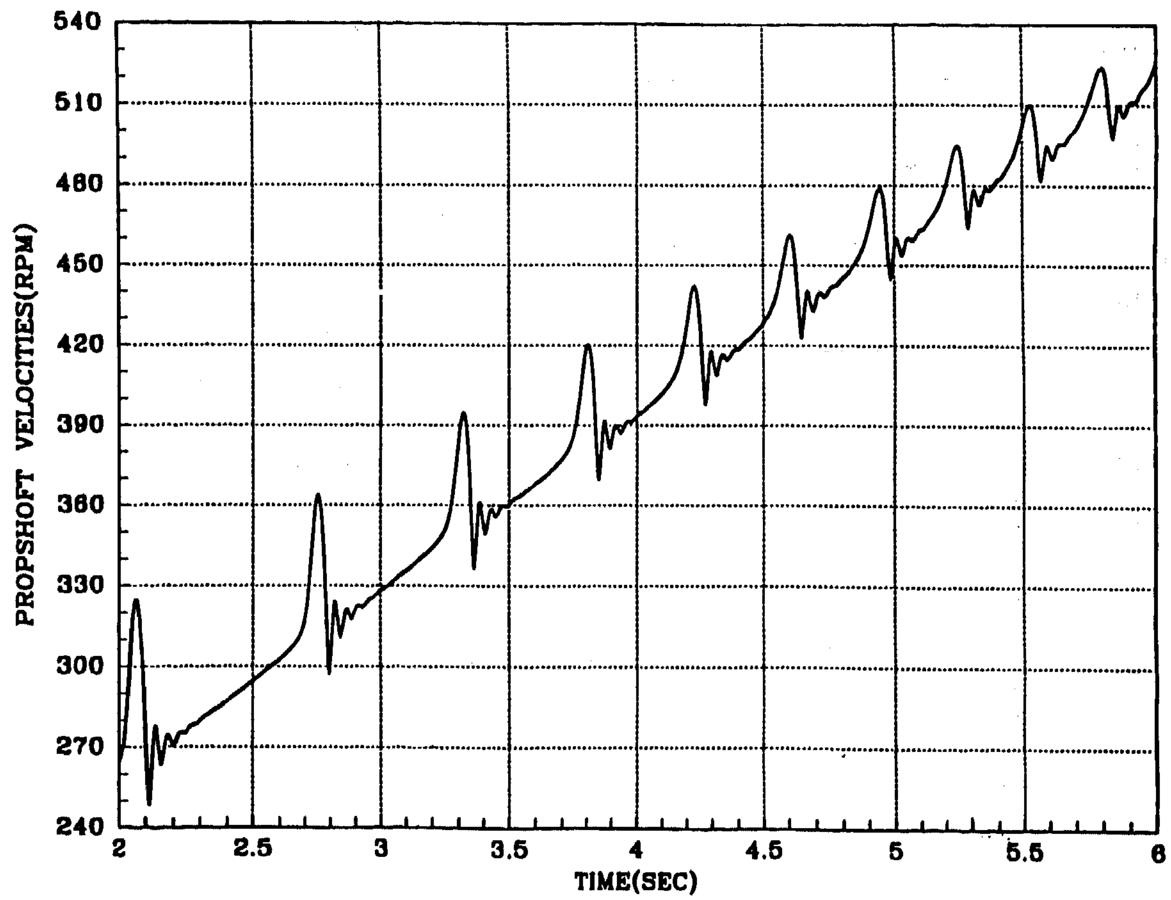


Fig. 18. Simulation Prop-Shaft Speeds for 4WD Lock on Snow.

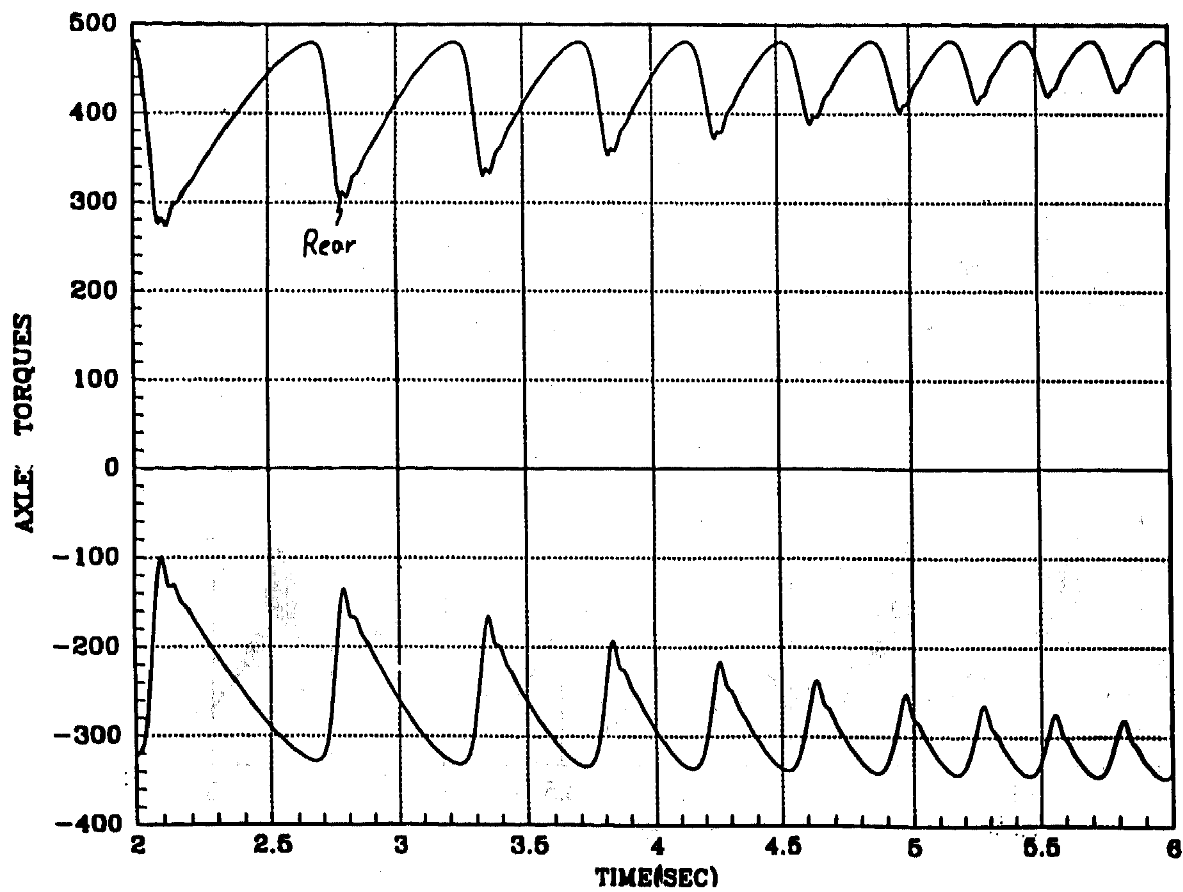


Fig. 19. Simulation Prop-Shaft torques for 4WD Lock on Snow.

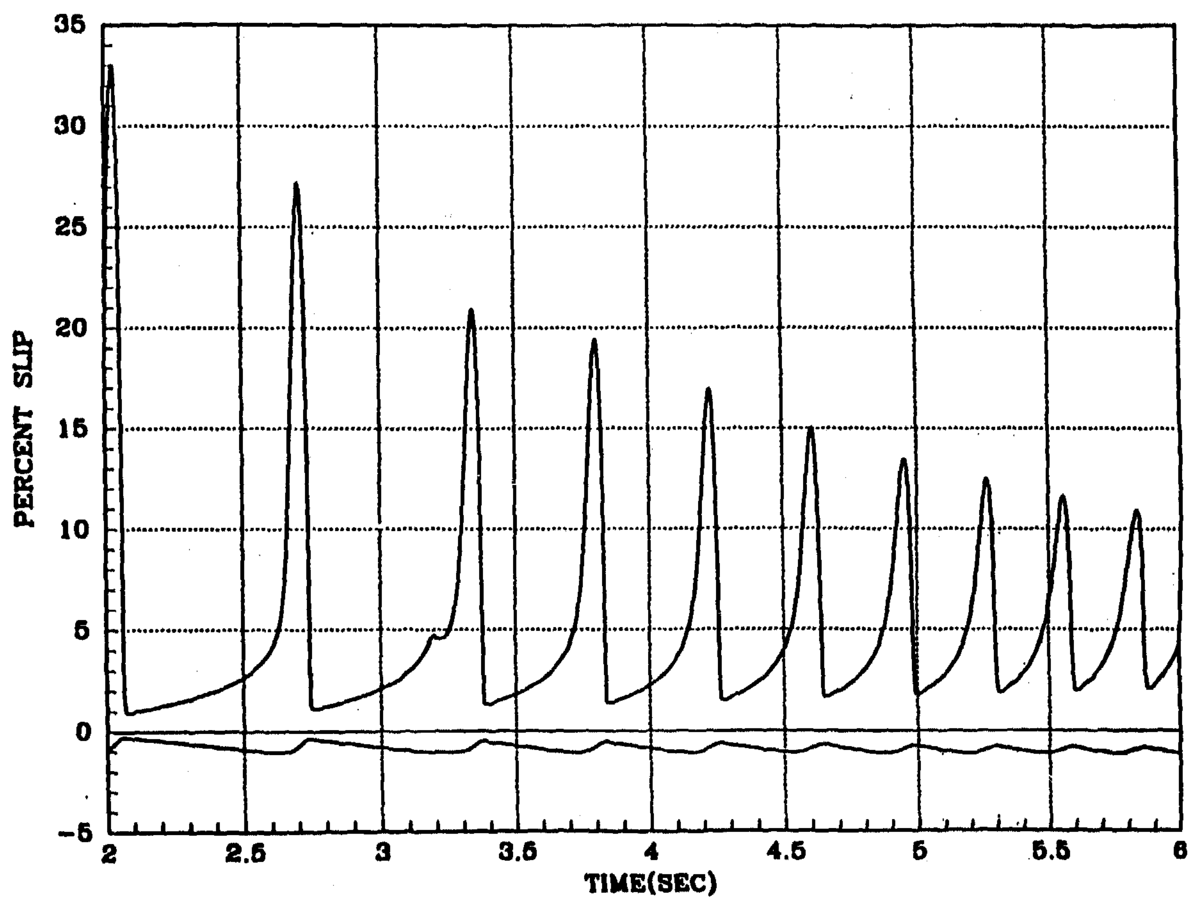


Fig. 20. Simulation % Slip for 4WD Lock on Snow.

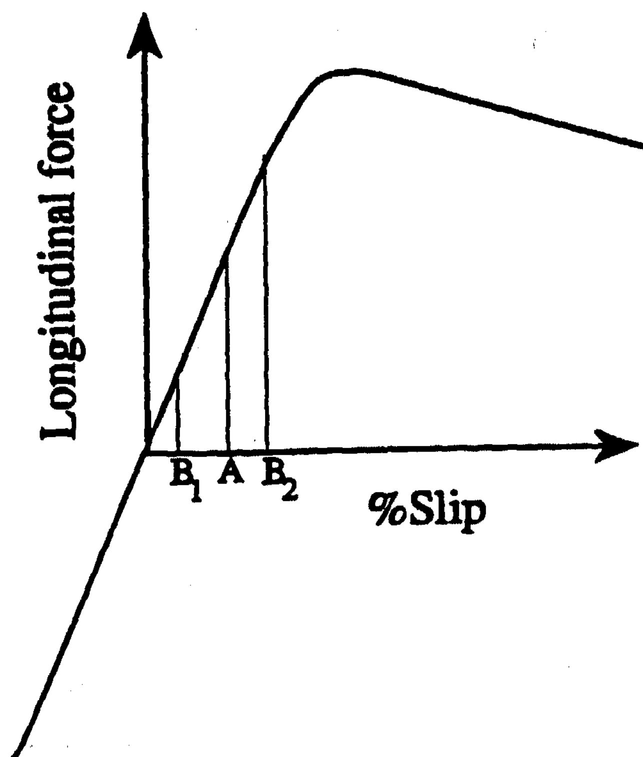


Fig. 21. Typical Longitudinal Force as Function of % Slip.

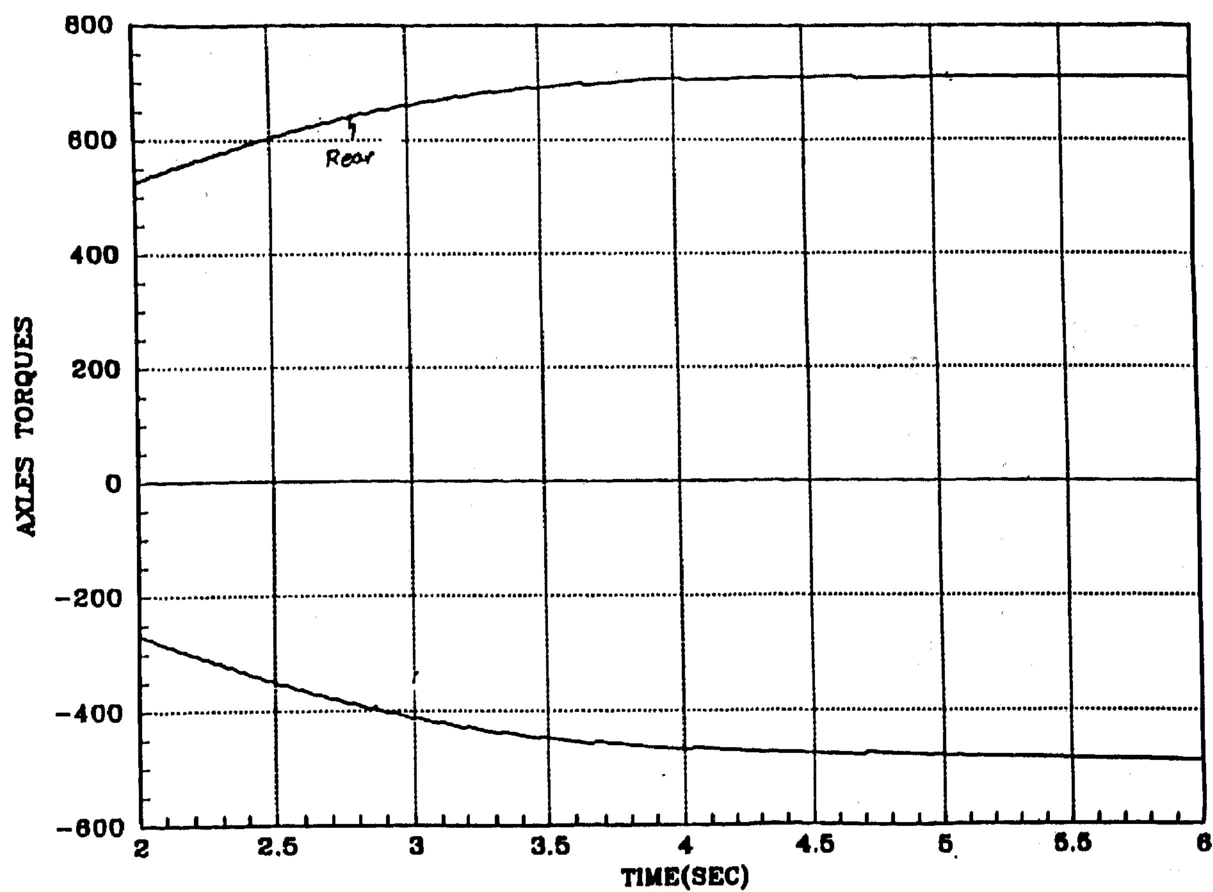


Fig. 22. As fig. 19 but with Smaller Difference in Tire Radius.

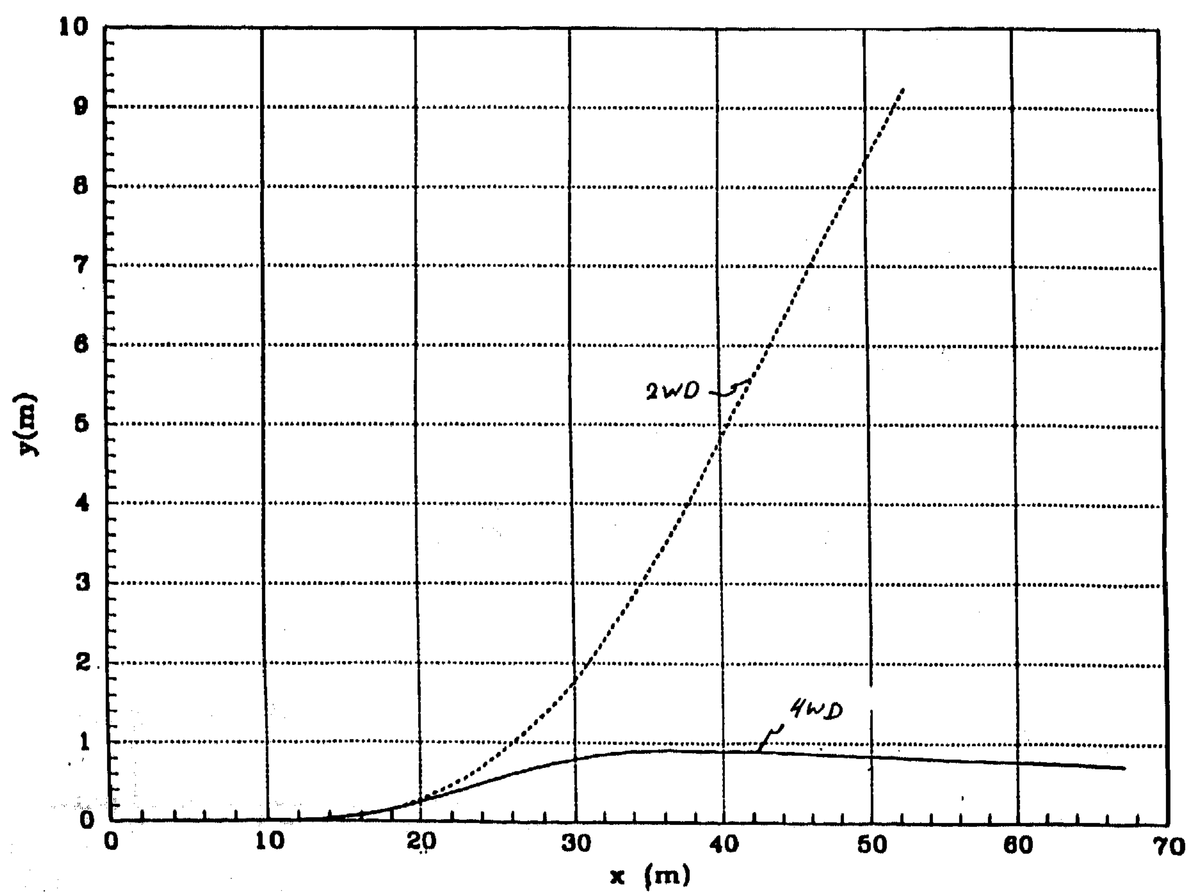


Fig. 23. Simulation Results for 2WD and Control Trac Handling Maneuver on Snow.

Although the above simulation results are very preliminary and warrant further investigations, one possible rationalization for the above phenomenon is given below. If one assumes equal loads and radii of front and rear tires -since the transfer case clutch is locked- both front and rear tires would operate at point A in Fig. 21. Increasing the rear tire radius (still we have the hub velocities $\omega_f = \omega_r$) will result in rear tires having larger percent slip than the front tires resulting in operation at point B_2 in Fig. 21. The corresponding front slip is at B_1 in Fig 21. This can also be seen from the corresponding simulated slip traces shown in Fig. 20. If the difference in tire radius is big enough it can cause the front tire to operate on the negative side of tire longitudinal force characteristics. An oscillation or chattering occurs when the rear tire forces pass over the maximum of the traction force diagram. This is illustrated by the slip traces of Fig. 20 for the above case of 7% effective tire radius difference. For the smaller difference, of 3.5%, these oscillations are not present (see Fig. 22).

6. MODEL USAGE

To study the effect of the transfer case on handling, a comparison test was done between RWD and a 4WD-equiped vehicle. The simulation was done on snow with a harsh steering input that is typical of, for example, emergency lane change maneuvers. Figure 23 shows the trajectory of the two simulations. As can be seen the 2WD vehicle was unstable while vehicle equipped with on-demand 4WD transfer case had very good handling characteristics under these circumstances.

7. CONCLUDING REMARKS AND FUTURE WORKS

A Matrixx model of an on-demand 4WD transfer case was developed and partially validated against available vehicle test data. Bond graphs were used to represent the components of the vehicle models, and causality was used to coordinate the assemblage of these components into an overall computer model. This model was used to gain an insight into the vehicle system dynamics. It was also used to to guide the overall system design and verification process. The model can be used to study the effects of parameter changes such as number of clutch plates, the number of coil turns, etc. and recommend possible improvements. Finally the model is also well suited for development and evaluation of different control strategies.

8. REFERENCES

1) D. C. Karnopp, D. L. Margolis, and R. C. Rosenberg, System Dynamics, a Unified Approach, John Wiley & Sons Inc., 1990.

2) D. Hrovat, Work notes on E8 Modeling and Control, Ford Motor Co., 1982-1984, also, "A Model of Clutch Pressure Control via Direct Feed PWM Solenoid Valve," Ford Motor co. Research Report SR-89-54, May 1989.

3) D. Hrovat and W. Tobler "Bond Graph Modeling of Automotive Power Trains," Journal of the Franklin Institute, special Issue on Bond Graph Modeling, 1991.

4) D. Hrovat, W.E. Tobler, and M.C. Tsangarides, "Bond Graph Modeling of Dominant Dynamic of Automobile Power Trains," Proceedings of 1985 ASME Winter Annual Meeting, Publication DSC Vol. 1, Nov. 1985.

5) E. Bakker, L. Nyborg, and Hans B. Pacejka, "Tyre Modeling for Use in Vehicle Dynamics Studies", SAE Paper No. 870421.

6) J. Asgari and D. Hrovat, "2D Vehicle Models Including Effects of Time Varying Load and Side Slip and Angle," Ford Motor Company Research Report, August 1990.

7) J. Asgari, "Integrated Brake and Suspension Control During Braking and Steering", Ph.D. Dissertation, Department of Mech. Engr., University of California, Davis, 1989.

970970

An Improved Real-Time Model of a Planetary Gear Train

B. Klages, R. J. Woermann, and H. J. Theuerkauf
University of Kassel

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Advanced automatic transmissions are controlled by electronic control units (ECUs) which perform the gear change and furthermore a lot of complex control and diagnostic functions. Hardware-in-the-loop simulation (HIL) is a powerful tool to develop and test the control algorithms implemented in the ECUs. The simulation is based upon mathematical models of the different power train components, adapted to the real-time requirements. In this paper an improved real-time model of an automatic transmission with a Ravigneaux planetary gear train is presented. This transmission contains among other components numerous clutches and brakes. The time-variant states of these nonlinear elements (sliding and static friction) lead to numerical problems with respect to real-time simulation. In the suggested model the transferred torques of the different clutches and brakes are calculated using the methods of multi-variable control.

INTRODUCTION

Today automotive ECUs are not only adapted to existing mechanical or hydraulic systems, like the combustion engine or the automatic transmission, but they are designed and developed as integral parts of these systems. This parallel development process is characteristic for the principles of simultaneous engineering and means a completely new working method for the engineers.

In the beginning the mechanical system may exist only as a concept, no hardware tests or measurement results are available. In a first step the general concept of control can be developed and tested using an off-line simulation. But the parallel development of the ECU and the mechanical components require tests and verifications in each stage of development with an increasing test amount. When the ECU or significant parts of the mechanical system work sufficiently, new developed control algorithms, as well as components, should be already tested under real-time conditions in the laboratory. This has significant advantages with respect to the quality of the software in the ECUs. Additional transfer errors, when programming the ECUs with detailed (off-line) simulation results, can be avoided. To combine real components of the system with simulated ones, means to simulate parts of the system in real-time.

The demand for a time and cost efficient development process leads to the following restrictions on the simulation system:

- availability of a hierarchical modeling structure, which allows a transition from a global simulation in the phase of specification up to a function specific simulation of special system components,
- availability of design methods for mathematical models, appropriate as well to off-line simulation methods as to real time computing,
- modular structured software with an interface for the implementation of user defined models,
- modular structured hardware, consisting of a basic system, which is adaptable by standard modules and scaleable computing power to different types of ECUs and simulation models.

The real-time simulation system measures the output signals of the ECUs and simulates the sensor signals of the corresponding power train components. This is the application field of the real-time simulator CARTS®, developed by the University of Kassel upon the basis of a cooperation with Volkswagen Research, Wolfsburg, Germany.

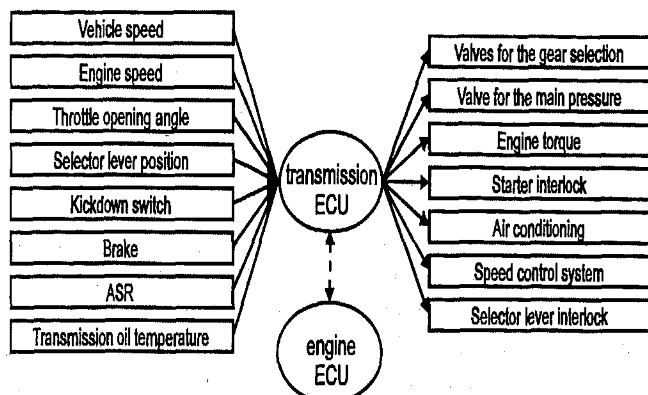


Figure 1 *Information flow between the real-time simulator and the ECU*

AUTOMATIC TRANSMISSION WITH ELECTRONIC CONTROL - Nowadays automatic passenger-car transmissions have 4 ore 5 gears. The main mechanical components are:

- a hydrodynamic (Foettinger) torque converter provided with a lockup clutch,
- a Ravigneaux planetary gear train,
- multi-disc clutches and brakes assigned to the different wheels and axles to allow gear shifting under traction,
- an oil pump for the power supply of the hydraulic shifting components and the torque converter.

The transmission control is realized by an ECU which chooses the gears and controls the oil pressure according to the torque to be transmitted. **Figure 1** shows the signal flow between the automatic transmission, the sensors, and actuators and the ECU. Electronic control allows the realization of driver dependent shift programs, the adaptation to environmental conditions, or different vehicles. The main sensors and actuators necessary for electronic transmission control are:

- load-sensors, e.g. throttle potentiometer or the according information by the engine ECU and kickdown-switch,
- speed sensors for vehicle speed and engine speed or the according information by the engine ECU,
- position sensor of the selector lever,
- oil-temperature sensors,

- solenoid valves to switch the clutches and brakes,
- modulation valve to control the oil pressure.

The transmission ECU processes the sensor information and controls shift point, shift quality, lockup clutch, and special safety features. In addition it communicates, e.g. via CAN, with the engine ECU in a mesh connected network.

Assuming a powerful simulation system and a suitable mathematical model, the interaction between the (real) transmission ECU and the (simulated) transmission can be tested and verified in the electronic laboratory, also with respect to data exchange with a (real) engine ECU.

MECHANICAL STRUCTURE OF A RAVIGNEAUX PLANETARY GEAR TRAIN

A Ravigneaux planetary gear train (*ref. [5]*) consists of a ring gear (R), a small and a large sun wheel (S1 and S2) and a short and a long planetary wheel (P1 and P2) which are attached to the planetary carrier (PC).

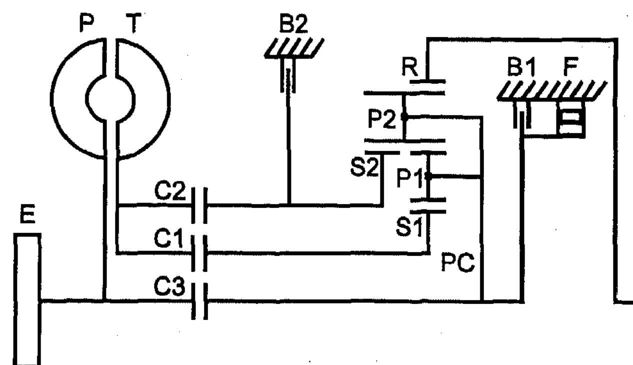


Figure 2 *Scheme of the Ravigneaux planetary gear train with engine (E) and torque converter (pump (P) and turbine (T))*

Figure 2 shows the arrangement of the gear wheels in the transmission. With the aid of the clutches C1, C2, and C3, the brakes B1 and B2, and free-wheel F it is possible to realize four forward gears and one reverse gear.

gear	B1	B2	C1	C2	C3	F
R	x			x		
1H			x			x
2H		x	x			
3H			x	x		
3M			x	x	x	
4M		x			x	

Table 1 *Combination of clutches and brakes* (*ref. [3]*)

CARTS (Computer Aided Real-Time Test System) is a registered trademark of the Institute of Power Engineering - Drive Engineering, Prof. Dr.-Ing. H. J. Theuerkauf, University of Kassel, Germany

Table 1 shows the combination of the activated clutches and brakes in the different gears. There exist further combinations (*ref. [3]*) between the third and the fourth gear which are not described in the table. They are used as transient states to improve the gear change.

NEWTON'S LAW OF MOTION FOR THE RAVIGNEAUX PLANETARY GEAR TRAIN

The differential equations of a planetary gear train can be derived using Kane's equations (Jourdains's principle). Jourdain's principle states, that the internal virtual power produced by the resulting constraint forces F^Z and torques M^Z in a system of n rigid bodies is equal to zero. The mathematical formulation yields:

$$\sum_{i=0}^{n}\left[\delta\,\vec{v}_{i,0}\cdot\vec{F}^{z}_{i,0}+\delta\,\vec{\omega}_{i,0}\cdot\vec{M}^{z}_{i,0}\right]=0 \qquad (1)$$

While the first index i characterizes the body, the second one indicates the reference vector base. The described Ravigneaux planetary gear train consists of six wheels performing only rotations and equation (1) can be reduced to

$$\sum_{i=0}^{n}\left[\delta\,\vec{\omega}_{i,0}\cdot\vec{M}^{z}_{i,0}\right]=0 \qquad (2)$$

Since the directions of all the axes of rotation and the external and internal torque coincide, the vectors $\vec{\omega}_{i,0}$ and $\vec{M}_{i,0}$ are replaced by the corresponding scalars $\omega_{i,0}$ and $M_{i,0}$.

The constraint torque $M^{z}_{i,0}$ can be obtained from Newton's equation

$$J_{i,0}\frac{d\,\omega_{i,0}}{dt}=J_{i,0}\dot{\omega}_{i,0}=M^{z}_{i,0}+M^{e}_{i,0} \qquad (3)$$

where $M^{e}_{i,0}$ represents the sum of external torques, acting on wheel i which has the moment of inertia $J_{i,0}$. Eq. (2) can now be written in the form:

$$\sum_{i=0}^{n}\left[\delta\,\omega_{i,0}\cdot\left(J_{i,0}\dot{\omega}_{i,0}-M^{e}_{i,0}\right)\right]=0 \qquad (5)$$

Writing the sum in detail for the planetary gear train yields:

$$\begin{aligned}&J_{S1}\dot{\omega}_{S1}\delta\omega_{S1}+J_{S2}\dot{\omega}_{S2}\delta\omega_{S2}+J_{PC}\dot{\omega}_{PC}\delta\omega_{PC}+\\&J_{P1}\dot{\omega}_{P1}\delta\omega_{P1}+J_{P2}\dot{\omega}_{P2}\delta\omega_{P2}+J_{R}\dot{\omega}_{R}\delta\omega_{R}\\&=M_{C3}\delta\omega_{PC}+M_{C1}\delta\omega_{S1}+M_{C2}\delta\omega_{S2}-M_{B1}\delta\omega_{PC}\\&-M_{B2}\delta\omega_{S2}-M_{R}\delta\omega_{R}\end{aligned} \qquad (6)$$

After deriving the kinematic relations between the angular velocities of the different gear wheels, the number of variables can be diminished to two, according to the number of degrees of freedom. Keeping ω_{PC} and ω_R as variables, ω_{S1}, ω_{S2}, ω_{P1} and ω_{P2} are replaced by the right terms of the following kinematic equations :

$$\omega_{S1}=\left(1-\frac{z_R}{z_{S1}}\right)\omega_{PC}+\frac{z_R}{z_{S1}}\omega_R \qquad (7)$$

$$\omega_{S2}=\left(1+\frac{z_R}{z_{S2}}\right)\omega_{PC}-\frac{z_R}{z_{S2}}\omega_R \qquad (8)$$

$$\omega_{P1}=\left(\omega_{PC}-\omega_R\right)\frac{z_R}{z_{P1}} \qquad (9)$$

$$\omega_{P2}=\left(\omega_{R}-\omega_{PC}\right)\frac{z_R}{z_{P2}} \qquad (10)$$

(z = number of teeth)

After introducing the abbreviations

$$P_1=1-\frac{z_R}{z_{S1}} \qquad H_1=\frac{z_R}{z_{S1}} \qquad (11)$$

$$P_2=1+\frac{z_R}{z_{S2}} \qquad H_2=-\frac{z_R}{z_{S2}} \qquad (12)$$

$$P_3=\frac{z_R}{z_{P1}} \qquad H_3=-\frac{z_R}{z_{P1}} \qquad (13)$$

$$P_4=-\frac{z_R}{z_{P2}} \qquad H_4=\frac{z_R}{z_{P2}} \qquad (14)$$

the substitution of Eq. (7) to (14) into Eq. (6) yields:

$$\begin{aligned}&\big(J_{S1}P_1^2\dot{\omega}_{PC}+J_{S1}P_1H_1\dot{\omega}_R+J_{S2}P_2^2\dot{\omega}_{PC}+J_{S2}P_2H_2\dot{\omega}_R\\&+J_{PC}\dot{\omega}_{PC}+J_{P1}P_3^2\dot{\omega}_{PC}+J_{P1}P_3H_3\dot{\omega}_R+J_{P2}P_4^2\dot{\omega}_{PC}\\&+J_{P2}P_4H_4\dot{\omega}_R\big)\cdot\delta\,\omega_{PC}\\&+\big(J_{S1}P_1H_1\dot{\omega}_{PC}+J_{S1}H_1^2\dot{\omega}_R+J_{S2}P_2H_2\dot{\omega}_{PC}+J_{S2}H_2^2\dot{\omega}_R\\&+J_R\dot{\omega}_R+J_{P1}P_3H_3\dot{\omega}_{PC}+J_{P1}H_3^2\dot{\omega}_R+J_{P2}P_4H_4\dot{\omega}_{PC}\\&+J_{P2}H_4^2\dot{\omega}_R\big)\cdot\delta\,\omega_R\\&=\left(M_{C3}+M_{C1}P_1+M_{C2}P_2-M_{B1}-M_{B2}P_2\right)\cdot\delta\,\omega_{PC}\\&+\left(M_{C1}H_1+M_{C2}H_2-M_{B2}H_2-M_R\right)\cdot\delta\,\omega_R\end{aligned} \qquad (15)$$

With further abbreviations

$$a_{11} = J_{S1}P_1^2 + J_{S2}P_2^2 + J_{PC} + J_{P1}P_3^2 + J_{P2}P_4^2$$
$$a_{12} = J_{S1}P_1H_1 + J_{S2}P_2H_2 + J_{P1}P_3H_3 + J_{P2}P_4H_4$$
$$a_{21} = J_{S1}P_1H_1 + J_{S2}P_2H_2 + J_{P1}P_3H_3 + J_{P2}P_4H_4$$
$$a_{22} = J_{S1}H_1^2 + J_{S2}H_2^2 + J_R + J_{P1}H_3^2 + J_{P2}H_4^2$$
$$M_1 = M_{C3} + M_{C1}P_1 + M_{C2}P_2 - M_{B1} - M_{B2}P_2$$
$$M_2 = M_{C1}H_1 + M_{C2}H_2 - M_{B2}H_2 - M_R$$

the equations of motion for the planetary gear train result:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \dot{\omega}_{PC} \\ \dot{\omega}_R \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} \tag{16}$$

CALCULATION OF INTERNAL TORQUE UNDER REAL-TIME CONDITIONS

The considered Ravigneaux planetary gear train contains three clutches and two brakes, allowing the gear shift under traction. In order to solve Newton's equation of motion for the planetary gear train (Eq. (16)), it is necessary to determine the transferred torque of clutches and brakes (M_{C1}, M_{C2}, M_{C3}, M_{B1}, M_{B2}) with respect to their actual state.

A HIL simulation requires mathematical models with appropriate size and structure in order to fulfill the real-time conditions. Real-time simulation means, that the computing of the mathematical equations must be finished in a given interval of time (e.g. in several milliseconds if integrating an ECU as real part).

Apart from the computing power of the simulation system, which has to be fitted to the model extent, the model structure itself must guarantee a determinate computing time. For this reason time-variant model structures and numerical methods, which require iterative calculations have to be avoided, because they are characterized by variable and unpredictable execution times, which could exceed the limited real-time interval.

A friction clutch is an example for a system with a time variant structure. If the friction discs slip the device is characterized by two degrees of freedom because the discs rotate independently. During adhesion the clutch represents a one mass system.

If the clutch slips the internal torque M_C is limited to the "torque capacity" $M_{C,cap}$ and is assumed as nearly independent on the rotational speed difference (**Figure 3**). $M_{C,cap}$ is a function of the clutch's surface coating, as well as a function of the pressure force acting on the discs. The torque capacity will become zero if the clutch is released. If the clutch adheres, the transferred torque is a function of the external torque acting on the discs.

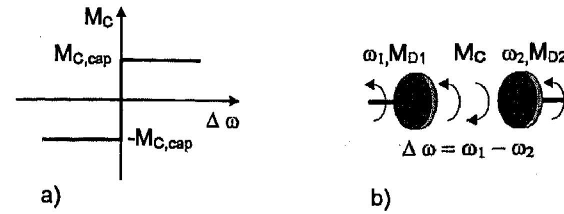


Figure 3 *a) Characteristic curve for static ($\Delta\omega = 0$) and sliding friction ($\Delta\omega \neq 0$) b) Speed and torque variables*

If the states and the external torques are known, the internal torque can be derived exactly from the equations of motion of the planetary gear train. A problem occurs if there are state transitions during gear change. In this case the states and torques have to be determined in two or more steps until all states remain stable. This iterative procedure should be avoided for real-time simulation. Therefore other approaches were developed in the past to realize simplified real-time models of the clutch. These models are also suited as brake models with the difference that one disc is fixed.

REAL-TIME MODELS OF A CLUTCH

Most real-time models of the clutch calculate the torque as a function of the rotational speed difference between the clutch discs, regarding the discs as separate masses even during static friction.

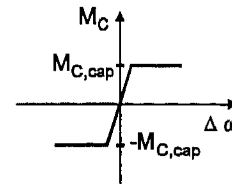


Figure 4 *Simplified characteristic curve*

Figure 4 shows a simple approach (*ref. [2], [4]*) represented by a limiter which models static friction by a proportional part for little speed differences ($M_C = c\cdot\Delta\omega$). While the clutch slips the torque is limited to the torque capacity "$M_{C,cap}$" ($|M_C| = M_{C,cap}$).

A block diagram, representing this model, is shown in **Figure 5**. The speed difference between the clutch discs is calculated by an integrator as a function of the internal clutch torque M_C and the external torque M_D acting on the clutch discs.

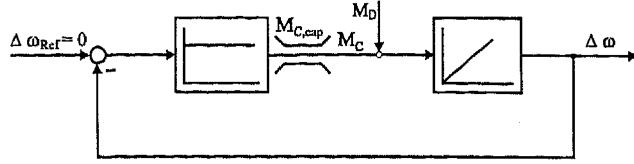


Figure 5 *P-algorithm with limiter (ref. [2],[4])*

The P-algorithm with saturation in **Figure 5** can be assumed as a "P-controller" with limiter because the torque M_C acts as "manipulated variable", which decreases the "system deviation", represented by the rotational speed difference $\Delta\omega$. The amplification of the controller, which is ideally infinite, must be limited to values not leading to numerical instability. This kind of torque determination has the disadvantage that the slip is not zero during static friction, even in steady-state, though the clutch discs physically adhere.

An obvious model improvement to decrease the "control error", i.e. the speed difference $\Delta\omega$, is the replacement of the P- by a PI-algorithm. This leads to a neglectable steady state speed difference during static friction. Nevertheless, if the external torque M_D, that is acting on the clutch, changes dynamically, there will be a transient speed error, until the PI-algorithm has adapted the clutch torque to its new steady-state value.

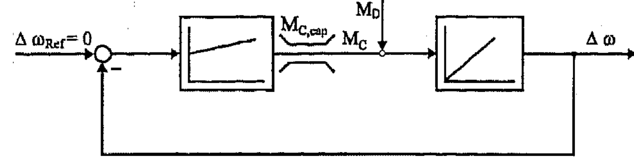


Figure 6 *PI-algorithm with limiter*

IMPLEMENTATION IN A MODEL OF A RAVIGNEAUX PLANETARY GEAR TRAIN

The implementation of the PI-algorithm in the planetary gear train model leads to five "PI-control-loops" calculating the torques of the clutches C1, C2 and C3 and the brakes B1 and B2. A remaining problem of this gear train model is the fact that the five "control loops" interact during simulation and form a complex dynamic system (**Figure 7**). Each torque does not only act as "manipulated variable" in its own "control loop" but also as a "disturbance" in the other ones.

Due to the internal coupling it is difficult to optimize each control loop with regard to stability, exactness, damping and speed. Moreover, in order to guarantee the stability of the numerical integration, the step size has to be adapted to the eigenvalues of the total system which are also determined by the PI-parameters.

Several methods exist to adapt controllers in a multi-variable control system. If the control loops can be completely decoupled, the problem is reduced to the adaptation of a single variable control loop. After decoupling the control loops a compensation of measurable or known disturbances would additionally improve the control.

Regarding the simulation system of a planetary gear train as a multi-variable control structure leads to the block diagram in **Figure 7.** It can be optimized, using the methods of decoupling and feedforward.

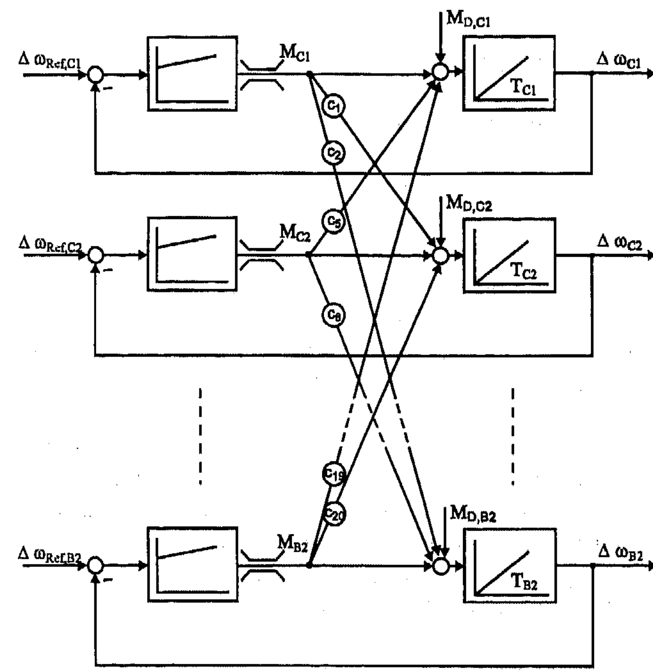


Figure 7 *Interaction of clutches and brakes*

DECOUPLING AND FEEDFORWARD CONTROL

Figure 8 shows an example of a control system with two inputs and outputs. Each input acts on both outputs via the transfer functions p_{11} and p_{12} respectively p_{12} and p_{22}. With the aid of the additional transfer elements k_{12} and k_{21} both control loops can be dynamically decoupled with the aim that a diagonal transfer matrix $\underline{Q}$ results. The Laplace transform results in the following matrix equation (*ref.[6]*):

$$\underbrace{\begin{bmatrix} p_{11}(s) & p_{12}(s) \\ p_{21}(s) & p_{22}(s) \end{bmatrix}}_{\underline{P}(s)} \underbrace{\begin{bmatrix} 1 & k_{12}(s) \\ k_{21}(s) & 1 \end{bmatrix}}_{\underline{K}(s)} = \underbrace{\begin{bmatrix} q_{11}(s) & 0 \\ 0 & q_{22}(s) \end{bmatrix}}_{\underline{Q}(s)}$$

The coefficients of matrix $\underline{K}(s)$ result from:

$$\underline{K}(s) = \underline{P}(s)^{-1}\underline{Q}(s) \quad \Rightarrow k_{ij}(s) = (-1)^{i+j}\frac{\det \underline{P}_{ji}(s)}{\det \underline{P}_{jj}(s)}$$

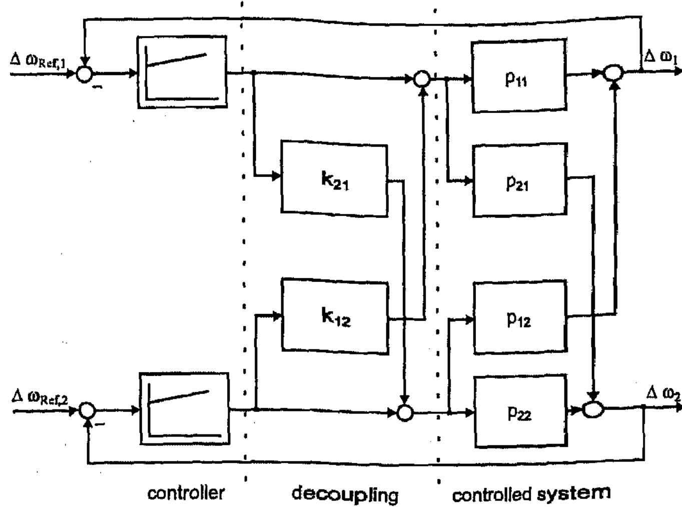


Figure 8 *Decoupling of interacting control loops*

After decoupling the control loops, measurable disturbances with a determinable disturbance point can be compensated, using feedforward control (**Figure 9**). In the example of **Figure 9** an ideal feedforward control must fulfill the equation

$$f_j(s) = -q_{ii}^{-1}(s)\, d_j(s)$$

in order to compensate the disturbance M_{Dj} completely. After the addition of decoupling and feedforward elements the dynamic of the resulting system is only determined by the poles of the transfer functions $q_{ii}(s)$. Therefore the algorithm can be adapted to a given integration step size or to a requested closed-loop behavior.

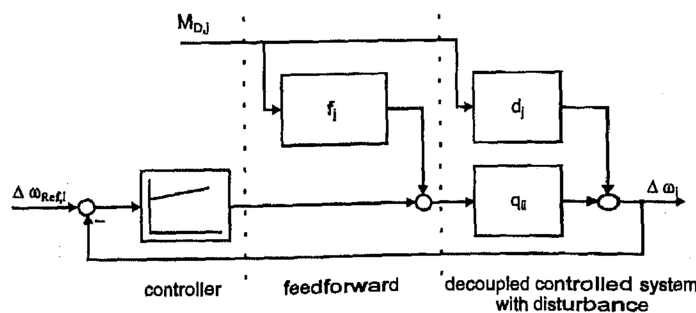


Figure 9 *Feedforward control of disturbances*

APPLICATION TO THE PLANETARY GEAR TRAIN

The decoupling matrix depends on the actual states (static or sliding friction) of clutches and brakes. If a clutch/brake slips the torque will be limited to the torque capacity $M_{C,cap}$ and will be represented by a real two mass system, which forms an open loop in the simulation model. Only adhering clutches and brakes lead to closed and consequently active "control loops" which can be decoupled. On the other hand the torque of an open loop can be regarded as a disturbance for the remaining active loops and can be compensated using feedforward control. The case of releasing the clutch is equivalent to the case of sliding friction setting the torque capacity to zero.

The torques of clutches and brakes are further determined by the external torques from engine, turbine, and drive shaft, acting on the planetary gear train. These external torques appear as further disturbances in the "control loops". Their effect on the control error can also be eliminated by feedforward control.

As an example the decoupling and feedforward control is shown in the second gear (**Figure 10**). The clutch C1 and brake B2 adhere, while C2, C3 and B1 are released. The control algorithms to calculate M_{C1} and M_{B2} are dynamically decoupled by the coefficients k_{21} and k_{12}. The feedforward coefficients f_{T1}, f_{T2}, f_{R1} and f_{R2} compensate the effects of the external torques from turbine (M_T) and gearbox output (M_R), which would otherwise cause faulty speed differences $\Delta\omega_{C1}$ and $\Delta\omega_{B2}$ in the simulation. The engine torque has no influence on the speed differences of clutch C_1 and brake B_2.

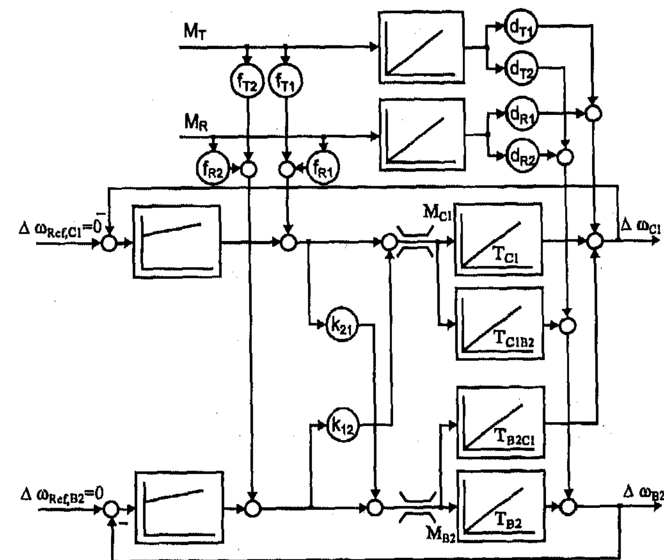


Figure 10 *Decoupling and feedforward control in the second gear*

In order to minimize the computing time, the matrices for the decoupling and feedforward control are calculated in an initialization routine before the real-time simulation.

HIL-SIMULATION OF A DRIVE TRAIN WITH AN AUTOMATIC TRANSMISSION

The transmission models are implemented in a drive train model, consisting of an engine, a Foettinger torque converter and a drive shaft with load (*ref. [8]*). The models of the engine and the torque converter are steady-state models based upon measured maps. The drive shaft is modeled as a damped spring, while the load model simulates the vehicle's aerodynamic drag, the rolling resistance, and the gradient of the road. The drive train model is implemented in MATLAB/ SIMULINK® as well as in the real-time simulator CARTS.

The HIL-simulator CARTS (**Figure 11**) is generally described in *ref. [7]*. Therefore this chapter concentrates on the specific simulator equipment for transmission-ECUs.



Figure 11 *Real-time simulator CARTS*

The transmission simulator consists of a VMEbus system with a separated I/O-subsystem and a PC. The VMEbus system based on OS-9® is the real-time unit of the simulator which is controlled by a Windows-PC. The VMEbus computer contains a master CPU (MC68060) executing the simulation program and four I/O-CPUs (MC68030, MC68340) for the parallel processing of the I/O-data. In dependence on the model extend sample rates of 2 ms can be reached.

The simulator is adapted to the required I/O-signals of the transmission ECU, using Mezzanine-modules (M-modules) for signal generating and measurement. Level adaptation is done in the I/O-subsystem which is placed in a separate box. It also contains an adjustable power supply, modulation valves as real parts and power resistors for load imitation.

SIMULATION RESULTS

A result of a HIL drive train simulation using CARTS is presented in **Figure 12**. It shows four variables during the gear change from second to third. After the clutch C2 adhered and brake B2 was released, the speed of the turbine, of the large sun wheel, and of the planetary carrier coincide. The torque at the gear box output shows several peaks which are caused by the elasticities of the drive shaft. Equivalent simulation results were validated by test-bench measurements.

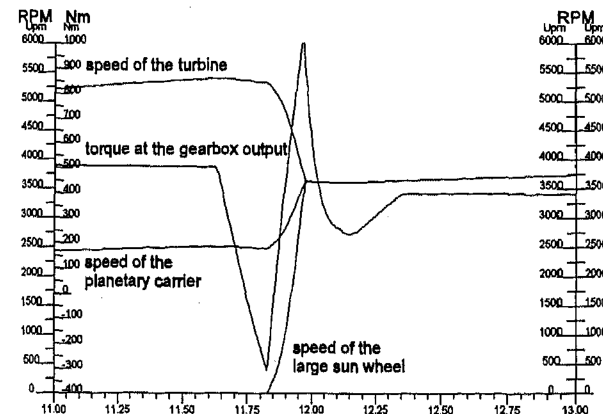


Figure 12 *Result of a real-time simulation*

Figure 13 points out the effects of the interacting control algorithms on the simulation results. The parameters of the appropriate PI-algorithms were experimentally optimized with respect to damping and stability. The plot shows as an example transient characteristics during the gear change from first to second. Though the clutch C1 actually adheres there exists a speed difference between the turbine and the small sun wheel in the simulation.

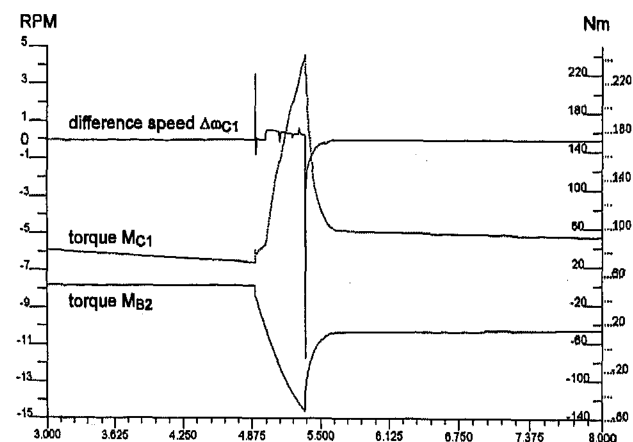


Figure 13 *Gear change without decoupling*

After applying decoupling and feedforward algorithms the parameters were individually determined for each clutch and brake in a theoretical analysis in advance. This routine also respects certain restrictions of the step size either by numerical stability of the mathematical model or the requirements of real-time processing. Using this method damping and stability of the PI-loops are guaranteed. As a result the faulty speed difference in the simulation between the adherent discs was significantly decreased (**Figure 14**).

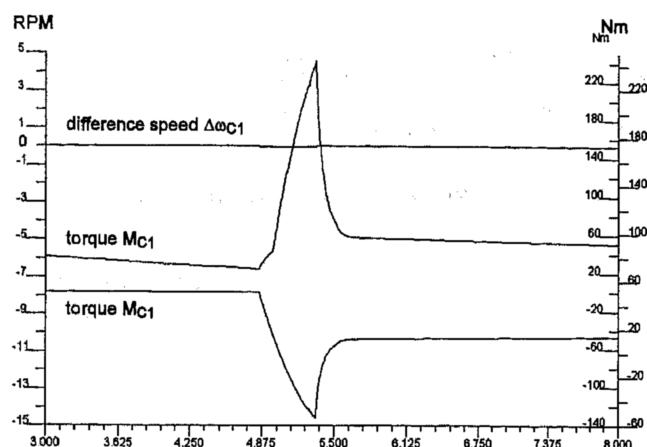


Figure 14 *Gear change with decoupling and feedforward*

SUMMARY

In this paper a real-time model of a Ravigneaux planetary gear train is presented which uses methods of the multi-variable control theory to calculate the internal torques of clutches and brakes. First the mechanical structure of the automatic transmission is described and the equations of motion are derived. Simplified models of a clutch/brake are explained which are suitable for real-time applications. The numerical problems of their implementation in a planetary gear train model are discussed and as a consequence an improved model is suggested. Simulation results of different model implementations are finally shown and compared.

ABBREVIATIONS

Variables

a, P, H	general abbreviation
c	coefficient
F	force
f, d, k	transfer functions
J	moment of inertia
M	torque
n	number
p	plant transfer functions
s	Laplace variable
t	time
T	time constant
v	velocity
ω	angular velocity
$\Delta\omega$	rotational speed difference between the clutch discs
z	number of teeth

Indices

0	reference vector base
B	brake
C	clutch
D	deviation
e	external force and torque
cap	capacity
i,m	general indices
z	constraint force and torque
P1	short planetary wheel
P2	long planetary wheel
PC	planetary carrier
R	ring wheel
S1	small sun wheel
S2	large sun wheel
T	turbine

REFERENCES

[1] Automotive Handbook, Robert Bosch GmbH, Stuttgart, 1996

[2] C. Klarhoefer et al.: "Real time engine and transmission simulator" ATA - New Frontiers for Automotive Electronics; 24./25. 03. 93; Turin, Italy

[3] L. Prüß: "Das neue Viergang- Automatikgetriebe von Volkswagen", ATZ, 1990

[4] G. Rill: "Simulation von Kraftfahrzeugen", Vieweg Verlag, Braunschweig/Wiesbaden, 1994

[5] J. Stüper: "Automatische Automobilgetriebe", Springer Verlag, Wien-NewYork, 1965

[6] H. Tolle: "Mehrgrößenregelkreissynthese", Oldenbourg Verlag, München-Wien, 1983

[7] R.J. Woermann et al.: "CARTS-A Hardware in the Loop Simulator for Test and Development of Automotive Control Units", SAE-No.961019, Detroit, 1996

[8] R.J. Woermann: "Ein Beitrag zur Echtzeitsimulation technischer Systeme hoher Dynamik mit diskreten Modellen" Dissertation, Universität Kassel, 1994

ABOUT THE MAIN AUTHOR

B. Klages is a member of the Institute of Power Engineering - Drive Engineering, University of Kassel, Germany. Klages works in the field of real-time simulation for automotive applications since 1993.

Address:	University of Kassel IEE-Antriebstechnik 34109 Kassel, Germany
Phone:	+49(561)804-6354
FAX:	+49(561)804-6378
e-mail:	kla@at.e-technik.uni-kassel.de

970971

Gear Contour Hardening by Micropulse® Induction Heating System

Yoshitaka Misaka, Yutaka Kiyosawa, Kazuhiro Kawasaki, and Takao Yamazaki
Neturen Co., Ltd.

Wilson O. Silverthorne II
Contour Hardening, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Contour hardening of small gears, such as automotive transmission gears, is beneficial due to the higher strength characteristics and lower distortion. In the past, conventional induction hardening was not used due to its inability to produce a contour case pattern. However, carburizing also has inherent problems such as abnormal carburized layer, distortion due to high temperatures and long heating time, and retained austenite.

The Micropulse system, using both high and low frequencies, has been developed to alleviate some of the problems with traditional induction hardening. This Micropulse process makes it possible to produce a contour hardened pattern on a variety of gearing. This process is very rapid, as final heat time normally does not exceed 0.250 seconds.

The advantages to the Micropulse process include: high hardness, fine microstructure, high residual compressive stress, and low distortion. In addition to these benefits the process is environmentally friendly, energy efficient and easily adaptable to in-line processing.

INTRODUCTION

Induction hardening is a surface hardening method of directly heating a component using a high frequency induced current. The Micropulse process has been found to improve wear resistance, fatigue strength and toughness of a steel component. In addition to these mechanical property advantages, the process is also very clean and environmentally friendly.

The unique characteristics that differentiate this process from traditional induction heating are the very rapid heating (<0.250s) and contour surface hardening. With traditional induction hardening, small module gears which are used in automotive transmission gearing are hardened through the entire tooth section. Because of the through hardened gear, it is understandable why higher strengths and lower distortion are not achievable for some applications. These disadvantages have led to the continued used of carburizing in small module gears. By carburizing the small module gears, a contoured surface hardening results, lending itself to higher strength gears. The diffusive carburizing process also has problems which are detrimental to gear life such as: abnormal carburizing layer, distortion due to high temperatures and long heat times, and retained austenite.

To alleviate some of the downfalls of traditional induction hardening, the Micropulse system has been developed which uses both high and low frequencies high power generators. This process makes it possible to achieve contour hardening of small module gears. This contour is achieved by a suitable pre-heating time and a very quick final heating time of approximately 0.10 - 0.25 seconds.

The resulting process produces higher fatigue strength due to higher hardness and lower distortion. Other advantages to the process are the energy savings and the ease of in-line processing.

Due to the rapid heat times of the Micropulse process, the prior microstructure plays a very important role in obtaining a uniform martensitic case pattern. The Micropulse process requirements will be discussed in the following report.

Contour Hardening Technical Development - According to patent applications, the contour hardening process was developed by "heating" using dual frequencies or "using two steps for heating" in Japan thirty years ago as shown in Figure 1.

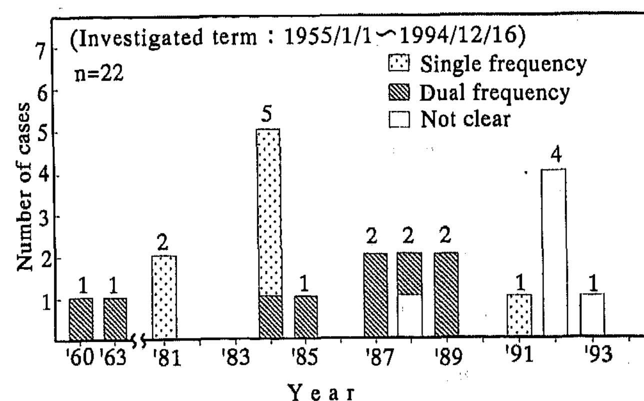


Figure 1 - Patents in Japan relating to the induction hardening process

The process did not advance during this period due to carburized gearing having higher overall strength. The technology was not available to adequately support the dual frequency method as it is today. Because the process lacked the technological support, the components produced using this method showed excessive distortion and lower strength than

carburized components. Due to this fact, this method was shadowed by carburizing for years. Within the past ten years, the contour hardening technology has advanced enough that induction is the preferred method to carburizing in some applications. One of the important breakthrough in technology other than better power supplies was the Micropulse system developed by Contour Hardening, Inc. USA. This process allows very accurate timing of high output generators with the patented A.P.C. (Absolute Pulse Controller). It is with these types of technological advances that induction can outperform carburized gearing. Induction hardening is in the infant stage of development and more research is needed to further improve this promising process.

Specifications of Contour Hardened Gear - Basic testing of the contour hardening process were conducted as shown in Table 1. The standard gear used for this test had O.D. of 126 mm., thickness of 20mm, and a module of 3. The evaluations of such tests were performed at Neturen's facility and the fatigue tests were performed in collaboration with Ryukoku University.

Table 1 - Chemical Composition and Dimensions of Test Gear

wt%

	C	Si	Mn	Cr	Mo	P	S
SAE1045	0.46	0.19	0.64	—	—	0.020	0.019
SAE4120	0.20	0.20	0.69	1.05	0.15	0.013	0.021

Module	3
Number of teeth	40
Pressure angle	20°
Addendum diameter	126
Tooth width	20
Helix angle	0

COMPARISON OF CONTOUR HARDENING VERSUS OTHER GEAR HEAT TREATING

Four types of gear hardening methods were compared: Micropulse contour hardening, soft nitriding, conventional induction hardening and carburizing. The results from each of these processes are shown in Figure 2. The root case depth was 0.5mm for the contour hardened and carburized gearing. The conventional induction hardening root case depth was 1.0mm. The soft-nitrided gear had a root case depth of 0.022mm.

The resulting microstructures, with the exception of the nitrided gear, all exhibited a uniform martensitic microstructure. The soft-nitrided gear microstructure was $Fe_3N + \gamma$ compound layer. The contour hardened gear showed the highest residual compressive stress and the least amount of distortion after heat treatment. The conventional induction hardened sample was through hardened from the tip down as shown in Figure 2.

The residual compressive stress in the root fillet was measured for all four samples by x-ray diffraction stress apparatus. The results of this test showed the Micropulse contour induction hardened sample to exhibit the highest residual compressive stress followed by carburizing, conventional induction hardening and finally soft-nitriding. The contour induction hardened sample was three hundred percent higher in residual compressive stress than the carburized gear. Figure 3 shows the compressive stress vs. depth of hardened zone for the contour induction hardened and carburized gears.

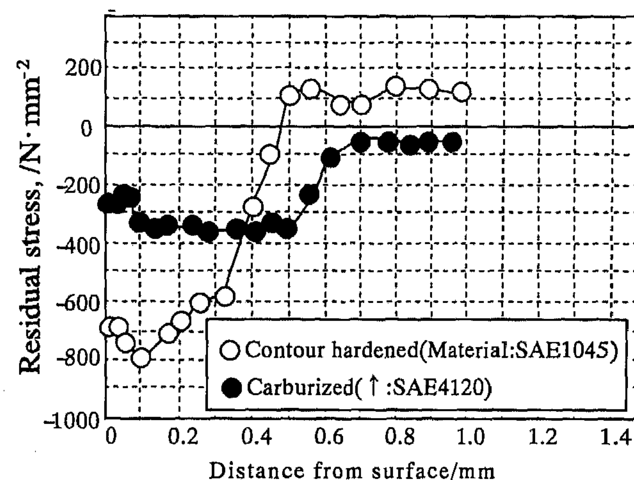
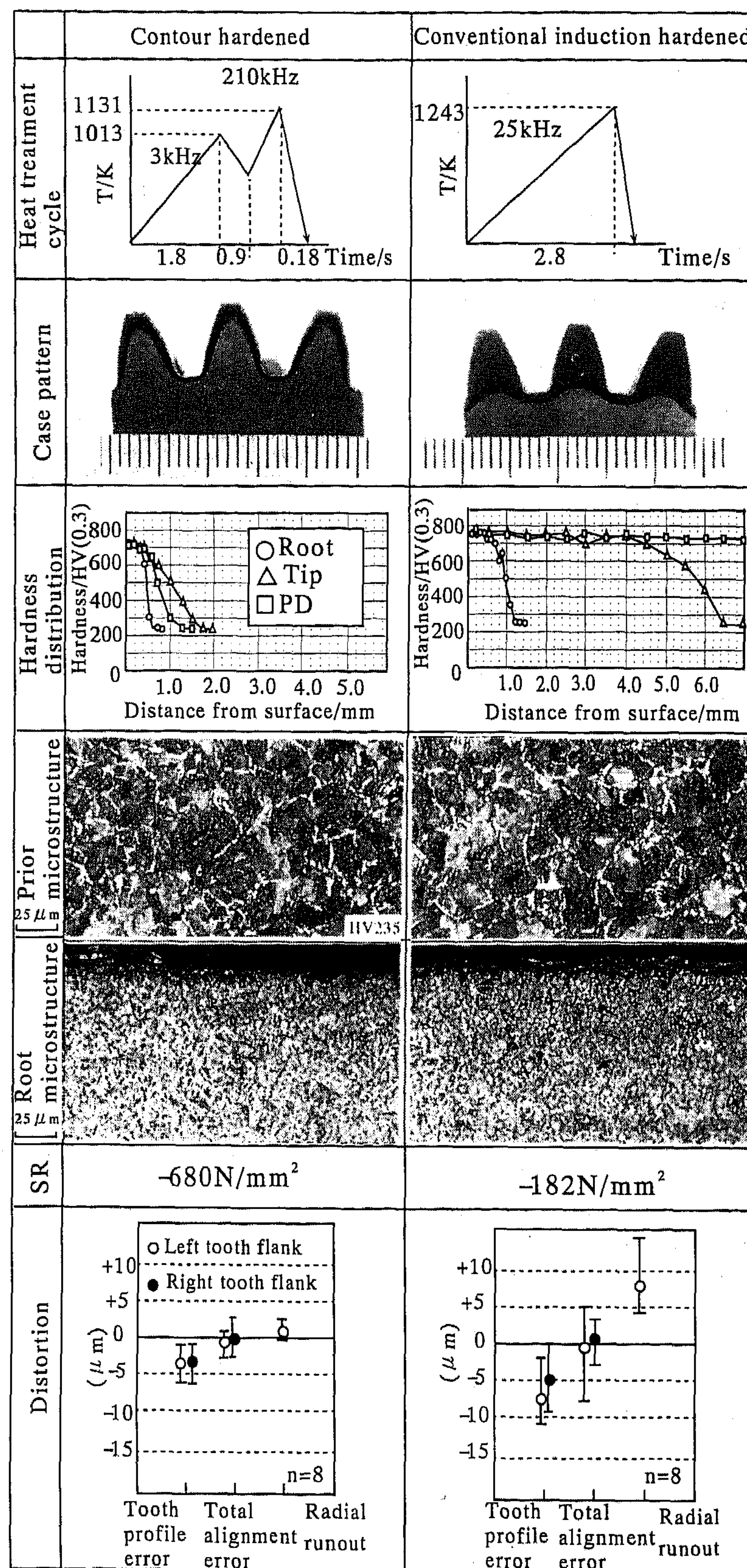


Figure 3 - Root Radius Residual Stress Comparison between Contour Hardening and Carburizing

Figure 2 also shows the amount of gear movement in tooth profile error, total alignment error, and radial run out. The soft-nitrided process was used when distortion had to be kept to a minimum while maintaining a hard wear resistance surface. The contour hardened sample showed distortion movement very similar to that of soft-nitriding. The contour hardened sample not only has low distortion, but strength unlike the soft-nitrided sample.

Tooth bending fatigue results for each of the gears tested are shown in Figure 4. The contour hardened gear was superior in strength to the soft nitrided sample and the conventional induction hardened gear and slightly stronger than the carburized gear. The high fatigue strength of the contour hardened gear is due to its very high residual compressive stress in the root fillet area. The apparatus used to test for bending fatigue can be seen in Figure 5.



SR: Root radius residual stress tangent to surface

Figure 2 - Comparison of Four Different Heat Treat Processes

	Carburized	Softnitrided
Heat treatment cycle	1183 Carburizing, Pre-heating, Diffusion, 1113 Soaking, Oil cooling, 433, CP1.0%, CP0.8%, T/K, 30, 70, 50, 40, 120, Time/min	NH_3 +Endothermic conversion gas, 853, T/K, 90, Time/min
Case pattern		————
Hardness distribution	Hardness/HV(0.3), 800, 600, 400, 200, 0, ○Root, △Tip, □PD, 1.0, 2.0, 3.0, 4.0, 5.0, Distance from surface/mm	Hardness/HV(0.025), 800, 600, 400, 200, 0, 1.0, 2.0, Distance from surface/mm
Prior microstructure 25 μm	————	————
Root microstructure 25 μm		
SR	-290N/mm^2	+12N/mm^2
Distortion	(μm), +10, +5, 0, -5, -10, -15, n=8, Tooth profile error, Total alignment error, Radial runout	(μm), +10, +5, 0, -5, -10, -15, n=8, Tooth profile error, Total alignment error, Radial runout

SR: Root radius residual stress tangent to surface

Figure 2 (cont). - Comparison of Four Different Heat Treat Processes

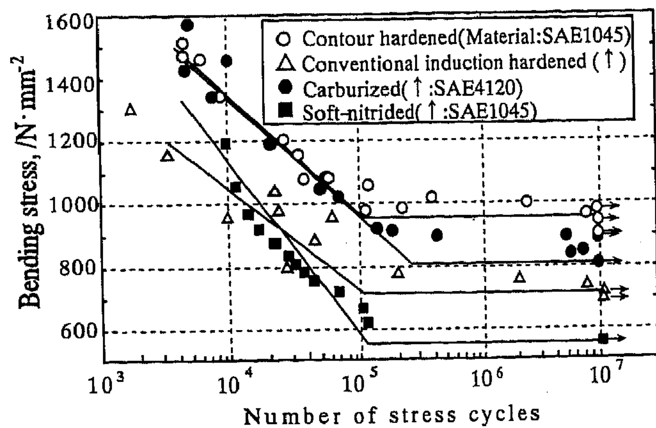


Figure 4 - Bending Fatigue test results of Heat Treated Gears

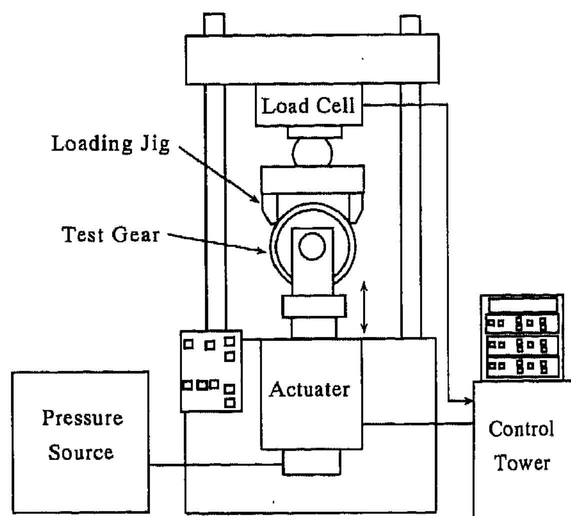


Figure 5 - Bending Fatigue Test Apparatus

The single overload bending test showed different results than the bending fatigue test. The strongest gear in the single overload bending test was the carburized gear followed by the Micropulse contour hardened, conventional induction hardened and soft nitriding. The test results are shown in Figure 6.

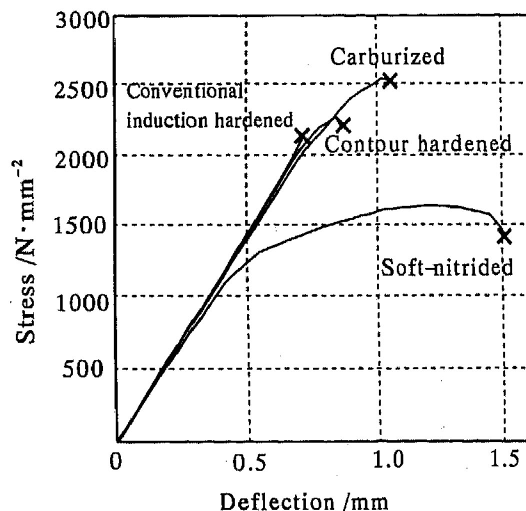


Figure 6 - Single Cycle Overloading Bending Test

The reason the carburized gear strength is greater in the overload test compared to the contour hardened gear is the core strength of each sample. The core hardness for the carburized gear was HV 350 where the contour induction hardened sample was HV 235. The core of the carburized sample was able to withstand higher subsurface stress due to the higher core hardness. To achieve the same level of performance, the core hardness of the contour induction hardened sample would have to be raised to a higher hardness. It is predicted that the performance would be much greater than the carburized gear if the core hardness was as high as the carburized sample. With this knowledge the Micropulse contour induction hardened gear has both higher strength and lower distortion.

Relationship of heat treatment condition and heat treatment quality - The Micropulse contour hardening is obtained by the use of the A.P.C. which accurately controls the time ($5X10^{-6}$sec.), and a high power generator using dual frequencies. The two frequencies are not administered simultaneously, but at two different times with a short dwell between them. To establish this new heat treating technique, the influence of each heat treating condition was compared to the resulting quality of the component. In Figures 7 and 8, a relationship between the heat treating processes and hardened depth, residual stress and distortion can be seen. These results show that when a contoured pattern is achieved or hardening depth is shallower, the component distorts less and has a higher residual compressive stress.

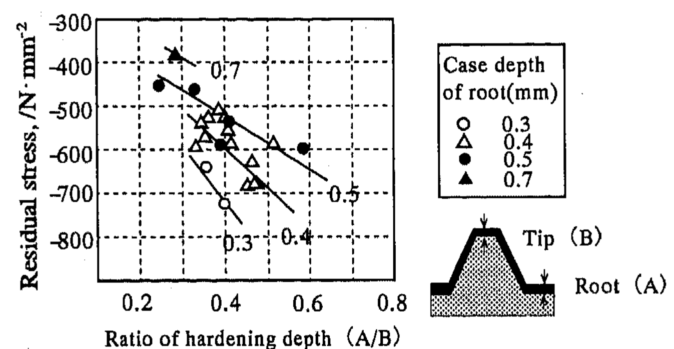


Figure 7 Case Depth versus Residual Stress

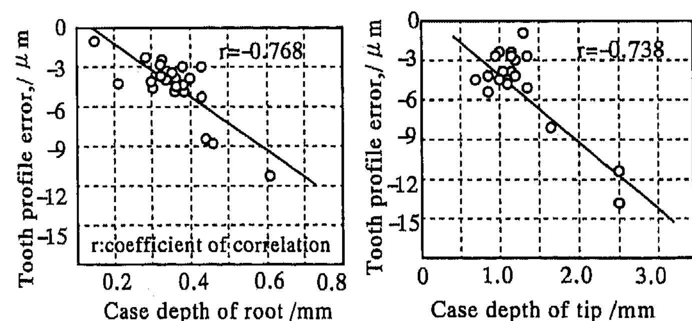


Figure 8 Case Depth versus Distortion

INFLUENCES OF PRIOR MICROSTRUCTURE ON CONVENTIONAL INDUCTION HARDENING

The final mechanical properties of an induction hardened component is influenced by prior microstructure. The Micropulse process is even more susceptible to prior microstructure due to the short heating times involved. In a typical Micropulse process, the time of heating is less than 0.5 seconds.

The material used for these tests was S45C or SAE1045.

The prior microstructure of this steel was controlled to several prior microstructural conditions. These conditions can be seen in Table 2.

Table 2 - Nomenclature of Microstructures

Symbol	Microstructure
F+P(F20)	20% Ferrite in Pearlitic Matrix
F+P(F32)	32% Ferrite in Pearlitic Matrix
F+P(F45)	45% Ferrite in Pearlitic Matrix
Q&T	Tempered Martensite
F+Spheroidal	Ferrite with spheroidal cementite

Figures 9 through 11 show the influence of prior microstructure on conventional induction hardening. Figure 9 shows the relationship between prior microstructures heating temperature and resulting hardness. The F32 and F+Spheroidal samples have low hardness at low temperatures and the hardness across the range is not consistent. The Q&T microstructure seen in the graph maintains a high level of hardness across the range of temperatures listed.

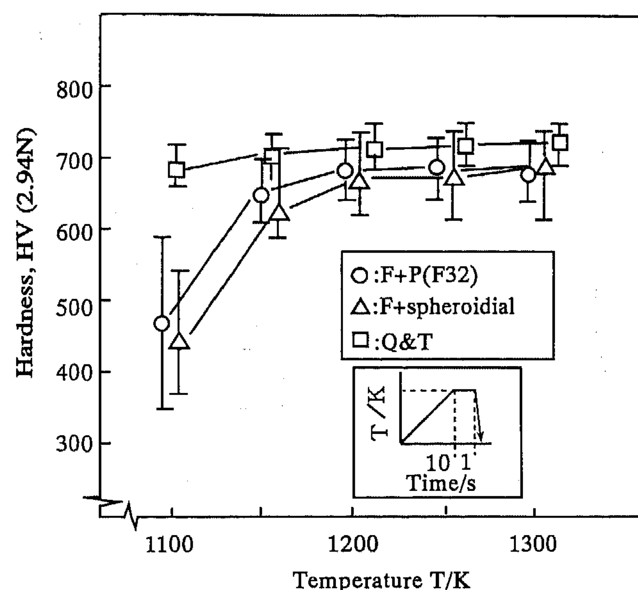


Figure 9 - Heating Temperature versus Hardness of Conventional Induction Hardened Specimens

The relationship between microstructure and lower limit of the TTA diagram can be found in Figure 10. The Q&T sample can be austenitized at a lower temperature and shorter time than microstructures that contain large amounts of ferrite or spheroidal cementite. The larger the amount of ferrite, the higher temperatures and longer heating time must be achieved before the sample can be austenitized. The same is true for spheroidal cementite as seen in Figure 10.

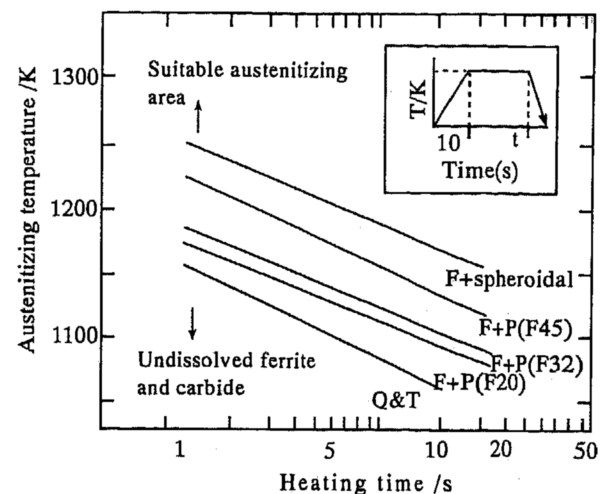


Figure 10 - Conventional Induction Hardening TTA (Time-Temperature-Austenitization) Diagram versus Microstructure

Figure 11 shows a microhardness traverse of all four microstructures given the same heating conditions. The Q&T microstructure showed the highest level of hardness and was deeper to an HV513. As the amount of ferrite increased, the hardness was lower and case depth shallower. The residual compressive stress followed trends similar to the case hardness analysis. As depicted by the data, induction hardening is highly dependant on prior microstructure. It is difficult to austenitize the material when there is a large amount of ferrite due to the long diffusion distances needed to attain a homogenous austenitic solid solution.

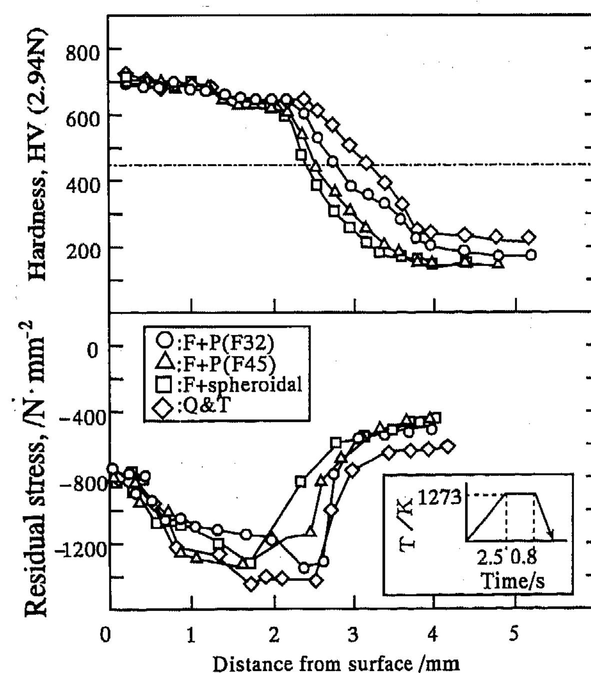


Figure 11 - Hardness and Residual Stress Profile of Induction Hardened Steel (ϕ30)

Some initial research has been done to support the theory that the larger the cementite sphere the longer time and higher temperature needed to dissolve the microstructure. In Figure 12, the radius of cementite that will dissolve under given conditions was simulated. From this data, we can conclude that the TTA curve is higher for large spheroidal cementite compared to smaller

cementite.

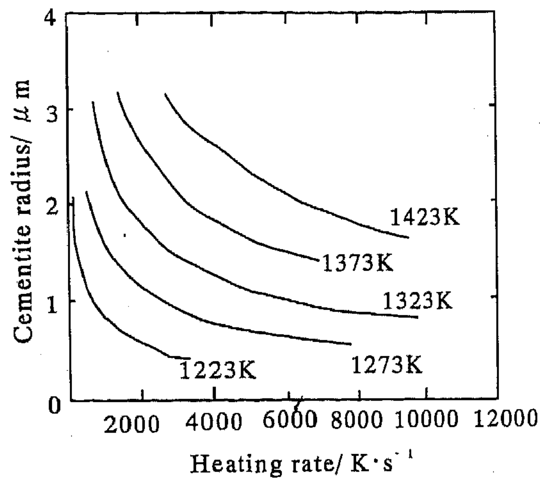


Figure 12 - Critical Cementite Radius which will Dissolve during Heating as a Function of Heating Rate

INFLUENCE OF PRIOR MICROSTRUCTURE ON CONTOUR INDUCTION HARDENING

Due to the short heating time, the contour hardening process is most responsive to the Q&T prior microstructure. This tempered martensitic structure is best when the percentage of spheroidal carbon is small. Unfortunately, the Q&T microstructure is not as easy to machine as a normalized or annealed structure, and this can be seen schematically in Figure 13. To alleviate this problem, experiments were conducted for normalizing samples by varying certain microstructural elements. The influence of ferrite and spheroidal carbon in F+Cementite were tested on static mechanical properties and fatigue strength of contour hardened gears.

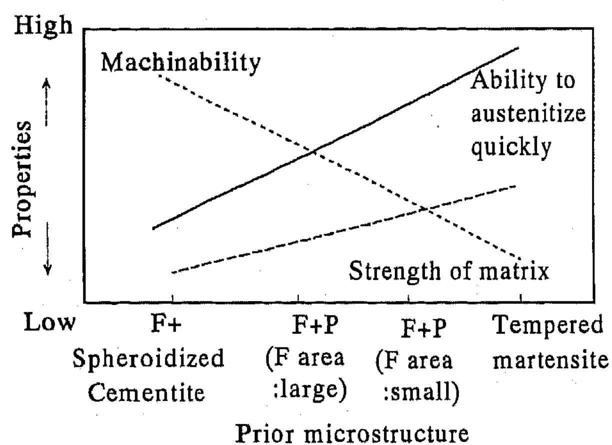


Figure 13 - Schematic Figure of Prior Microstructure versus Machinability, Ability to Austenitize, and Strength of Matrix

The material used to make the gears and conduct the various prior microstructure tests was SAE1045. The goal of varying the microstructure was to produce a uniform martensitic case after contour hardening. Figure 14 shows the various microstructures tested with concentration focused on the amount of free ferrite.

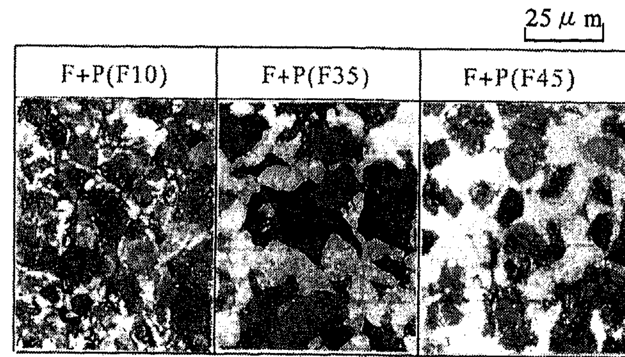


Figure 14 - Varied Microstructure of SAE1045

Results from these tests show that when the ferrite area is large, the temperature to achieve austenitization is much higher and the time is also extended. Large areas of ferrite inhibit ideal contour patterns and decrease the residual compressive stress. The distortion is also increased due to the higher temperatures and longer heat times needed to completely austenitize the sample. Figure 15 shows the lower line limit in the TTA diagram for the various prior microstructures. As the percentage of ferrite and spheroidal cementite increases, higher temperatures and longer heating times are needed to austenitize homogeneously.

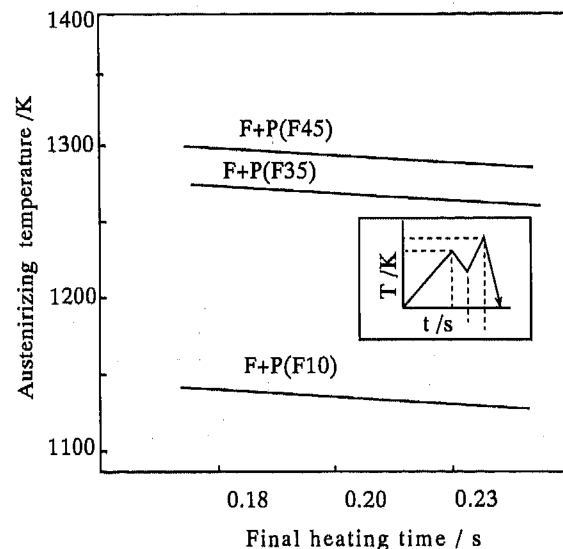


Figure 15 TTA Diagram versus Varied SAE 1045 Microstructure

Testing was performed on the various microstructures and the bending fatigue results can be seen in Figure 16.

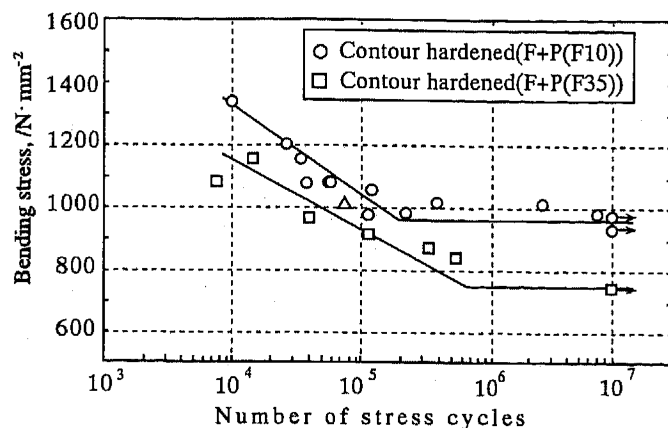


Figure 16 - Bending Fatigue of Contour Hardened Gears

The SAE 1045 microstructures had very different results with respect to hardness, stress and distortional characteristics. The three main samples that were Micropulse contour induction hardened can be seen in Figure 17.

Steel Development for rapid induction heating and quenching - The steel that is easy to austenitize is not easy to machine, and the steel that is easy to machine does not readily transform to homogenous austenite during rapid heating. At the present, the machinability issue is sacrificed somewhat to produce the best component possible. Several steel makers are currently working on developing a steel that can meet both requirements of quick austenitization and machinability.

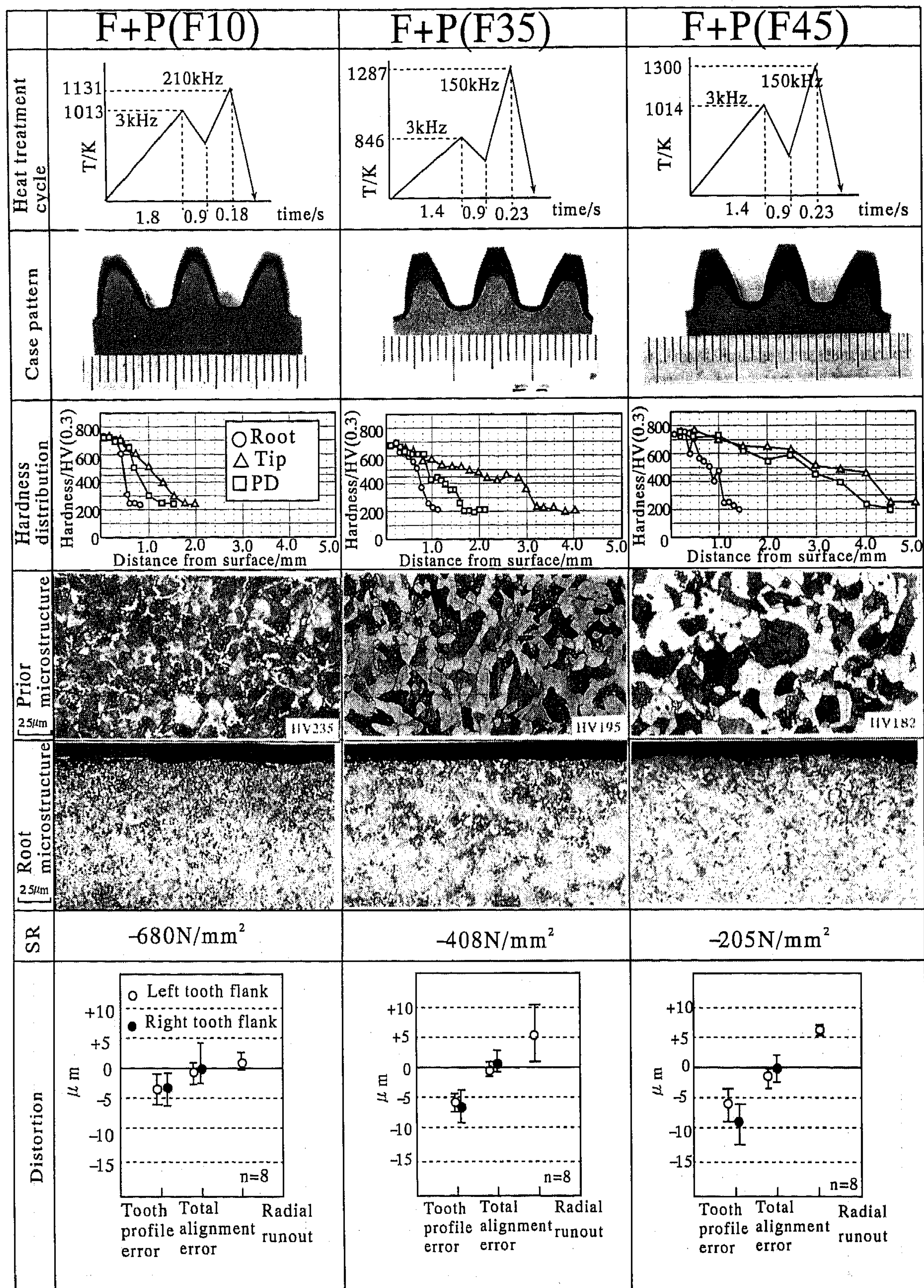
FUTURE RESEARCH

Demand for automotive small gearing - In the automotive industry there is a great need for gearing that is accurate, strong, low friction, lightweight, quiet and fuel efficient. These types of gears not only demand fatigue strength of tooth root, but also need good impact and contact fatigue properties. These properties are outlined in Figure 18. Carburizing has a long history and an abundance of data on the effect of material composition, hardened pattern, microstructure and compressive stress to name a few. The future goal is to establish a database for rapid induction hardening that can compare to the historical expertise of carburizing.

Technical development of process unit - One of the many goals in the future of rapid induction hardening is to improve productivity and develop a production technique that meets the users requirements. To better accommodate larger gears, a high output generator of the proper frequency must be designed.
With the higher power generators, large gears can be heated very quickly and achieve the same results as smaller gearing.

	Bending fatigue	Higher stress fatigue (impact)	Contact fatigue
Reasons for failures	Root radius receives bending stress in maximum torque repeatedly and fails. Stress Breakage	Root radius receives impact stress and fails. Stress Breakage	Tooth receieves rolling and sliding contact stress in mesh therefore pitting and spalling. Pitting Spalling
Solution to failures	1.Higher hardness of surface 2.Higher residual compressive stress 3.Optimum case depth 4.Higher core hardness	1.Shallower case depth (impact resistance ductility) 2.Optimum core hardness 3.Higher residual compressive stress	1.Higher surface hardness 2.Sufficient case depth for Heltz stress 3.Optimum amount of retained austenite
Schematic figure	1,2 Material strength Loading stress of root radius Stress 3 4 Distance from surface	Impact fatigue Bending fatigue Stress 10^3 N	1,3,4 Material strength Shear stress of tooth Shear stress 2 Distance from surface

Figure 18 - Desired Fatigue Properties by Automotive Gear Makers



SR: Root radius residual stress tangent to surface

Figure 17 - Comparison of S45C (SAE1045) Microstructures and Properties

CONCLUSION

Today's many technological advances in the induction power supply industry, combined with the Micropulse A.P.C. technology, now allows accurate contour induction hardening of automotive gears. This process has shifted gear hardening from carburizing and nitriding processes to a contour induction hardening process that easily can be adapted to in-line processing. Gears that are contour induction hardened with the Micropulse process have higher overall strength than gears processed by carburizing, conventional induction hardening or soft-nitriding. These results are the findings of initial testing and are not generalized of all gearing. The main focus of this study was small automotive type gearing.

To obtain the best results with contour induction hardening, a Q&T microstructure with a small percentage of spheroidal carbon is ideal. The Q&T microstructure is rather hard and is difficult to machine. In order to facilitate good machining characteristics, several microstructures were tested for quick austenitization and machinability. The main focus in this study was on the amount of ferrite in the pearlitic matrix. The best results of the normalized samples contained the least amount of ferrite in the core microstructure. The microstructure that contained ten percent free ferrite had good machinability and physical properties. Residual stress, dimensional stability, and case depth degraded as the amount of ferrite was increased in the core microstructure.

This study resulted in important discoveries about such things as: strength compared to other processes, machinability, prior microstructure, and optimizing the current process of Micropulse induction hardening. Many factors influence the outcome of contour induction hardening, but with a good database from research and development, the scope of this process is unlimited.

REFERENCES

1. John M. Storm, Michael R. Chaplin; Heat Treat 19(6) (1987)
2. Y. Misaka, Y. Kiyosawa, K. Kawasaki; 35th Conference of the JPN. SOC. Heat Treat, 17-18 (1992)
3. Y. Misaka, Y. Kiyosawa, K. Kawasaki; JPN. SOC. Mech. Eng. 187-188 (1995)
4. T. Horikawa, Y. Misaka, K. Kawasaki; 44th Conference of JPN. SOC. Material Science, 103-104 (1995)
5. Wilson O. Silverthorne II; First International Induction Hardening Conference Proceedings, pp 171-178 (1995)
6. Komodori, Shimizu, Y. Misaka, K Kawasaki; 37th Congerence of JPN. SOC. Mech. No. 2223 (1996)
7. G.D. Pfaffmann; "Heat & Surface", 429 (1992)
8. M. Umemoto, Y. Misaka, K. Kawasaki; 132nd Conference of Iron and Steel Institute of Japan, 1462 (1996)

970972

Effect of Normalized Microstructure in Alloy Steel on the Performance of Planetary Gear Set of Automatic Transmission

Seung Cheal Jung, Seung Gyun Ahn, Hak Jin Kim, and Young Woo Kim
Hyundai Motor Co.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The banded microstructure of pearlite and ferrite in normalized alloy steel is susceptible to thermal distortion during carburizing process due to its unidirectional orientation parallel to rolling direction. The planetary gears with material of banded microstructure have been experienced in high thermal distortion during carburizing and quenching process and result in uneven surface hardness and effective case depth at the inside of pinion gear after honing. These defects played failure initiation site roles in durability test during development of new automatic transmission. The galling between the contacting components in severe lubricating system was the main failure mechanism. Double normalizing at 920 ℃ was designed to resolve the banded microstructure of normalized alloy steel. The microstructure and grain size of the double heated steel became equiaxed and fine due to homogenizing and recrystallization through double heat treatment. The thermal distortion and subsequent surface properties at the inside of pinion gear were also improved. Finally the durability of the planetary gear set at maximum speed test is enhanced by 2.91 times.

INTRODUCTION

As the high performance engines have been developed, the driving conditions of automatic transmission have become more severe. The lubrication and cooling system are important factors to enhance the performance of automatic transmission. The temperature of ATF oil was reached to critical point under maximum speed condition. This condition causes loss of lubricating characteristics of ATF oil and results in the failure of planetary gear set of automatic transmission. Improvement of oil flow system and dynamic analysis of moving component of automatic transmission have been studied to solve these problems (1, 2). Regardless of the importance of lubrication and cooling system, failures of moving parts in automatic transmission are frequently related with some defects of materials which consist of planetary gear set. Improper surface hardness and effective case depth, abnormal microstructure, thermal distortion and roughness are among these faults (3, 4). In this research, we examined failed planetary gear set of developing automatic transmission during high speed test, revealed the failure route and mechanism. The thermal distortion of pinion gear during carburizing process increased the ovality and cylindricity of inside of pinion and resulted in uneven surface hardness, effective case depth and microstructure after honing. These defects were major sources of failure in planetary gear set.

The segregation of chemical composition during steel melting, soaking, forging and rolling process could cause the banded microstructure of alloy steel (5). Because of predominant directionality, the banded microstructure showed high thermal distortion during carburizing heat treatment. To improve the banded microstructure, double normalizing at increased temperature from 880 ℃ to 920 ℃ was introduced and well equiaxed and fine microstructures were obtained. The double normalized pinion gear showed less thermal distortion after carburizing than the single normalized pinion. With improved microstructure of alloy steels which used in pinion gear, the performance of planetary gear set could be enhanced without any supplementary attachment of cooling system for gear set.

EXPERIMENTAL PROCEDURE

PLANETARY GEAR SET OF AUTOMATIC TRANSMISSION - Figure 1 shows the planetary gear system of automatic transmission which was investigated in this research. Part ②, ③, and ④ are pinion shaft, long pinion and needle roller bearing respectively, which are major components in contacting system. The material specifications for the components are shown in Table 1. The above three parts are subjected to severe driving atmosphere at maximum speed condition. When the lubrication and subsequent cooling system are not proper, the contacting parts are heated, weared and finally seized as shown in Figure 2. Improvements of lubrication routes have been made to increase oil supply as shown in Figure 3. Supplementary oil cooler was attached to decrease temperature of ATF from 125 ℃ to 110℃

MICROSTRUCTURE IMPROVEMENT OF PINION GEAR MATERIAL - The other improvement for the planetary gear set was made in microstructure modification of pinion gear by double normalizing before carburizing heat treatment. The pinion gear is directly machined from smooth turned as-rolled bar with normalizing heat treatment as shown in Figure 4. The single normalized microstructure of pinion gear showed very predominant orientation parallel to rolling direction with banded microstructure, pearlite and ferrite. Figure 5 represents the process of improved double normalizing heat treatment to obtain equiaxed microstructure of pearlite and ferrite.

BIAS MEASUREMENT OF PINION GEAR - The distortion of tooth profile can be determined by measuring profile and lead bias which mean the difference between maximum and minimum of pressure angle error and helix angle error respectively as shown in Figure 6. Two kinds of pinion gears with different microstructures, one directional banded and the other equiaxed, were

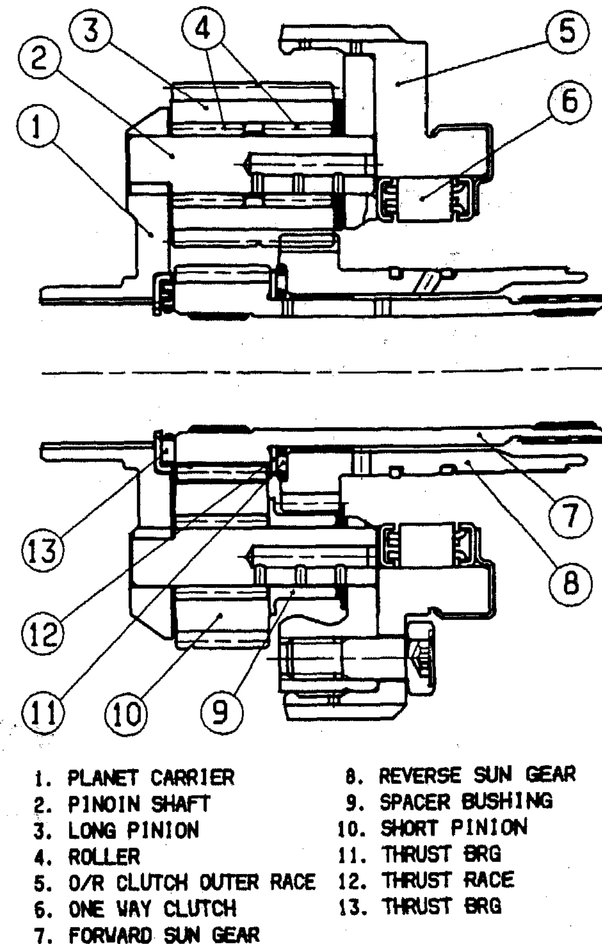
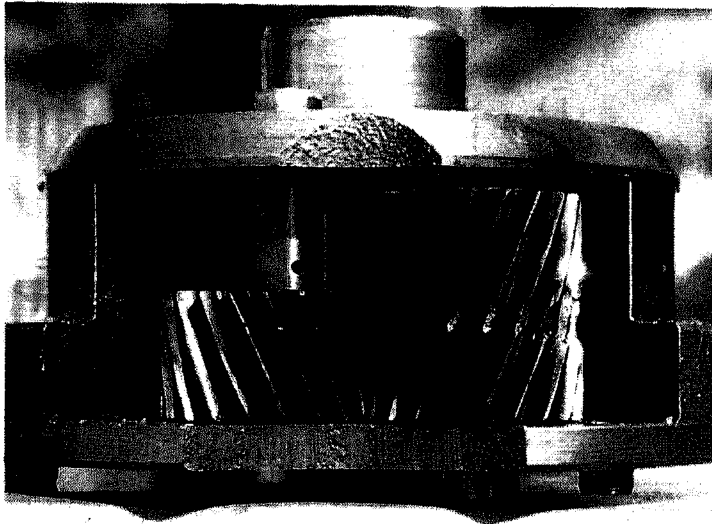


Figure 1. Planetary gear set of automatic transmission

TABLE 1 - Chemical composition and Heat Treatment of Components

Component	Materials	Chemical Composition (Wt%)										Heat Treatment
		C	Si	Mn	P	S	Cr	Mo	Ni	Cu	Fe	
Pinion Shaft	SUJ 2	1.05	0.20	0.30	0.017	0.009	1.50	0.007	0.045	0.035	BAL.	Carbonitriding
Roller Bearing	SUJ 2	1.10	0.16	0.29	0.016	0.006	1.45	0.006	0.047	0.039	BAL.	Carbonitriding
Long Pinion	SNCM420H -ESR	0.23	0.15	0.67	0.016	0.001	0.52	0.21	1.80	0.077	BAL.	Carbonitriding



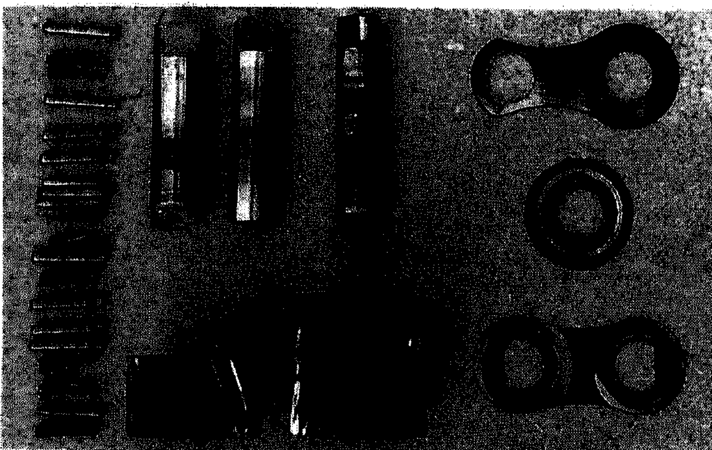


Figure 2. Failed planetary gear set of automatic transmission

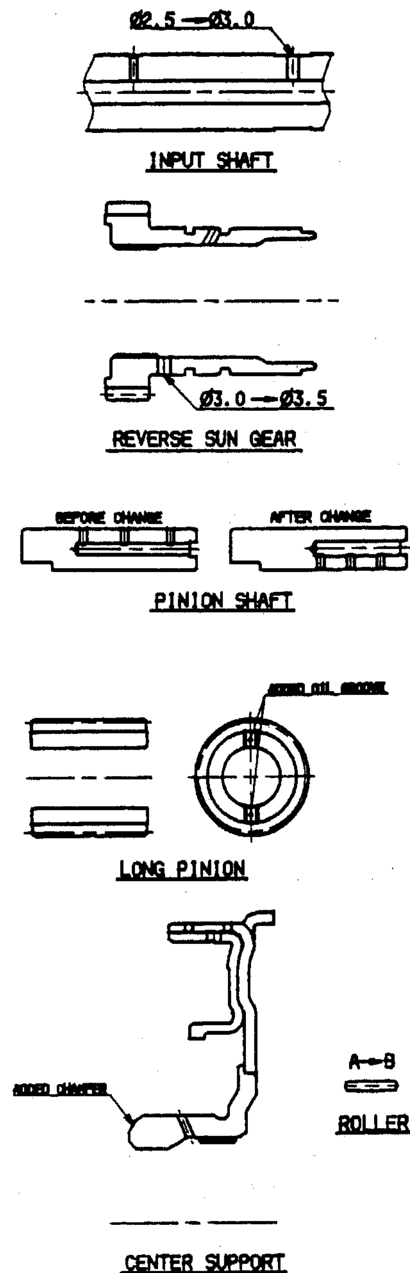


Figure 3. Improvement of lubrication root of planetary gear set of automatic transmission

prepared. After carburizing and honing the inside of pinion gear, profile and lead bias were measured to estimate the influence of microstructure on thermal distortion during the carburizing process.

THERMAL DISTORTION - To investigate the effect of microstructure on thermal distortion during carburizing process, navy-c type and cylindrical type specimens which are shown in Figure 7 and Figure 8 were machined from single and double normalized samples respectively.

The dimensional changes of c-gap, thickness, internal and external diameters after carburizing in navy-c type specimens were measured. Whereas, in cylindrical samples, the dimensional changes of internal and external radius at every 36 degree along the radial direction at hight A, B and C of Figure 8 were also measured.

METALLOGRAPHY AND HARDNESS TESTING - Metallographic samples were cut from failed gear, single normalized and double normalized new gears. The surface microstructure of failed pinion gear carefully examined to identify the failure mechanism and surface hardness distributions especially inside region were examined. The criterion of surface hardness is above 650 HV and that of effective case depth is the distance from surface to the subsurface which shows 550 HV hardness. The failed and honed surface were examined by scanning electron microscopy (SEM). The microstructure of single and double normalized samples were also examined by optical microscope. The prior austenite grain sizes were determined in accordance with ASTM E112-88, after etching in a saturated picric acid and hydrochloric acid solution in which sodium tridecylbenzene was used as a wetting agent (6).

DYNAMOMETER TEST OF PLANETARY GEAR SET - Dynamometer tests of automatic transmission gear sets were conducted to confirm the effect of improved microstructure of pinion gear on the performance of automatic transmission at maximum speed condition.

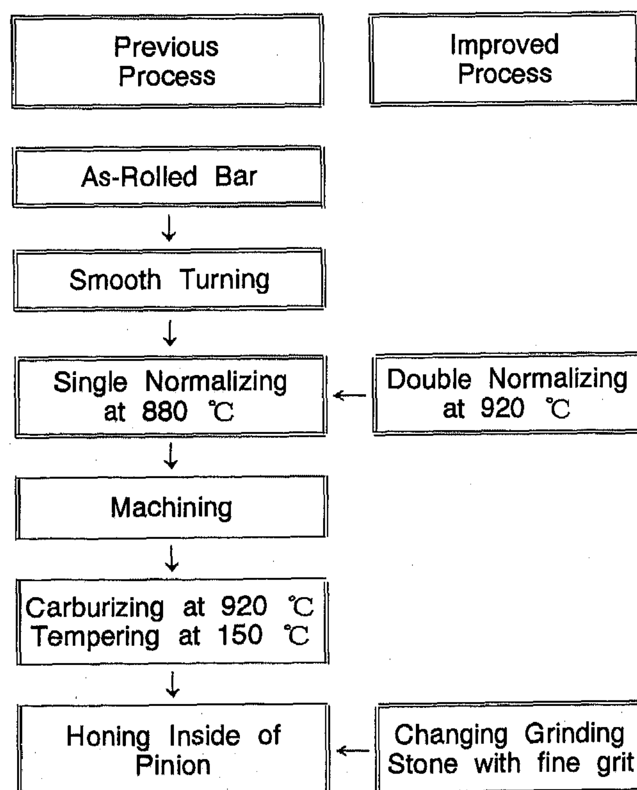


Figure 4. Comparison of previous and improved processes for pinion gear preparation

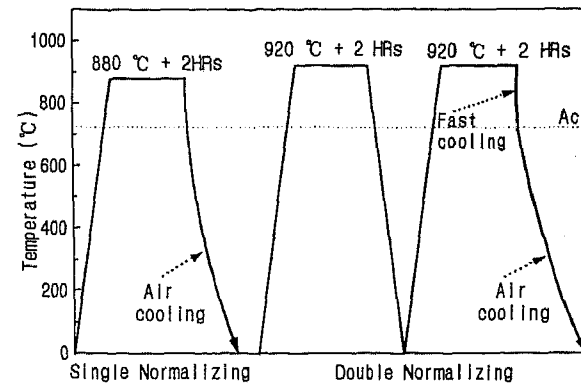


Figure 5. Comparison of single and double normalizing conditions.

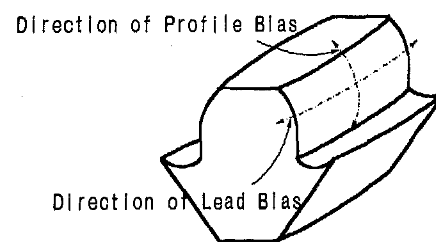


Figure 6. Comparison of tooth profiles for direction of profile bias and direction of lead bias

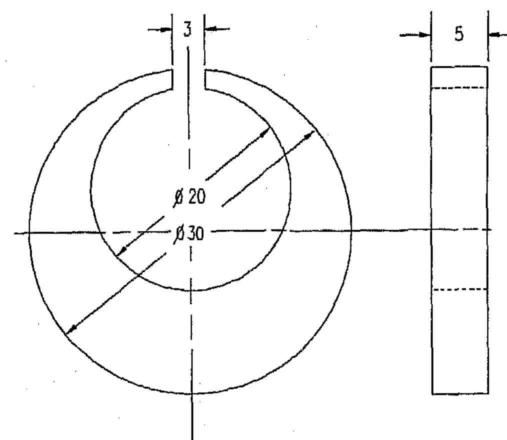


Figure 7. Navy-C type thermal distortion specimen. All dimensions in millimeters.

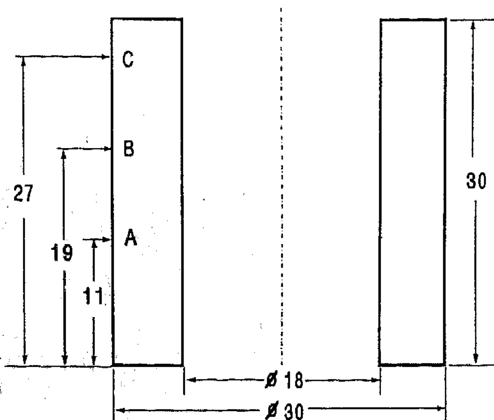
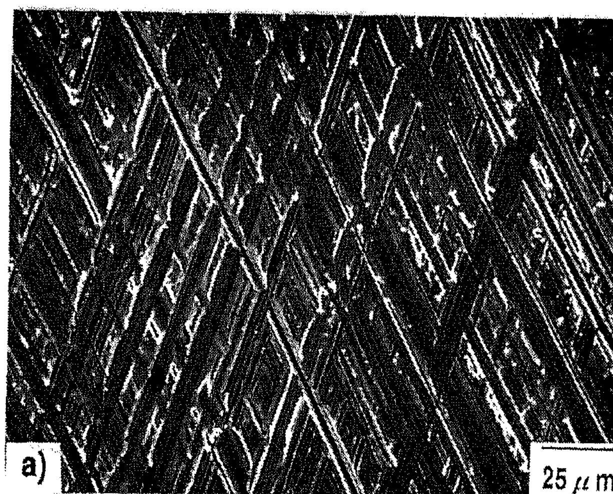
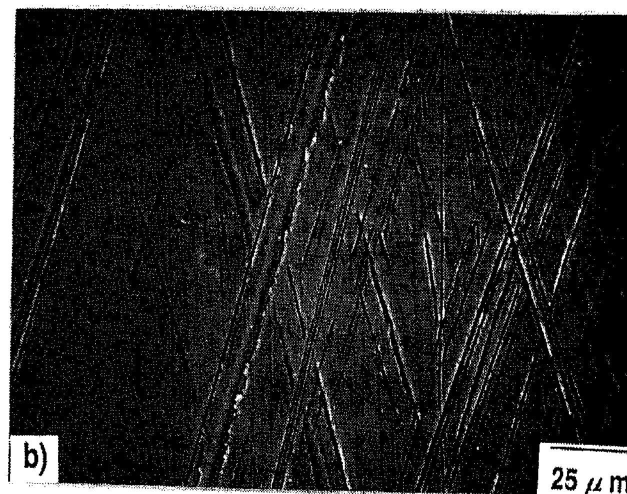


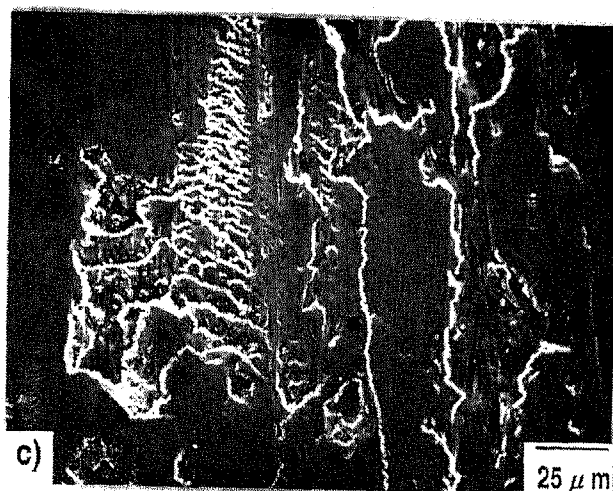
Figure 8. Cylindrical type thermal distortion specimen. All dimensions in millimeters.

RESULTS AND DISCUSSION

FAILURE MECHANISM OF PLANETARY GEAR SET - Failed parts of planetary gear set were examined to explain the failure mechanism of automatic transmission. The failure originated from the interface between inside of long pinion gear and needle roller bearing during dynamometer test under maximum speed condition. SEM analysis for failed parts showed the failure process as following galling failure ; Severe Wear → Adhesive wear → Spalling → Seizure. General causes of galling are insufficient galling resistance of material due to low hardness, ease of welding, lack of protective layer, severe sliding in contact, severe normal load, overheating, unfavorable surface microgeometry ; roughness and inadequate boundary lubrication (7, 8). Figure 9 shows the failure process from (a) to (d) for the inside of failed pinion gear. At first stage of failure, sharp feather edge at inside of pinion







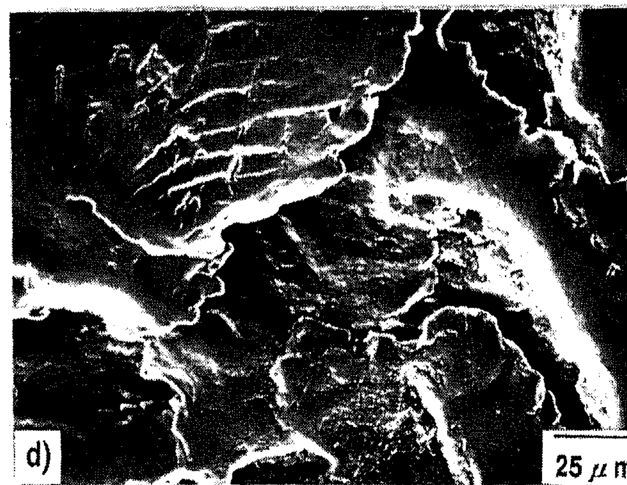


Figure 9. SEM analysis of galling failure process (a) as-honed condition (b) Severe wear (c) Adhesive wear (d) Macrospalling and Seizure.

gear after honing weared and disappeared by asperity contact with needle roller bearing as shown in Figure 9 (a) and (b). After this stage, prow type wear resulted from metal to metal adhesive wear was observed in Figure 9 (c). The prow rings were experienced additional wear and work hardened and finally resulted in spalling as Figure 9 (d). This failure process drove the planetary gear set to seizure with contacting components, needle roller bearing and pinion shaft. Surface microstructure of pinion gear inside area was observed by optical microscope as shown in Figure 10. At the early stage of failure, galling phenomena were accompanied with decarburizing as shown in Figure 10 (a) and adhesive welding wears were distributed over the entire range along the axis direction of pinion gear. Spallings were occurred after asperity contacting abrasive wear as shown in Figure 10 (b) and lubrication layer was destroyed and resulted in seizure failure.

Among the factors influencing galling failure, low surface hardness and effective case depth, coarse surface roughness, loss of oil characteristics due to excessive high oil temperature were main sources of this automatic transmission failure.

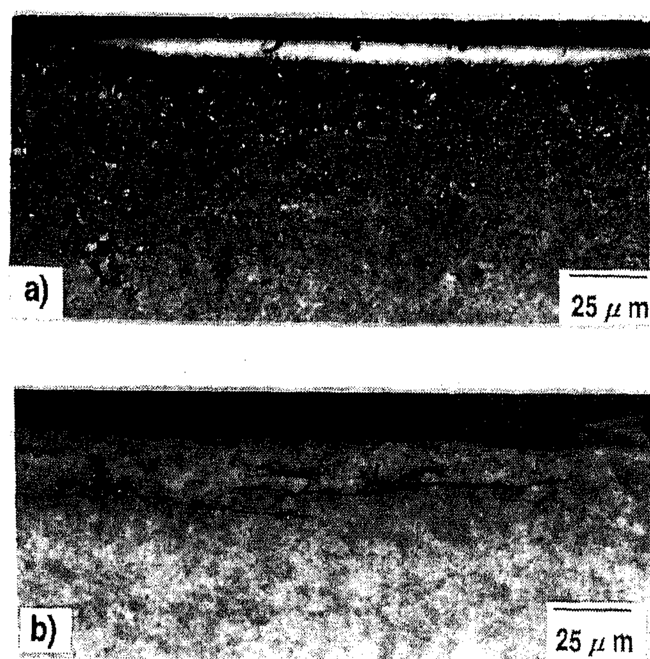


Figure 10. The microstructures of (a) surface galling and (b) spalling which are shown as white decarburized layer and subsurface microcracks respectively.

MICROSTRUCTURE IMPROVEMENT OF PINION GEAR MATERIAL - Microstructure of single normalized pinion gear material consisted of banded structure with pearlite and ferrite. This microstructure has strong directionality so that high thermal distortion during carburizing was expected. Through investigating failed pinion gear, it was confirmed that some surface area of pinion gear showed very low surface hardness and effective case depth in Figure 11 and Figure 12. The surface hardness was ranged from 550 HV to 780 HV and effective case depth was also not uniform from 0.20 mm to 0.40 mm. These inadequate mechanical properties were driven from non-uniform grinded depth for inside of pinion gear after carburizing. Because of severe ovality and cylindricity of pinion inside after carburizing, uneven grinding could be inevitable. This oval type thermal distortion is directly related with banded microstructure of smooth turned pinion gear material.

Banded microstructure comes from the segregation of chemical composition during steel melting, soaking, forging and rolling process and is basically influenced from chemical composition of alloy steel. From the results of research by Fujita, as the content of Cr, Ni, and Mn increase and those of Si and Mo decrease, the banded microstructure becomes more predominant (5). Considering the chemical composition of pinion material ; SNCM420H-ESR, it was difficult to avoid the banded structure during current steel-bar making process without modifying chemical composition. Modification of normalizing process was introduced to find alternative method for modification of alloy chemical composition. Basic concept of modified heat treatment was based on homogenization at austenite and rapid cooling through two phase, austenite and ferrite, region during normalizing. Enough thermal energy at austenite phase region was supplied to resolve chemical segregation by double normalizing with 2 hours at 920 ℃. Rapid cooling through austenite

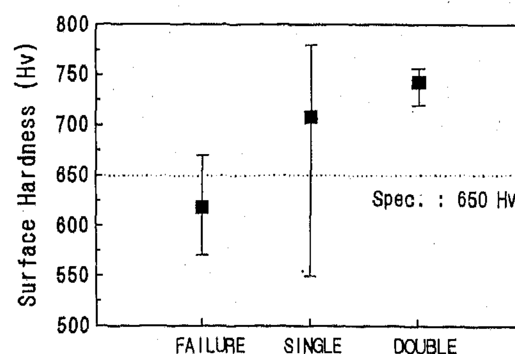


Figure 11. Distribution of surface hardness at inside of pinion gear. The marks of failure, single and double mean failed pinion during test, single normalized new pinion and double normalized new pinion respectively.

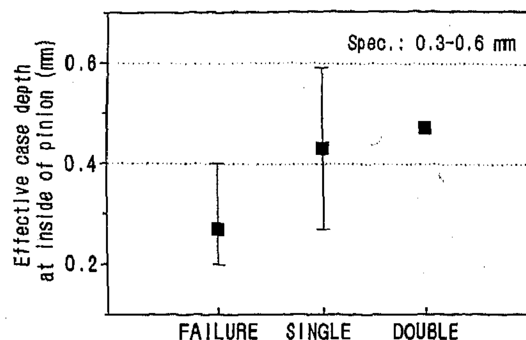


Figure 12. Distribution of effective case depth at inside of pinion gear. The marks of failure, single and double mean failed pinion during test, single normalized new pinion and double normalized new pinion respectively.

and ferrite two phase region after homogenizing was designed to avoid chemical segregation due to directional precipitation and growth of proeutectoid ferrite during cooling process through this region. Normalizing temperature was also increased from 880 ℃ to 920 ℃ to coincide with the temperature of carburizing to make stabilizing the microstructure during subsequent carburizing at 920℃.

Double normalizing was designed to transform the banded microstructure to equiaxed microstructure with pearlite and ferrite by diffusion process. By double normalizing at elevated temperature, the banded microstructure was transformed to equiaxed as shown in Figure 13. The surface hardness and effective case depth of double normalized pinion gear showed excellent uniformity with surface hardness from 720 HV to 757 HV and effective case depth from 0.45 mm to 0.48 mm respectively. The prior austenite grain size became more fine after double normalizing because of recrystallization process during second heating. The numbers of grain sizes were ASTM number 9.20 and ASTM number 11.40 for single and double normalized samples before carburizing respectively. The tendency of the difference had been maintained during carburizing. As shown in Figure 14 (a) and Figure 14 (b), the average number of grain size became ASTM number 8.68 and ASTM number 10.60 for single and double normalized samples after carburizing respectively. Figure 14 (c) shows the abnormally coarse grain of single normalized sample after carburizing. This kind of abnormal grain growth was randomly distributed and deteriorate thermal distortion and strength of pinion gear. The bending fatigue performance of carburized alloy steels can follow the Hall-Petch type relationship and carburized steels with fine austenite grain size exhibit superior bending fatigue performance. From previous research results as shown in Figure 15, the bending fatigue strength of pinion gear can be improved approximately 10 % by double heating (9, 10). The relationship between contact fatigue strength and grain size has not been studied quantitatively enough.

Tsushima reported that prior austenite grain size did not seem to influence the rolling contact fatigue life (11). But, recent studies have shown that the increase of contact strength of case hardened steel could be achieved by grain refinement with microalloying element. In dynamometer test with fine grained pinion, surface pitting failure was also improved so that even though quantitative correlation has not been proposed, the Hall-Petch type relationship between grain size and contact fatigue strength can be also estimated. This improvement of thermal distortion, bending and contact fatigue strength of double normalized pinion gear could contribute to increase the performance of planetary gear set of automatic transmission.

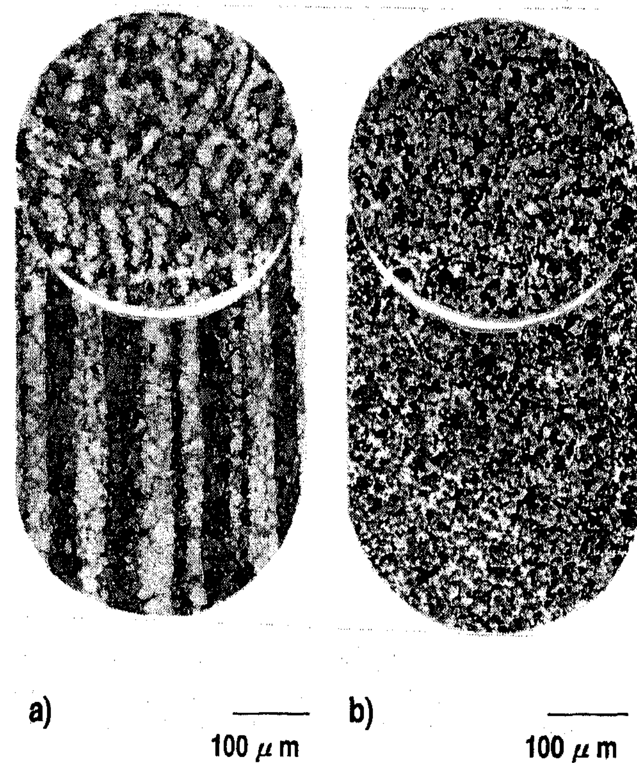


Figure 13. Microstructure of single and double normalized pinion gear materials (a) single normalized and (b) double normalized

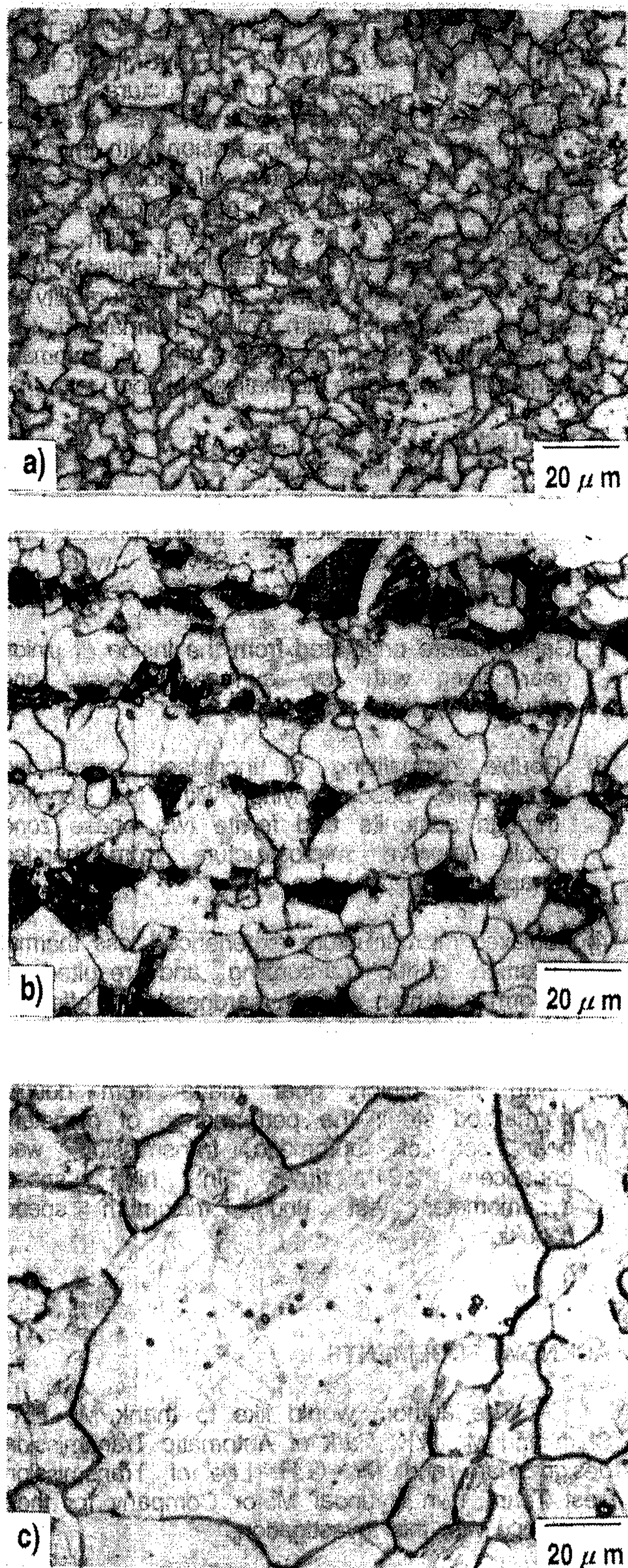


Figure 14. Prior austenite grain sizes after carburizing. (a) single normalized pinion, (b) double normalized pinion and (c) Abnormal grain growth observed at single normalized pinion after carburizing.

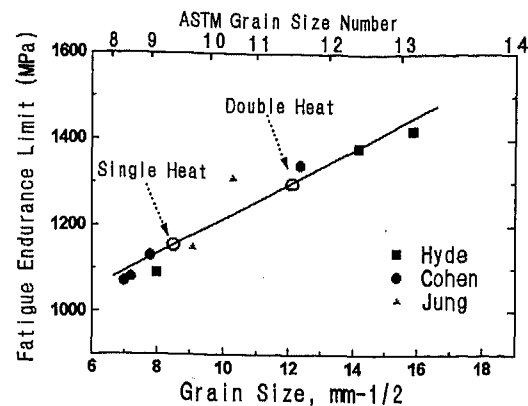


Figure 15. Bending fatigue endurance limit versus prior austenite grain size

BIAS MEASUREMENT OF PINION GEAR - The results of profile and lead bias measurement for carburized pinion gears made from single and double normalized steels were shown in Figure 16. The direction of profile bias is from root to tip of gear and perpendicular to rolling direction of material and the direction of lead bias is from bottom to top of gear and parallel to rolling direction of bar steel.

The single normalized pinion gear showed very wide deviation for profile bias and lead bias ranged from 3 μm to 30 μm and from 4 μm to 73 μm respectively as shown in Figure 16. But, double normalized pinion gears represented well stabilized bias ranged from 2.5 μm to 19.30 μm for profile direction and 8.00 μm to 29.40 μm for lead orientation. The average distribution of profile and lead bias were decreased 27 % and 33 % respectively by double heating.

The microstructure of pinion gear in parallel to rolling direction exhibited strong directionality than that of perpendicular direction in single normalized material. This means that lead bias is more susceptible to thermal distortion and the result of measurement are also coincide with this expect.

THERMAL DISTORTION - Figure 17 shows the dimensional changes of navy-c type specimen from single and double normalized samples after carburizing. The 87.8 % and 69.3% Improvements of thermal distortions were made on external and on internal diameter respectively. The improvements of thermal distortions on c-gap and thickness were also 29.3 % and 12.0 % respectively. Distortion measurement of cylindrical specimens were made on the points A, B and C of Figure 8. Cylindricity of internal and external diameter on single normalized cylinders were 8.3 μm and 23.5 μm respectively. After double normalizing, cylindricities of internal and external diameter were improved

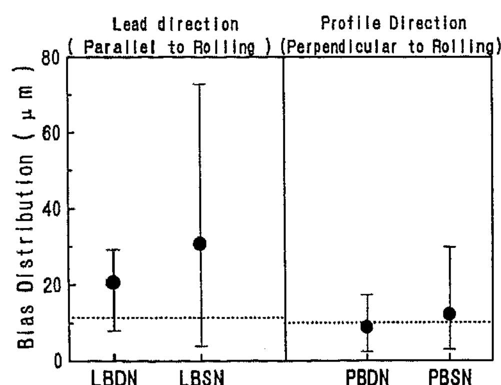


Figure 16. Comparison of bias distributions for pinion teeth which made from single and double normalized steels. LBDN and LBSN represent lead bias of double normalized and simple normalized gear respectively. PBDN and PBSN also represent profile bias of double normalized and simple normalized gear respectively

17.5 %, from 8.3 μm to 6.8 μm and 19.1 %, from 23.5 μm to 19.5 μm , respectively. Dimensional changes of internal and external diameters for single and double normalized cylinders were shown in Figure 18 and Figure 19. Single normalized samples showed higher contractive dimensional change than double normalized samples after carburizing in both internal and external diameter. The large contraction required more grinding depth after carburizing and results in localized improper low surface hardness and effective case depth at the inside of pinion gear. These localized improper surface properties of pinion inside could be the weakest points between the contacting pinion shaft and roller bearing and original sites of failure during durability test.

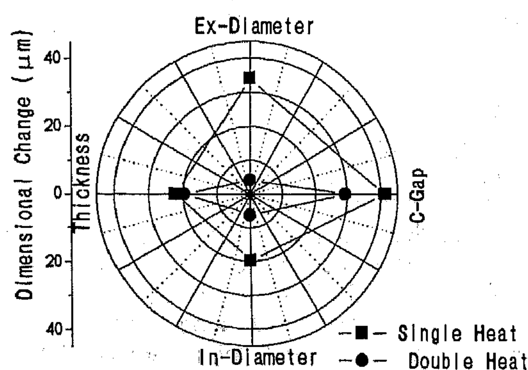


Figure 17. Dimensional change in navy-c type specimen after carburizing on single and double normalized samples

DYNAMOMETER TEST FOR PLANETARY GEAR SET OF AUTOMATIC TRANSMISSION - The effect of improved microstructure on the performance of planetary gear set was evaluated for assembled automatic transmission with improved pinion gear. The supplementary oil cooler in cooling system of transmission was not supplied in this dynamometer test. The results of high speed dynamometer tests for automatic transmission gear sets were shown in Figure 20. The durability of automatic transmission with double normalized gear was enhanced 2.91 times than that of automatic transmission with single normalized pinion.

CONCLUSION

1) Failure mechanism in planetary gear set of automatic transmission was galling process with following procedure ; Severe wear → Adhesive wear → Macrospalling → Seizure.

2) Galling failure originated from the inside of pinion gear along with low surface hardness and improper effective case depth.

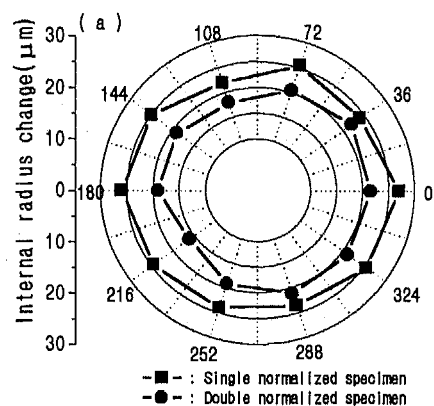
3) Double normalizing at increased normalizing temperature accompanying with fast cooling through austenite and ferrite two phase zone could improve microstructure from banded shape to equiaxed one.

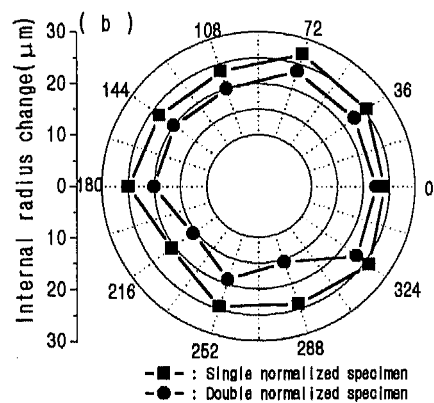
4) Equiaxed microstructure experienced less thermal distortion during carburizing and resulted in uniform and high surface hardness and effective case depth at the inside of pinion after grinding.

5) With the pinion gear made from double normalized steel, the performance of planetary gear set of automatic transmission was enhanced 2.91 times in high speed dynamometer test under maximum speed condition.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. H.T. Shim and Mr. R.K. Park of Automatic Transmission design team and Mr. G.H. Lee of Transmission Test Team 1 in Hyundai Motor Company for their assistance with this investigation.





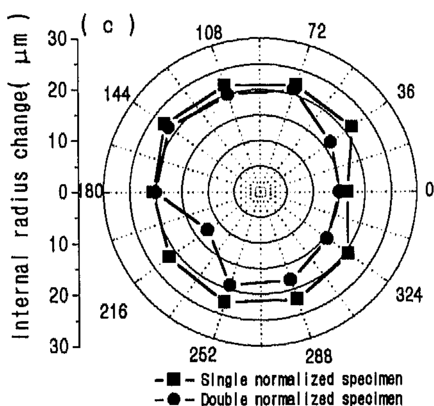
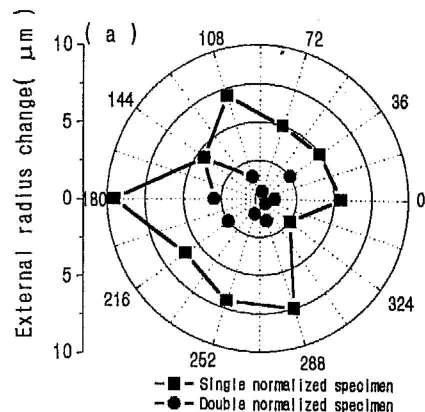
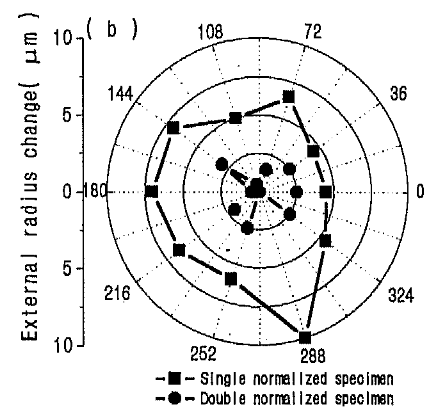


Figure 18. Dimensional changes of internal radiuses after carburizing for single and double normalized cylinders. Figure (a), (b) and (c) represent changes of internal radiuses at every 36 degree along radial direction at hight A, B and C of Figure 8.





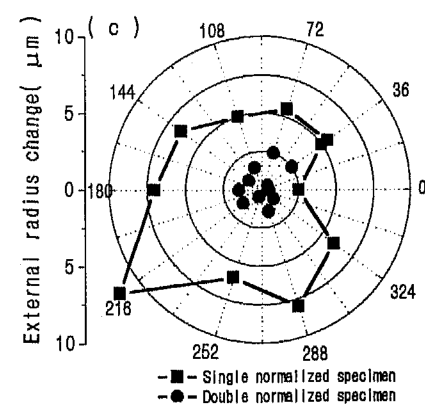


Figure 19. Dimensional changes of external radiuses after carburizing for single and double normalized cylinders. Figure (a), (b) and (c) represent changes of external radiuses at every 36 degree along radial direction at hight A, B and C of Figure 8.

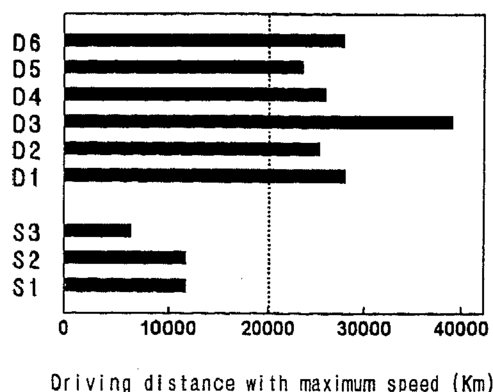


Figure 20. Results of high speed dynamometer durability test for planetary gear set of automatic transmission. S1, S2 and S3 are the results of test with single normalized pinion, whereas D1, D2, D3, D4, D5 and D6 represent the results of test with double normalized pinion.

REFERENCES

1. M.Kojima, K.Hukumura : A Study on the lubricating Oil Flow in the Automatic Transmission " SAE Report No.910801

2. Hiroshi Ikeda : Influence of Lubricating on Pitting of Gear - In respect of oil film thickness - Isuzu Technical Report Vol. 93 pp 108-113

3. G. Parrish and G.S. Harper : Production Gas Carburizing, Pergamon Press, Oxford, England, 1985

4. G. Krauss: " Microstructure and Properties of Carburized Steels " in Metal Hand Book, Vol.4 : Heat Treating, ASM International, Materials Park, Ohio, 1991, pp.363-375

5. Shyohei Fujita and Takaharu Morlyn : " On the Banded Structure in Low Alloy Steels " Nissin Steel Technical Report 40, 1979, Jun. pp. 1-14

6. A.W.Brewer, K.A.Erven, G.Krauss : " Etching and Image Analysis of Prior Austenite Grain Boundaries in Hardened Steels " Metal characterization 27, 1991, pp 53-56

7. T.E.Tallian : " Failure Atlas for Hertz Contact Machine Element " ASME Press, 1992, pp.121-140

8. E. Schedin and B. Lehtinen : " Galling Mechanism in Lubricated Systems ; A Study of Sheet Metal Forming " Wear 170, 1993, pp.119-130

9. S.C.Jung, D.J.Medlin, and G.Krauss " Effects of Subzero Treatments on the Bending Fatigue Performance of Carburized SAE-4320 and SAE -9310 Steels " SAE Report No.960313

10. R.S. Hyde, G. Krauss and, D.K. Matlock : "The Effect of Reheat Treatments on Fatigue and Fracture of Carburized Steels" SAE Report No. 940788

11. N. Tsushima and, I. Sugiura : " Influence of Various Material Factors of One Percent Chromium Carburizing Steel upon Rolling Contact Fatigue Life" SAE Report No.820633

970975

The Influence of Fillers on Paper-Based Friction Materials Relative to Wet Clutch Slip Characteristics

Hideto Nakagawa
Aisin Chemical Co., Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Various wet friction materials, in which different fillers are compounded on the surface of fibrillated fiber base, were prepared. The friction properties of these materials were pursued at various slip speeds, loads and oil temperatures. Streibeck Diagrams of the data were obtained, and the effect of fillers on the frictional characteristics was investigated.
The following results were obtained;

(1) The slope of the Streibeck Diagram should be positive to have smooth frictional characteristics.
(2) Abrasive powders effectively improved the slip characteristics of friction materials.
(3) The effect of the abrasive fillers was further improved by using less compliant paper.

1. INTRODUCTION

In recent years, positive slipping of paper-based friction material (friction material, hereafter) has become popular for the improvement of toque transmission efficiency and shift control of automatic transmissions. Along with this trend, vibration due to the friction properties of friction material is becoming a significant problem.

Since the friction of friction material involves boundary lubrication, theoretical prediction of friction characteristics is difficult and therefore properties are often analyzed by experiments.

Both mechanical and chemical actions generate the friction force of friction material. The effect of porosity, actual contact area (1)~(3) and ATF (4)~(7) have been studied extensively.

However, research has not advanced much on the chemical action of fillers in friction material. The study of effect of filler to friction material is very important not only for the practical purposes, but also for understanding the mechanism of the friction force generated in wet clutch systems.

In this research, several friction materials with different fillers compounded only on the friction surface were prepared, and the effect of fillers on slip characteristics was examined.

2. EXPERIMENT

2.1 OUTLINE OF TEST APPARATUS

Figure 1 shows the outline of the test apparatus used in this experiment. It is composed of a motor, a test head, a pipe heater, a torsion bar and a torque meter. The slip speed can be changed by a motor from 1 to 100 rpm. The oil temperature also can be controlled by a pipe heater from 40 to 120℃. An output torque is measured by a torque meter. Actual wet clutches and separators are assembled in the test head. A method to supply ATF can be selected from shaft center method, dipping method or combination of both methods. Friction vibration during sliding can be measured using a torsion bar.

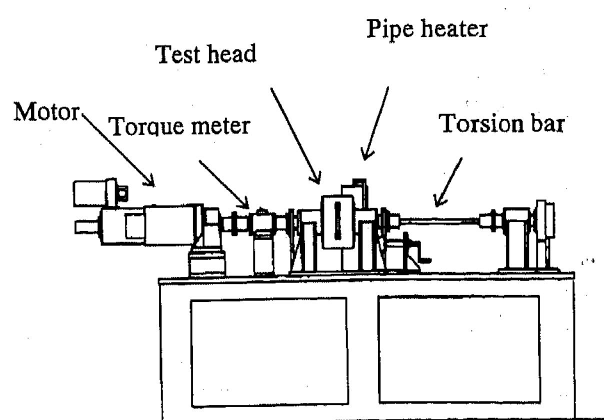


Fig.1 Test apparatus

Figure 2 shows the internal structure of the test head. Wet clutches are engaged with operating piston by increasing air pressure. The snap ring is installed in order to simulate distribution of surface pressure on the friction materials in actual vehicles. Load can be varied from 400 to 8000N.

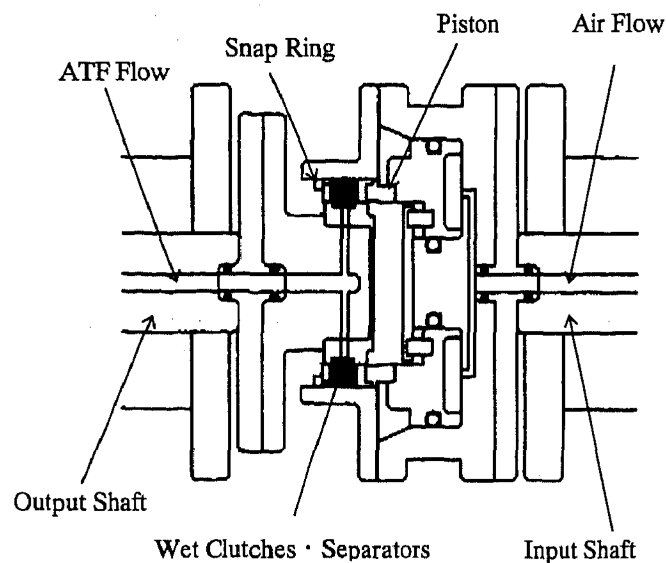


Fig.2 Internal structure of test head

2.2 PROTOTYPE SAMPLES AND FILLERS

Table 1 shows the compositions and properties of the test friction materials. Samples B to G and B' to G' are prepared by compounding various fillers only on the surface of sample A and A'. The amount of fillers are about 2% by weight. The friction properties can be affected by porosity, oil absorption speed and strain (8) ~ (10). Therefore, porosity and oil adsorption speeds of these samples were made equal to each other and a magnitude of strain was the only difference among A~G and A'~G'. The strain is measured using Rockwell Hardness Tester. The value indicates the amount of deformation in μ m after applying the 18.7MPa load for 15 seconds.

These are typical fillers used for friction materials. Hardness and abrasion of fillers are represented as relative compared with separator. The hardness of separator is Hv250. The hardness and abrasion are measured as follows. Fillers are distributed onto a separator with mirror-like polished surface. Fillers and separator are loaded to 18.7MPa.
Abrasive powder, Carbon group powder and Carbon group fibers made the surface of separator rough. Abrasive powder has the strongest abrasion as shown in Figure 3.

Table 2 shows characteristics of fillers. Figure 4 are the SEM (Scanning Electron Microscope) photos of the surface of friction materials.

Table 1 Compositions and Properties of Friction materials

Symbol	Composition	Fillers Quantity (%)	Strain (μ m)	Porosity (%)	Oil absorption speed (sec)
A	Pulp of low fibrillation degree	0.0	120	42	28
B	A + Abrasive powder	1.8	120	40	28
C	A + Carbon group powder	2.0	113	42	25
D	A + Phenol group powder	1.7	113	39	27
E	A + Solid lubricant	2.1	115	39	30
F	A + Carbon group fibers	1.9	123	39	30
G	A + Phenol group fibers	1.8	123	39	25
A'	Pulp of high fibrillation degree	0.0	75	42	28
B'	A' + Abrasive powder	1.7	70	39	35
C'	A' + Carbon group powder	2.3	80	40	35
D'	A' + Phenol group powder	2.2	80	39	33
E'	A' + Solid lubricant	1.9	80	40	25
F'	A' + Carbon group fibers	1.7	72	40	30
G'	A' + Phenol group fibers	2.2	80	41	28

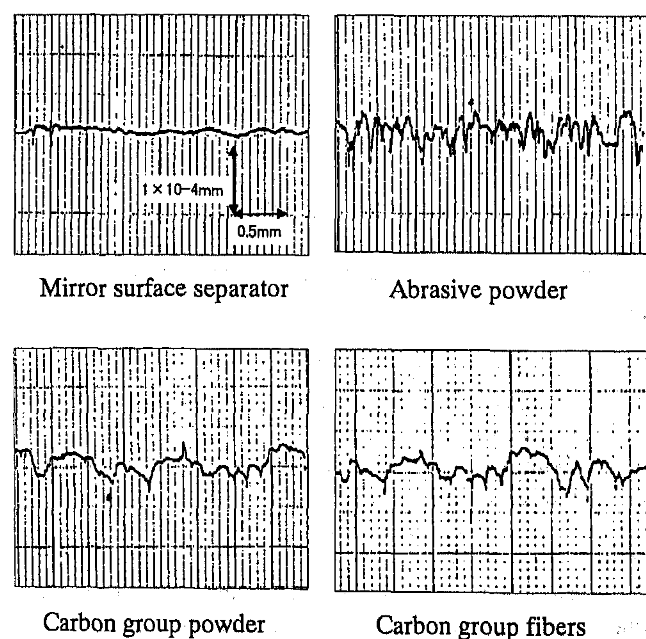


Fig.3 Surface Roughness of Separators

Table 2 Characteristics of Fillers

Filler	Temperature resistance (℃)	Hardness (Hv)	Diameter (μm)	Characteristics
Abrasive powder (B,B')	>1000	>250	6	Rough surface, More abrasive
Carbon group powder (C,C')	500	>250	15	Smooth surface, Less abrasive
Phenol group powder (D,D')	220	<250	74	Tend to form a film by friction heat
Solid lubricant (E,E')	450	<250	150	Self lubricant
Carbon group fibers (F,F')	500	>250	13	Smooth surface, Less abrasive
Phenol group fibers (G,G')	300	<250	14	Smooth surface, Less abrasive

2.3 BREAKING-IN

Because the change in actual contact area has the effect on friction characteristics, (11) the friction material was broken in, in advance, under the following conditions for stabilization of actual contact area.

Table 3 Conditions of breaking-in

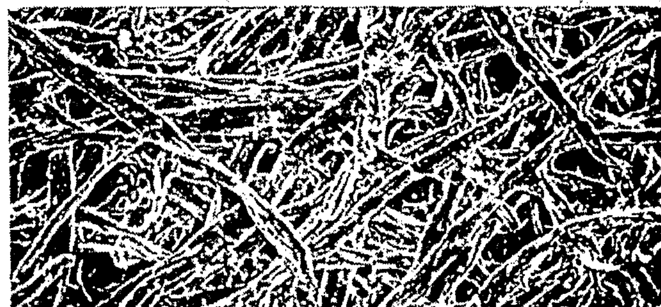
Speed	: 100rpm
Surface pressure	: 0.49MPa
Slip time	: 100sec
Oil temperature	: 100℃
Oil volume	: 1/2 dipping

2.4 TEST CONDITION

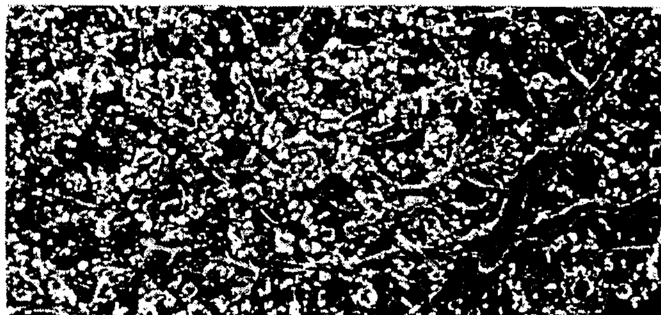
Table 4 shows the test conditions. Friction materials was slipped for 10 sec under various combinations of surface pressure , speed and oil temperature and friction coefficient and friction vibration were measured.

Table 4 Test Conditions

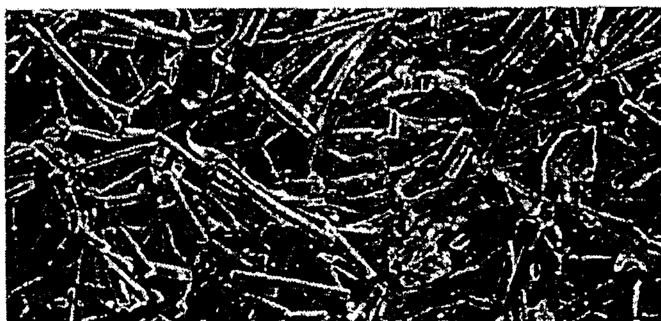
Friction material	Paper thickness: 0.55(mm) Inner/outer diameters: 105.48 × 127.3(mm)
Separator	Material : SAE1035 Inner/outer diameters: 106.35 × 126.25(mm)
ATF	Commercial ATF Kinetic Viscosity: 41.5/40℃ ,7.79/100℃ (mm^2/s)
Surface pressure(MPa)	0.49, 0.98, 1.98
Speed(rpm)	1, 5, 10, 30, 50, 70, 100
Oil Temperature (℃)	40, 100



Sample A



Sample B



Sample F

Fig.4 SEM observation of Friction Materials

Coefficient of friction is calculated by the equation below.

$$\mu = \frac{T}{n * F * \frac{2}{3} * \frac{(Ro^3-Ri^3)}{(Ro^2-Ri^2)}}$$

where T :Torque(Kgm)
n :Number of surface
F :Load(N)
Ro :Radius of outer lining
Ri :Radius of inner lining

Figure 5 shows examples of torque traces. Left side of the torque traces is an example when no vibration is observed. Right side of the torque traces is an example when vibration is observed. The average torque is used for the calculation of μ at vibration.

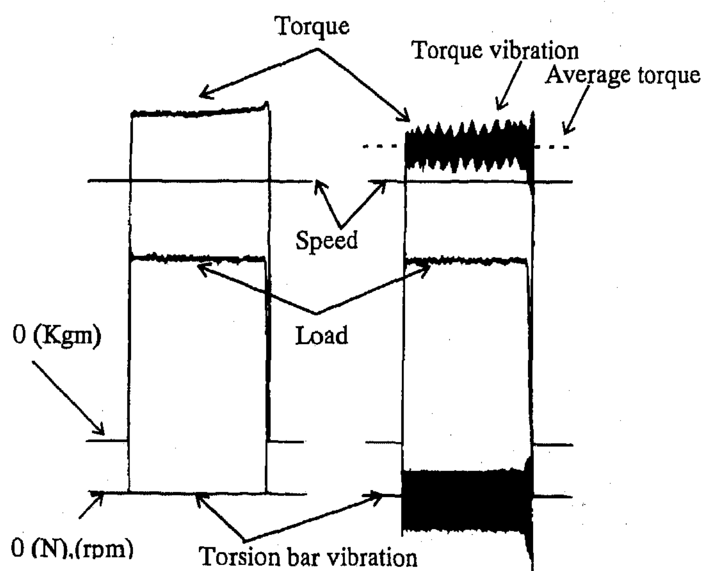


Fig.5 Torque Trace

3. EXPERIMENTAL RESULT

Friction of filler itself must be generated with as little change in porosity and actual contact area as possible, as mentioned in 2.2 and 2.3, for the understanding of filler effect itself on friction characteristics, which is the purpose of this research. Therefore, porosity, oil adsorption speed, strain, pore-size, bearing curve that indicates flatness of friction surface, and observation of adhesion of ATF additives onto friction surface were examined before and after testing. And, as a result, the following have been clear to the extent that the friction characteristics obtained from this experiment have been judged to indicate the characteristics of filler itself.

Examinations of porosity, oil adsorption speed, and pore-size —— Table 5 shows the pre-and post-test changes in the properties of friction materials. Only very small changes were observed. Figure 6 shows Sample A's pore-size distribution as a representative example.
This indicates the pore-size did not change during the experiment.

Table 5 Pre-and Post-test Change in Friction Materials

Symbol	Composition	Porosity (%)	Oil absorption speed (sec)	Strain (μ m)
A	Pulp of low fibrillation degree	0	3	5
B	A + Abrasive powder	1	5	6
C	A + Carbon group powder	2	7	6
D	A + Phenol group powder	1	7	3
E	A + Solid lubricant	1	3	5
F	A + Carbon group fibers	1	8	7
G	A + Phenol group fibers	2	5	8
A'	Pulp of high fibrillation degree	2	6	3
B'	A' + Abrasive powder	2	3	4
C'	A' + Carbon group powder	1	7	4
D'	A' + Phenol group powder	2	7	5
E'	A' + Solid lubricant	1	4	5
F'	A' + Carbon group fibers	2	8	1
G'	A' + Phenol group fibers	1	5	3

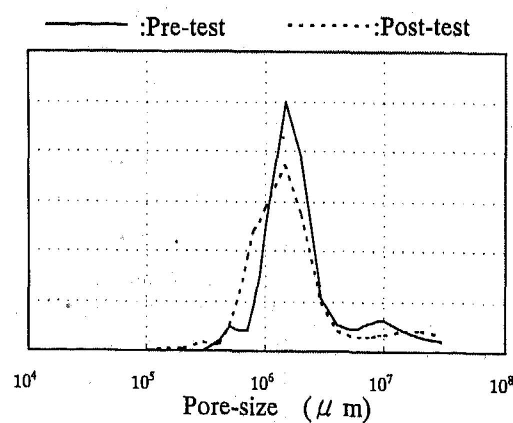


Fig.6 Distribution of Pore-size of Sample A

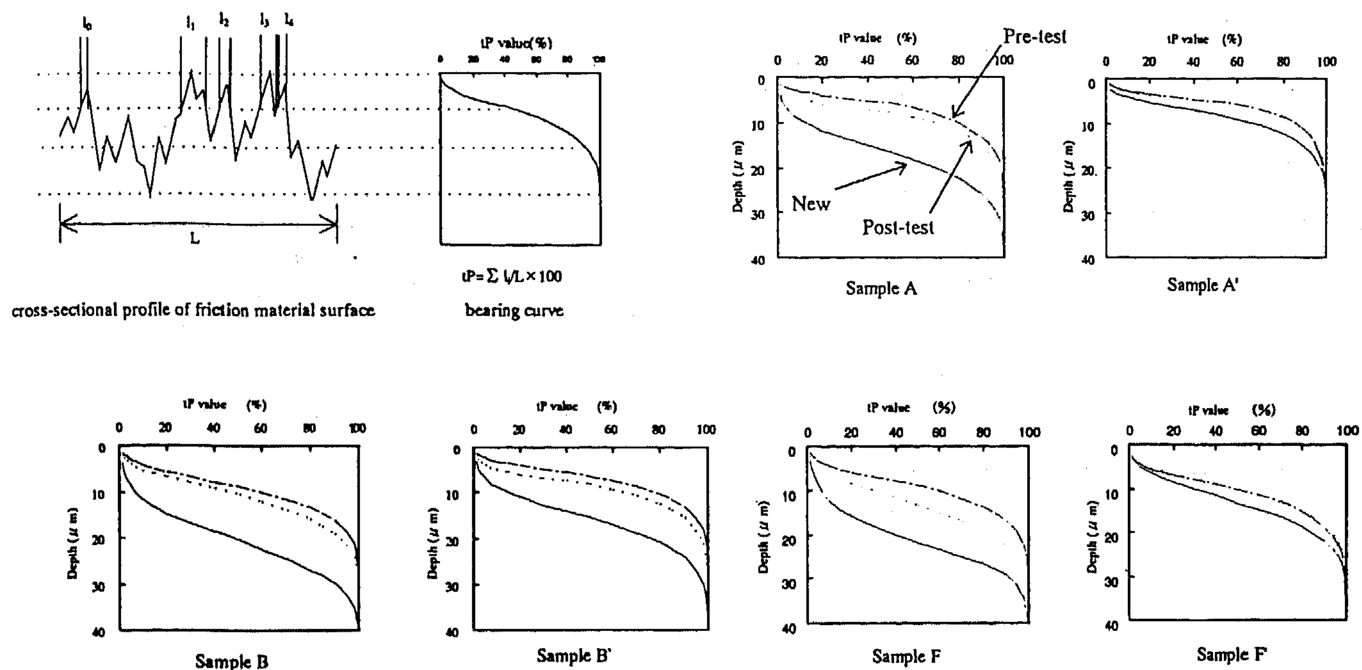
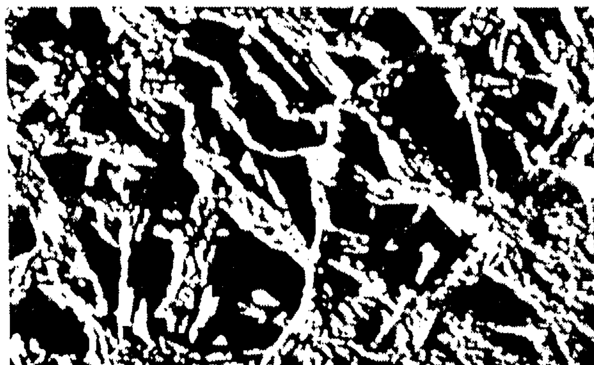
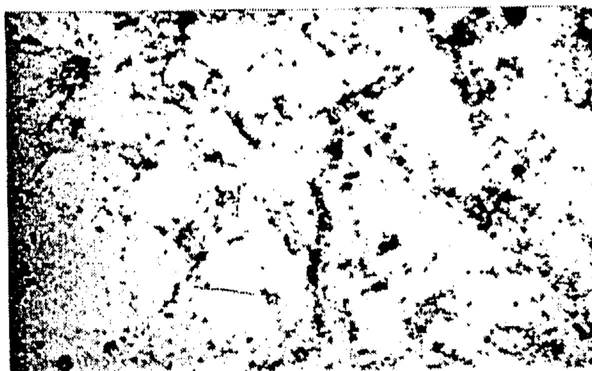


Fig. 7 Changes as time elapse in Bearing Curve



SEM observation



EPMA observation

Fig.8 Adhesion of Ca on Sample A

Examination of flatness of friction surface —— Figure 7 shows the bearing curves of surface conditions for new friction materials, pre-test material and post-test materials. Very small changes were observed between pre- and post test, (pre-test: break in material).

Examination of actual sliding material on friction surface —— Ca, which is one component of ATF additive, is adhered to pulp in friction material by friction in ATF. This phenomena is utilized to examine materials in actual friction.

Adhesion of Ca on post-test friction surface was examined by an EPMA(Electron Probe Micro Analyzer). Elements researched by this analysis are Ca, S, P and B. Only Ca is observed on the surface by this analysis. It proves that Ca will adhere to fibers. Adhesion of Ca is noticeable on friction surface of friction material(sample A) which compounded with only fibrillated fibers. The result is shown in Fig. 8. Little Ca adhesion was observed on fibrillated fibers of the friction surface of the sample compounded with filler on friction surface. This indicates no friction of pulp in friction material of which pulp surface is compounded with filler. In other words, the friction properties of friction material compounded with filler are the properties of filler material itself.

3.1 SPEED DEPENDENCY

Figure 9 shows the change(Δ μ V) in friction coefficient due to speed increase in comparison with the friction coefficient at low slip speed:1rpm(V=6.1 mm/s). This set of date is the average including variations due to oil temperatures and surface pressures.

Speed dependency of fillers is more pronounced for smaller strain of friction material. This is considered to be because less tangential force on fillers is absorbed by elasticity of base fibers when strain is small and shear force added to fillers is larger as compared to that on base fibers of larger distortion.

In regard to the speed dependency of fillers, friction coefficient decreases with increasing slip speed for filler E', with lubrication action under shear force, and filler D' that tends to become film-like, and on the contrary, friction coefficient of filler B', with abrasive action ,increases with increasing slip speed. This is considered to be due to the difference in shape stability of fillers against shear force.

3.2 TEMPERATURE DEPENDENCY

Figure 10 shows the change(Δ μ T) in friction coefficient due to oil temperature increase in comparison with the friction coefficient at 40 ℃ oil temperature. This set of data is the average including variations due to slip speeds and surface pressures.

Similar to the speed dependency, the temperature dependency of fillers is also prominent when strain of friction material is small, and this, too, is considered to be due to the difference of shear force added to fillers which is caused by the different strain of base fibers. Temperature dependency of fillers is related to the thermal stability of filler material and the temperature dependencies of filler B',C', and F' of high thermal stability are low.

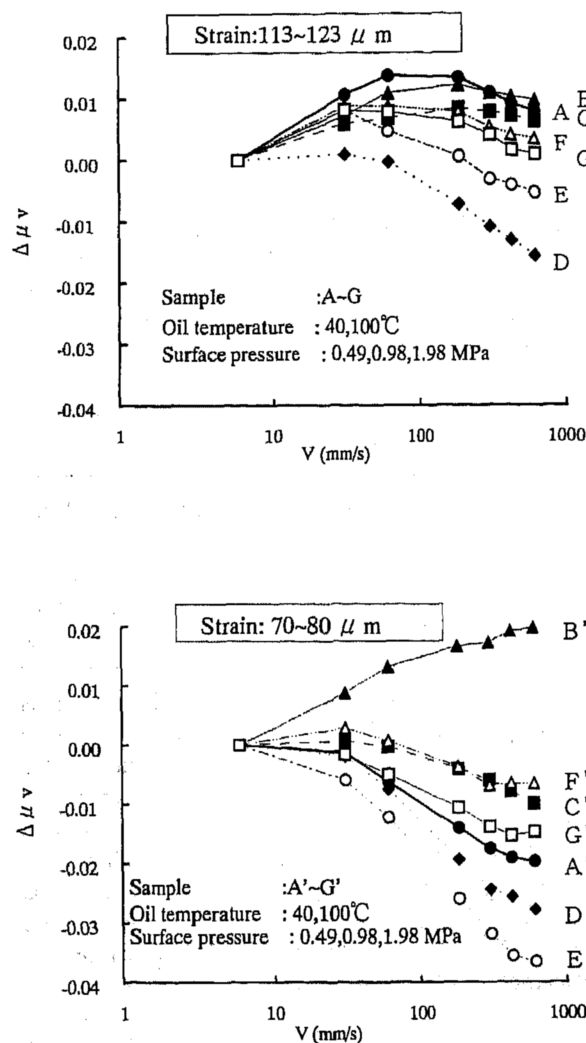


Fig.9 ⊿ μ v-V Characteristics

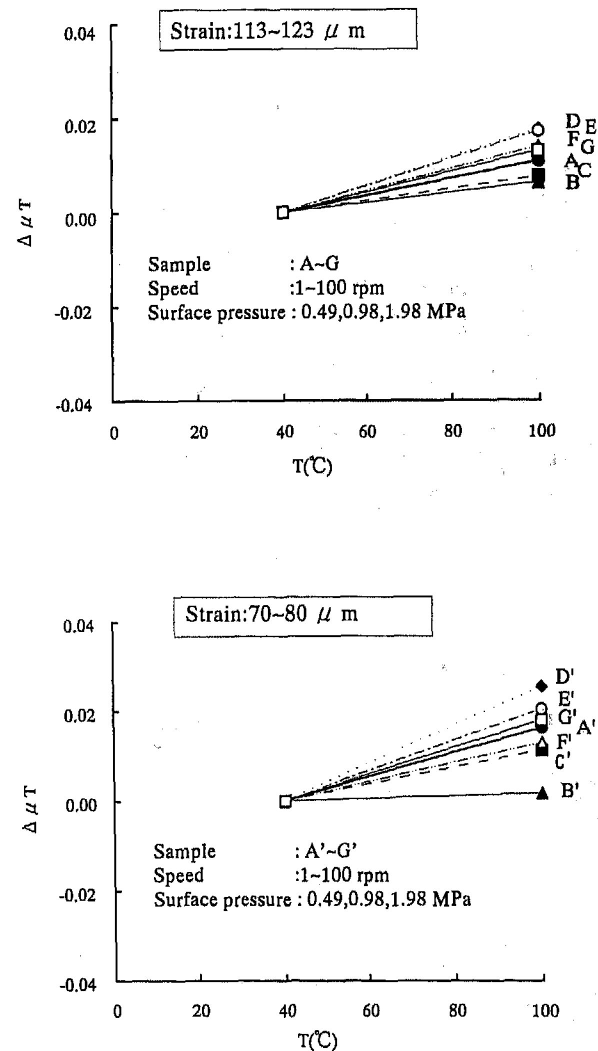


Fig.10 ⊿ μ t-T Characteristics

3.3 SURFACE PRESSURE DEPENDENCY

Change($\Delta\mu$ P) in friction coefficient due to increase in surface pressure is shown in Fig. 11, in comparison with the coefficient at 0.49 MPa surface pressure. This set of data is the average including variations due to oil temperatures and slip speeds.

Surface pressure dependency of fillers is also affected by the strain of friction material, and this character of fillers appears on the side of lower surface pressure when strain is smaller. This is considered to be because of a broader difference, under lower surface pressure, of shear force added to fillers which is caused by different strain of base fibers. Surface pressure dependency of fillers is lower for higher hardness of fillers, and filler B' shows little dependency. This is considered to be due to the difference in shape stability of fillers against load.

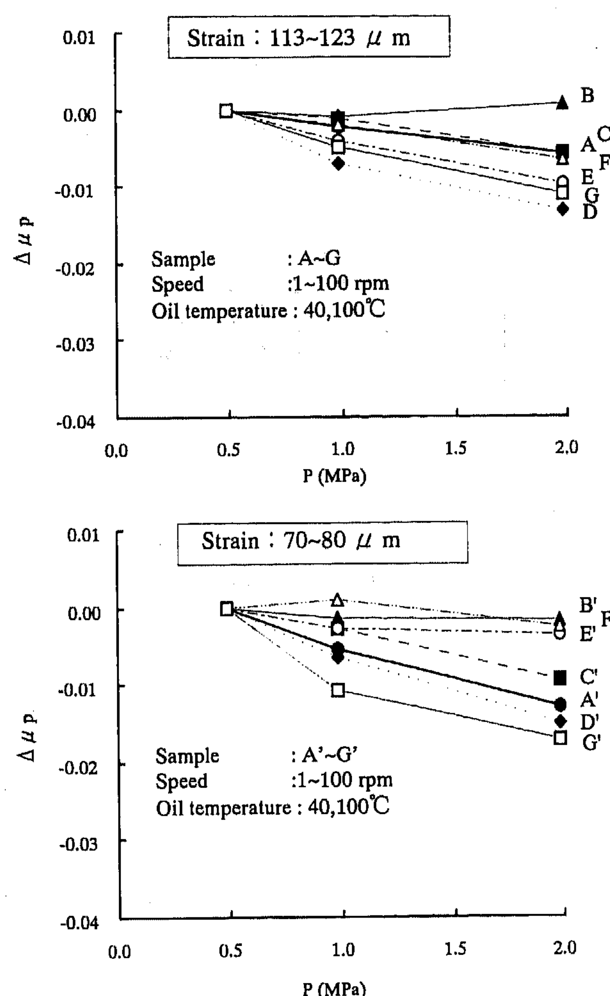


Fig.11 $\Delta\mu$ p-P Characteristics

4. EXAMINATION USING THE STREIBECK DIAGRAM

Friction properties of samples compounded with fillers is dependent on slip speed, oil temperature and surface pressure. These tests results are plotted into a Streibeck Diagram. Figure 12 shows the Streibeck Diagrams of the test samples. The value S is defined by the following equation.

$S = \eta V/P(m)$

where η : ATF viscosity(Pa*S)

P : Surface pressure(Pa)

V : Slip speed(m/S)

Black plots in the diagram indicate the data points which generate friction vibration. μ-S curves are shown are obtained by regression.

4.1 INFLUENCE OF FILLERS ON LUBRICATION CHARACTERISTICS

Figure 12(a) shows the Streibeck Diagram of A-A'. The difference of lubrication characteristics between A and A' is thought to be caused by the difference of thickness of boundary lubrication oil film, because the strain of sample affects the thickness of oil film. At low to medium S value, μ increases along with the increase of shear force of oil film. At medium to high S value, the data indicates the characteristics of fibrillated fibers. Lubrication characteristics of test samples compounded with fillers are compared with the ones of samples A and A'.

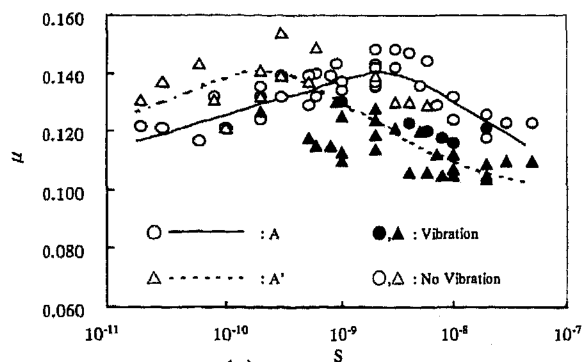
1)Fiber-like fillers (F-F' and G-G')---Figure 12 (f) and (g) show the Streibeck Diagrams of F-F' and G-G'. The differences of characteristics between Carbon group fibers and phenol group fibers are the hardness and temperature resistance. However, Carbon group fibers and Phenol group fibers have similar lubrication characteristics. These fillers do not affect the characteristics of A-A'. This result proves that fiber-like fillers do not affect the lubrication characteristics of the original paper base A-A'.

2)Powder fillers (B-B', C-C', D-D' and E-E')---Figure 12(b), (c), (d) and (e) show the Streibeck Diagrams of B-B', C-C', D-D' and E-E'. The lubrication characteristics of powder fillers depend on the characteristics of fillers.

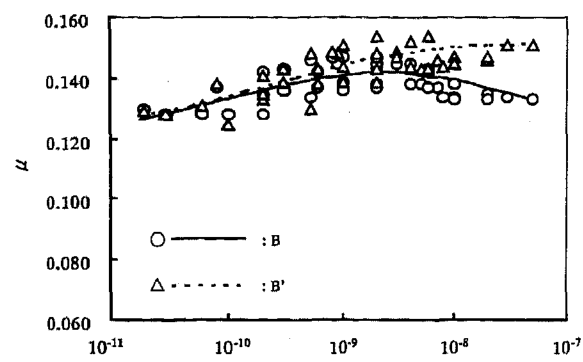
-Abrasive powder (B-B') improves the lubrication characteristics of the original paper base A-A' to positive slope.

-Carbon powder (C-C'), which has high hardness and low abrasion, does not affect the lubrication characteristics of the original paper base A-A'.

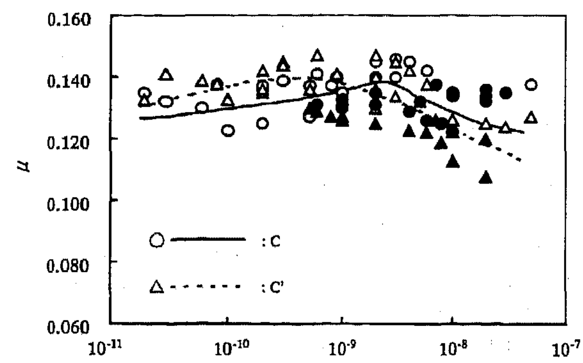
-Materials E, E', D and D', which have low resistance to shear force, show more negative slope on the lubrication characteristics of the original paper base A-A'.



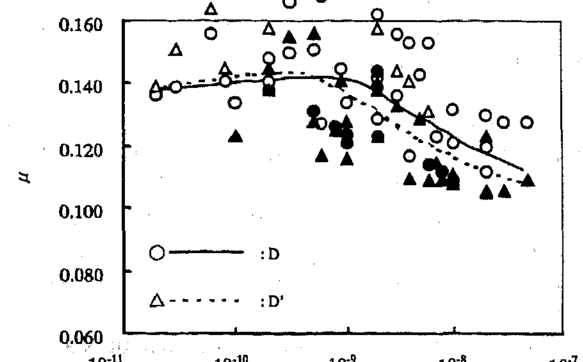
(a) Friction materials No. A, A'



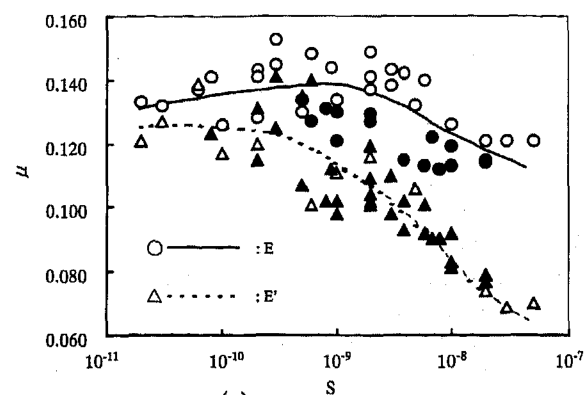
(b) Friction materials No.B,B'



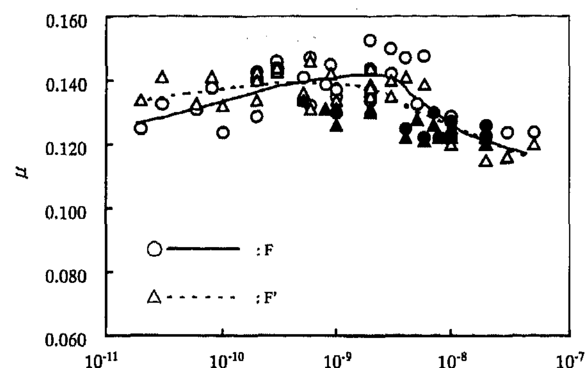
(c) Friction materials No.C,C'



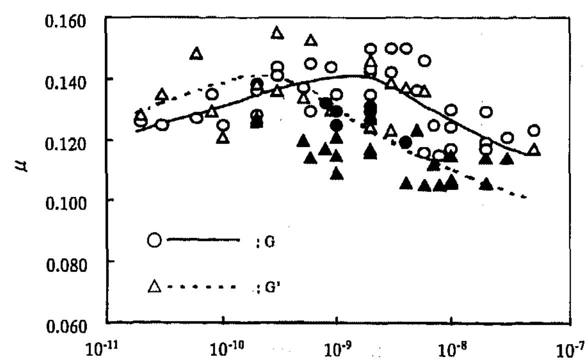
(d) Friction materials NoD,D'



(e) Fiction materials No. E, E'



(f) Fiction materials No.F,F'



(g) Friction materials No.G,G'

Fig.12 The Streibeck Diagrams

3)Comparison of strains---Figure 13 shows the Streibeck Diagrams classified according to the strain. The less compliant paper obviously represents the characteristics of fillers. As mentioned earlier, the difference of strain affects the forming of boundary lubrication oil film. The friction force of fillers has more effect on less compliant paper because thinner oil film forms on the surface.

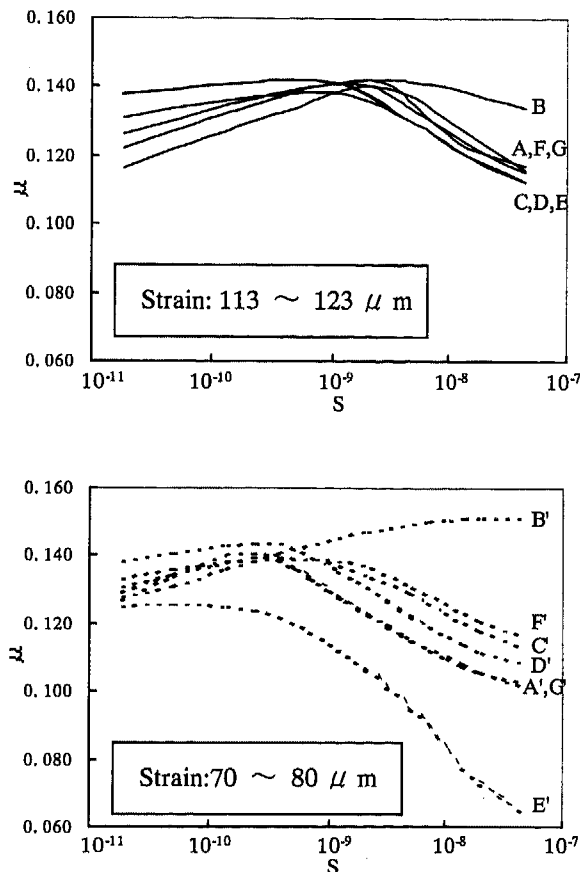


Fig.13 The Streibeck Diagram Classified according to the Strain

4.2 FRICTION VIBRATION CHARACTERISTICS

Black plots in the Streibeck Diagram of Fig.12 means the data was observed with occurrence of friction vibration. Figure 5 shows the torque trace with friction vibration. There is no vibration where the curve has positive slope. The vibration starts to occur around the maximum of the curve where the slope changes from positive to negative. This means friction materials, which have positive slopes on the μ-S curve, are effective in preventing vibrations. In this study, friction material B' is the most effective paper to prevent vibrations.

5. Conclusion

To summarizes the results:

(1)Friction materials, which have positive slope on the μ-S curve in the Streibeck Diagram, effectively improve the slip characteristics. The characteristics of fillers contribute to the slope. μ-S curves can be classified into three types by the effect of fillers.

μ – S curve	filler
	abrasive powder
	carbon group fibers phenol group fibers, carbon group powder
	solid lubricant phenol group powder

(2)The abrasive powder is effective in improving the slip characteristics of friction materials, because the fillers have the shape stability to withstand shear force, load and temperature.

(3)The abrasive fillers work more efficient by using less compliant paper. The friction force created by the fillers increases as shear force of oil film decreases. This phenomena is caused by the formation of a thin boundary lubrication oil film. The oil film formation is affected by the strain of paper. The fillers, like solid lubricant or phenol group powder, affect the slip characteristics of friction material so as to make more negative slope. Therefore, it is very important to understand the characteristics of the fillers to improve the slip characteristics.

We think the amount of abrasive fillers used in the wet friction materials will increase to the slipping clutch applications.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Mr. Asai for his support and cooperation in preparing this study.

REFERENCES

1) Proceeding of JAST Tribology Conference, Kanazawa,1994-10,P903-906
2) Proceeding of JAST Tribology Conference, Tokyo,1995-5,P453-456
3) Y.Kato,R.Akasaka,T.Shibayama,Experimental Study on Lock-Up Shudder Mechanism of an Automatic Transmission, Journal of Japanese Society of Tribologists, Vol.39,No.12(1994) P.1067-1072
4) M.Eguchi,M.Takesue,T.Yamamoto,Friction Characteristics of a Paper-Based Facing for a Wet Clutch, Journal of Japanese Society of Tribologists, Vol.36,No.7(1991) P.535-542
5) S.Shirahama,Adsorption of Additives on Wet Friction Pairs and Their Frictional Characteristics, Journal of Japanese Society of Tribologists, Vol.39,No.12(1994) P.1037-1041
6) Proceeding of JAST Tribology Conference, Kanazawa,1994-10,P915-918
7) Proceeding of JAST Tribology Conference, Kanazawa,1994-10,P927-930
8) T.Matsumoto,The Influence of Paper-Based Friction Material Porosity on the Performance of a Wet Clutch, SAE Paper 941032 (1994)
9) A.Grzesiak,R.C.Lam,Application of Fluid Transport Phenomenon to Transmission Band Friction Material, SAE Paper 922098 (1992)
10) H.Sano,M.Takesue,Friction Characteristics of Wet Clutch in the Process of Engaging by SAE No.2 Tester, Journal of Japanese Society of Tribologists, Vol.39,No.12(1994) P.1089-1096
11) S.Sanada,Frictional Characteristics of a Wet Clutch Composed of Paper-Based Facings during Running-In Process, Journal of Japanese Society of Tribologists, Vol.39,No.12(1994) P.1047-1053

970976

Wet Clutch Performance in a Mineral-Based, and in a Partial-Synthetic-Based Automatic Transmission Fluid

Bülent Cavdar* and Robert C. Lam
Borg Warner Automotive

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The physical and chemical properties of a mineral-based (Fluid-M) and a partial-synthetic-based (Fluid-PS) automatic transmission fluid were compared by the analyses of Gel Permeation Chromatography (GPC), Thermogravimetry (TG), Fourier Transform Infrared Spectroscopy (FTIR), Nuclear Magnetic Resonance (NMR), Gas Chromatography (GC), viscometry, thermo-oxidative stability, torque response curve shape and metal-to-metal wear preventive characteristics. The effects of various properties of Fluid-M and Fluid-PS on wet friction material performance were investigated from the viewpoints of compressibility, durability, tensile strength, surface interactions and friction-pressure-speed-temperature characteristics. Friction material specifications for partial-synthetic fluid applications will be different from those for mineral-based fluid applications.

GPC showed that Fluid-PS has a higher concentration and a lower molecular weight of VI Improver than Fluid-M. Chemical analyses indicated major differences in the detergent, anti-wear and anti-oxidant additive packages of Fluid-M and Fluid-PS. Thermal oxidation tests indicated different degradation paths for the fluids. Both fluid formed acid, ketone, and ester. However, Fluid-M showed more tendency to polymerize to diester or polyester than Fluid-PS. Fluid-PS reduced friction and wear in metal-to-metal contacts in 4-Ball tests as compared to Fluid-M. The less metallic wear in Fluid-PS was explained with the superior viscometrics and anti-wear additive characteristics of the fluid. Fluid-M and Fluid-PS had similar torque response curve performance in SAE tests at low temperatures. However, Fluid-PS showed an increasing friction with decreasing speed (negative μ-v slope) whereas Fluid-M showed a positive μ-v slope at high temperatures and low speeds. Hence, Fluid-PS tended to cause more "shudder" at high temperatures, especially with rigid friction materials. The factors contributing to the negative μ -v slope in Fluid-PS are the narrow molecular weight distribution of the synthetic base oil, and the detergent free additive package of the fluid. In general, the friction materials wetted with Fluid-PS had less compressibility than the ones wetted with Fluid-M. The difference in compression modulus in the fluids were explained with the differences in base oil type and VI improver package of the fluids. Fluid-PS' tendency to increase the rigidity of a wet clutch system can be compensated by tailoring the compressibility of friction materials.

INTRODUCTION

Original Equipment Manufacturers (OEM) design their new Automatic Transmission Fluid (ATF) specifications for better fuel economy and longer fluid life. The new MERCON®V specification for automatic transmission fluids has been announced recently (1). MERCON®V specification requires superior viscometry, higher shear, and oxidative stabilities, and less metal-to-metal wear than the current MERCON® specification (2). Similar trends are expected in the new DEXRON® and overseas specifications (3). For better fuel economy, new generation ATFs are required to provide high friction coefficients during clutch engagements, and low drag losses during freewheeling. Sub-zero temperature-drag losses can be reduced by increasing the fluidity of ATF at low temperatures. High friction coefficients and little metal-to-metal wear can be achieved with synthetic lubricants and effective anti-wear additives. Synthetic-based fluids meet most of the requirements of low pour point, high viscosity index, high shear viscosity, and high oxidative stability. However, synthetic-based fluids are

* Author to whom correspondence should be addressed.

considerably more expensive than the mineral-based fluids. A necessity to compromise between the performance and cost leads to partial-synthetic fluids.

The performance of a wet clutch system depends on the compatibility of ATF with the friction material and the metal mating surface. The study of physical and chemical interactions of ATFs and friction materials is a formidable task mainly due to the complexity and secrecy of the formulations of ATFs and friction materials. Nevertheless, the design of better wet clutch systems requires a solid understanding of the interaction phenomena. Recently, there is a trend to use partial-synthetic based ATFs because of their superior viscosity characteristics. The interactions of various friction materials with a mineral-based commercial ATF and a partial-synthetic-based commercial ATF were studied in this research so that an informed selection of friction materials for these fluids can be made. The systematic approach of the research and general discussions of ATF and friction material properties are of interest to all OEMs and component suppliers. The physical and chemical properties of Fluid-M and Fluid-PS were compared by the analyses of viscometry, thermo-oxidative stability, and metal-to-metal wear preventive characteristics. Furthermore, effects of Fluid-M and Fluid-PS on wet clutch performance were investigated from the viewpoints of compressibility, durability, and friction-pressure-speed-temperature (μ -P-v-T) characteristics. The difference in compression modulus in the fluids was explained with the differences in base oil type and VI improver package of the fluids. The effects of base oil type, VI improver, friction modifier, detergent, dispersant and antiwear additives on the torque curve shape were also discussed.

EXPERIMENTAL METHODS

The properties of fluids were analyzed by viscometry, Inductively Coupled Plasma (ICP), Fourier Transform Infrared Spectroscopy (FTIR), Gas Chromatography (GC), Gel Permeation Chromatography (GPC), Nuclear Magnetic Resonance Spectroscopy (NMR), Thermogravimetry (TG), and mechanical tests. Mechanical tests included 4-Ball Wear Tests, compressibility tests, and SAE No.2 Test Machine tests. Two types of tests were run with the SAE No.2 Test Machine: durability test and μ -P-v-T test. The fluid viscosities were measured at 40°C and 100°C with a capillary viscometer, then, viscosity index was calculated. High temperature-high shear viscosity was measured with a tapered bearing simulator according to ASTM D4683. Spectrochemical analysis was done with the Spectroflame ICP module. FTIR spectra were obtained with 200 scans at 4 cm^{-1} resolution. The spectroscopic measurements of fluids were taken with a KBr cell of 0.2 mm fixed path length. GC was used for the identification and quantification of the synthetic portion of Fluid-PS. The GC system consisted of a Varian 3400 instrument equipped with a flame ionization detector and Star Workstation. GPC analysis was done to determine the molecular weight distribution of the fluids. The GPC system was equipped with Waters 610 isocratic solvent delivery system and a Waters 410 refractive index detector. The GPC analyses were conducted with four columns having the exclusion limits of 1500, 5000, 20000, and 400000 molecular weight. A GPC calibration curve was obtained with nine narrow standard samples of polystyrene and polymethylmethacrylate with molecular weights ranging from 200 to 375000 dalton.. NMR analyses were done with ^{31}P and ^{13}C isotopes on a Varian UNITYplus 400 system.

High resolution thermogravimetry was done with a TA-2950 apparatus. The heating rate was 200°C/min up to 100°C, then it was 20°C/min up to 950°C when no derivative was sensed. In high resolution TG, heating is switched to a very small temperature increase when a derivative is sensed. A flow of air and nitrogen was supplied to the sample chamber at a rate of 100ml/min. Wear preventive characteristics of the fluids were tested with a four-ball machine according to ASTM D2266 with one repetition. The submerged static compressibility tests of friction materials in ATFs were done after a 5-minute soaking at 93°C. Compressive load was applied in a mechanical test machine (MTS 810 Axial Test Machine) for 60 seconds before the displacement was measured. The load was removed for 60 seconds, and permanent compression set was measured. Then, the compression load was stepped up to the next level. Displacements and compression sets were measured at 11 load steps from 0.35 MPa to 10.35 MPa. A graphical representation of the test procedure is presented in Figure 1. The μ-P-v-T tests were run at four temperatures. At each temperature the speed and pressure were increased simultaneously in four steps as presented in Table 1. The durability tests were carried out in a SAE No.2 Test Machine with two friction plates

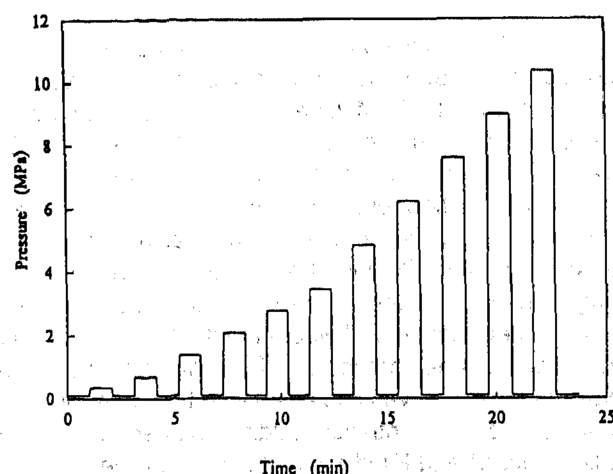


Figure 1. Static-stepwise-compression test procedure

Table 1. μ-P-v-T test procedure for SAE No. 2 Test Machine. The following procedure is repeated at 30°C, 80°C, 100°C and 120°C. There are 3 friction plates and 4 separator plates in a clutch pack Friction plate lining dimensions are OD/ID (mm): 126/106.

Level	Rpm	Pressure (kPa)	Inertia (Nms^2)	Energy (J)	Number of cycles
A	800	190	0.348	1220	50
B	1400	380	0.348	3750	25
C	2600	760	0.348	12900	25
D	3600	1140	0.348	24800	25

and three separator plates and with 1 liter of ATF. The inner and outer diameters of friction plates were 104 mm and 126 mm respectively. The absorbed energy was 36946 J and the energy density was 1.66 W/mm^2 at each engagement cycle. Thirty ml of oil sample was collected at 500 cycles, and the rest of the oil was collected at the end of the durability tests for chemical analyses.

The thermo-oxidative stability of pure ATFs and ATFs mixed with various resins was studied with an in-house developed test. The resins were cured and ground before the catalytic thermal oxidation tests. One part of resin was mixed with two parts of ATF, by weight. Then, the mixture was heated up to 150°C and held at that temperature for 300 hours. A constant air flow of 3 l/min was supplied to the mixture through a capillary tube. The mixture was continuously stirred during the heat tests. ATF was separated from resin by filtering, solvent washing, and oven drying after the thermo-oxidation tests. The residual ATF and resin were analyzed by FTIR and TG.

RESULTS

ANALYSES OF THE MINERAL AND THE PARTIAL-SYNTHETIC FLUIDS - The physical and elemental properties of Fluid-M and Fluid-PS are listed in Table 2. The viscosity characteristics of the fluids are quite different. Fluid-PS has a higher viscosity index and a high-temperature-high-shear viscosity than Fluid-M. The elemental analysis showed significant differences in the additive packages of the fluids. Fluid-PS has more phosphorus but less boron and sulfur than Fluid-M. Fluid-M has a high concentration of calcium and sulfur. Since alkylsulfonates are frequently added into oils as detergent additive, Fluid-M probably contains calcium sulfonate which commonly acts as a buffer. The total base number (TBN) of Fluid-M is significantly higher than the TBN of Fluid-PS supporting the presence of an overbased-calcium sulfonate in Fluid-M (Table 2). Fluid-PS does not have any calcium and has a lower concentration of sulfur. Phosphorus-, sulfur-, and boron-containing additives are frequently added to oils as anti-wear additives. Both fluids have a very small amount of silicon (5-6 ppm). Silicon-containing additives are added as anti-foaming agents. Fatty acids are added into oils as

Table 2. Physical and elemental properties of Fluid-M and Fluid-PS.

	Procedure	Fluid-M	Fluid-PS
Viscosity (cSt) at 40°	ASTM D445	41.98	37.11
at 100°		8.27	8.10
Viscosity Index	ASTM D2270	148	157
Brookfield Viscosity (cP) at -18°	ASTM D2983	1360	805
High Temperature (150°C)/High Shear (10^6 s^{-1}) Viscosity (cP)	ASTM D4683	2.26	2.68
Mol. Weight (Base Oil)	(by GPC)	483	~520 (partial-synthetic)
Mol. Weight (VI Imp.)	(by GPC)	165000	42000
4-Ball, Wear Dia. (μm), run1	ASTM D2266	447	380
run2		513	400
4-Ball, Avg. Fric. Coef. run1	ASTM D2266	0.079	0.049
run2		0.080	0.061
Total Acid Number (mg KOH / g)	ASTM D664	1.05	0.74
Total Base Number (mg KOH / g)	ASTM D4739	2.24	0.96
pH		7.7	7.46
Spectrochemical Analysis	(by ICP)	(ppm)	(ppm)
Iron		0	0
Silicon		5	6
Boron		156	98
Sodium		2	1
Calcium		922	1
Phosphorus		98	175
Zinc		0	0
Sulfur Concent. (%)	ASTM D4294	0.218	0.148

friction modifiers. The amount of fatty acids in the fluids was determined according to the modified Twitchell method, which is described in Reference 4. The carboxylic acid concentrations are 2.8% and 2.7% in Fluid-M and Fluid-PS respectively. At this moment, it is not known whether there were other kinds of friction modifiers in the fluids.

GPC analysis showed that Fluid-M is a hydrocarbon based fluid with a broad Molecular Weight Distribution (MWD) centered at 483 dalton (Figure 2). On the other hand, Fluid-PS is a blend of a narrow MWD and a broad MWD base stock (Figure 3). The narrow and broad MWDs are characteristic of synthetic and mineral base oils, respectively. The synthetic base stock has two narrow bands: an intense peak centered at 600 dalton, and a small peak centered at 770 dalton. The synthetic base of the Fluid-PS was identified as polyalphaolefin by

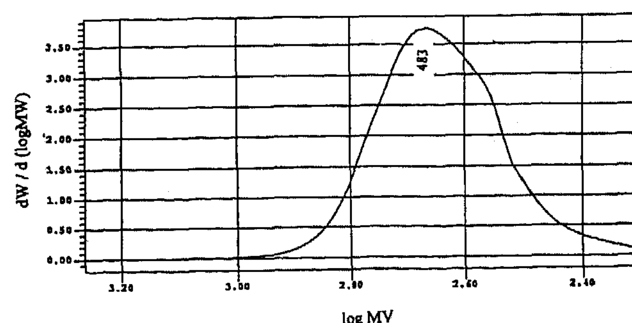


Figure 2. Molecular weight distribution of Fluid-M by GPC.

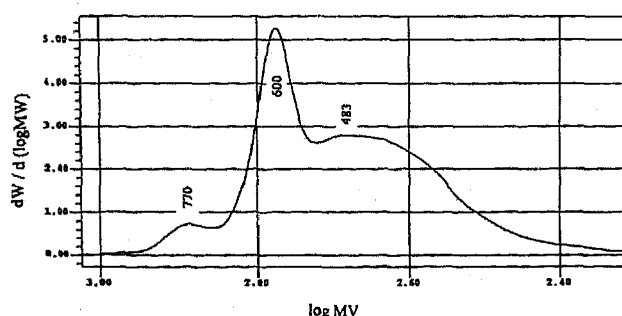


Figure 3. Molecular weight distribution of Fluid-PS by GPC.

GC and quantified as 20.4% decene dimer and 2% decene trimer in the entire fluid. Probably 600 and 770 dalton peaks in the GPC chromatogram correspond to decene dimer and decene trimer, respectively. GPC chromatograms have also indicated that both fluids have a high molecular weight and a medium molecular weight polymer which are probably the viscosity index improver and the dispersant, respectively. The molecular weight of the heavier polymer is about 42000 in Fluid-PS and 165000 in Fluid-M. The molecular weight of the dispersant is about 5000 in both fluids. The polymer fractions which were eluted from the GPC columns were deposited on a Germanium disk and analyzed by FTIR. The VI improver in Fluid-PS has a very similar spectrum to that of polyhexadecylmethacrylate. Since the VI improver amount in Fluid-M was quite small, the signal-to-noise ratio of the infrared spectrum of the fractionate was not very good. Nevertheless, FTIR spectrum of VI improver of Fluid-M showed peaks at 1730 cm^{-1}, 1467 cm^{-1}, 1451 cm^{-1}, 1221 cm^{-1} and 1161 cm^{-1} which corresponded to polyalkylacrylates or polyalkylmethacrylates. Carbon NMR spectrum of Fluid-M suggested that VI improver on Fluid-M is probably a polyalkylacrylate. The FTIR spectrum of the 5000 molecular weight polymer matched that of the polyisobutylene esters, which are added to ATFs as dispersants. Carbon NMR confirmed the presence of polyisobutylene structure in both fluids in equal amounts.

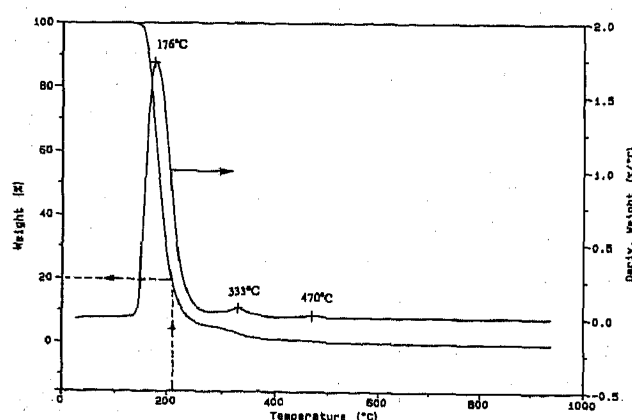


Figure 4. TG graph of Fluid-M.

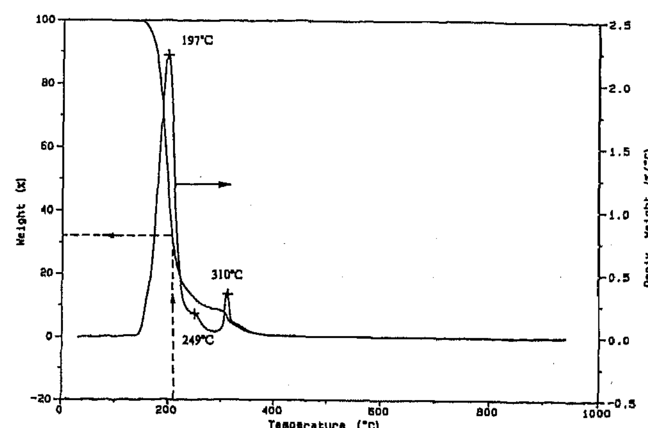


Figure 5. TG graph of Fluid-PS.

The interfacial temperatures between the separator plate and friction plate during a clutch engagement can reach several hundred degrees centigrade (5). The applied pressure in a wet clutch is on the order of a few MPa which is not high enough to prevent ATF from evaporation, hence, during each engagement, some portion of ATF evaporates readily. Therefore, it is very important to know the temperature-weight loss characteristics of an ATF, in order to understand its wet clutch performance. High resolution TG analysis of Fluid-M and Fluid-PS indicated major differences in the weight loss versus temperature characteristics of the fluids. TG derivative curves, shown in Figures 4 and 5 for Fluid-M and Fluid-PS, respectively, indicate that the peak temperature of the weight loss curve of Fluid-PS is higher than that of Fluid-M. However, Fluid-PS has a narrower and less homogeneous weight loss distribution than Fluid-M. The TG derivative curve of Fluid-PS skewed towards high temperature with a peak located at 197°C. The peak width at half the height of the peak of Fluid-PS, ranges from 177°C to 210°C, whereas, that of Fluid-M ranges from 155°C to 210°C with a peak located at 176°C. This indicates that during a clutch engagement, Fluid-PS will remain liquid longer than the Fluid-M at the interface. However, the rate of evaporation (% weight loss per degree centigrade) of Fluid-PS will be approximately twice as high as that of Fluid-M in the temperature range of 180°C and 210°C.

The TG curve of Fluid-PS has a shoulder centered at 249°C. The shoulder is probably due to that part of the synthetic base oil, which has a narrow molecular weight distribution centered at 770 molecular weight, as shown in Figure 3. High resolution TG also showed that Fluid-PS has about 9% VI improver, which was lost in a narrow temperature range centered at 310°C. Fluid-M has about 3% VI improver, which evaporates and/or degrades in a broader temperature range centered at 333°C. Furthermore, Fluid-M contains 1% of an additive which is lost at 470°C. This may be an inorganic material such as calcium sulfonate. Fluid-PS does not have this additive.

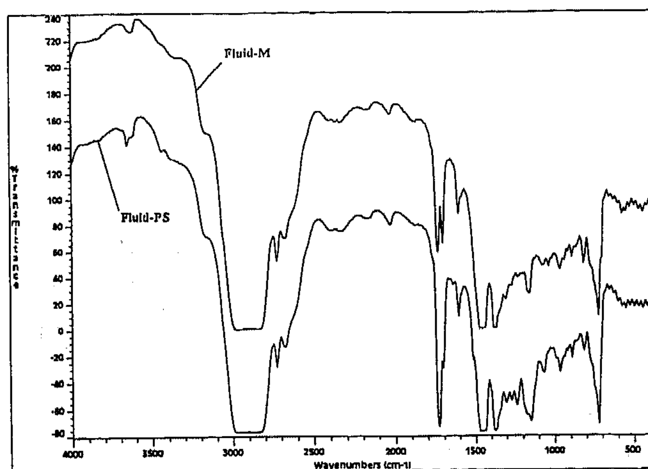


Figure 6. FTIR spectra of Fluid-M and Fluid-PS.

The composition of the residual fluids at 210°C of TG tests gives an idea about the homogeneity of the residual fluids. Here, homogeneous means "composed of the same kind or nature." There remains 20% of Fluid-M, and 26% of Fluid-PS as residuals at 210°C. The derivative weight loss curves suggest that the VI improvers (333°C and 310°C bands) and the inorganic additive (470°C band) will remain in the liquid phase at 210°C (Figures 4 and 5). The residual Fluid-M at 210°C consists of 15% VI improver, 4% inorganic additive, and 81% base oil and other additives. On the other hand, the residual Fluid-PS at 210°C is less homogeneous: 35% VI improver, and the rest are mineral base oil, synthetic base (decene dimer and decene trimer) plus other additives. Obviously, the residual of Fluid-M is more homogeneous than the residual of Fluid-PS at 210 °C.

FTIR analysis of Fluid-M and Fluid-PS indicated that Fluid-PS has a much stronger peak at 1740 cm^{-1} than Fluid-M (Figure 6). Moreover, Fluid-PS had stronger peaks from 1000 cm^{-1} to 1300 cm^{-1} than Fluid-M. The peak of 1740 cm^{-1} corresponds to the carbonyl vibrations of an ester. Esters are frequently used to improve additive solubility and/or elastomer compatibility. Ester linkages can also be found in VI improvers and dispersants. The FTIR finding is consistent with the GPC and TG data in that Fluid-PS has considerably more VI improver than Fluid-M. Furthermore, Fluid-PS has a sharp peak at 3650 cm^{-1}, unlike Fluid-M. Presence of a 3650 cm^{-1} peak in Fluid-PS indicates there is a hindered phenol-type antioxidant in Fluid-PS, which was also confirmed by Carbon NMR.

Four-Ball-Wear Tests showed that the friction and wear of sliding steel balls were reduced considerably in Fluid-PS, as compared to Fluid-M. The additives of phosphorus and boron chemistry are blended into automatic transmission fluids to reduce wear in metal-to-metal contacts. The concept of boundary lubrication suggests that the anti-wear additives function by forming protective surface films over the sliding surfaces to

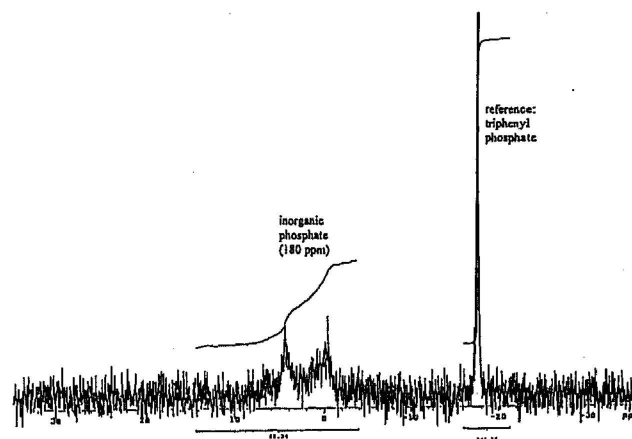


Figure 7. ^{31}P-NMR spectrum of Fluid-PS.

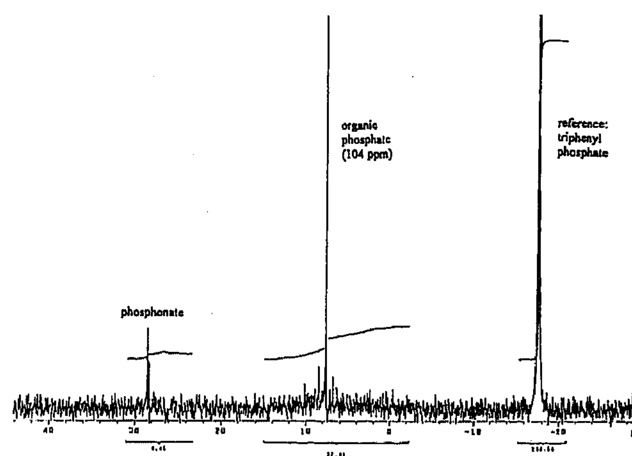


Figure 8. ^{31}P-NMR spectrum of Fluid-M

reduce wear (6,7). The chemistry of phosphorus anti-wear additive affects the durability of protective films. Phosphorus NMR analysis of the fluids yielded that Fluid-PS has an inorganic phosphate (Figure 7), whereas, Fluid-M has both organic phosphate and phosphonate (Figure 8), as anti-wear additives. The fluid samples were spiked with triphenylphosphate to quantify the phosphorus concentrations in the fluids. The phosphorus concentrations of the fluids measured by NMR, 104 ppm and 180 ppm for Fluid-M and Fluid-PS, respectively, match the ones obtained by ICP analysis (Table 2).

THERMAL OXIDATION TESTS OF THE MINERAL AND THE PARTIAL-SYNTHETIC FLUIDS - Thermal oxidation tests showed that Fluid-PS and Fluid-M follow different oxidation paths. The FTIR spectra of the fresh fluids were subtracted from those of the tested fluids to observe the changes that occurred in the fluids during the thermal oxidation tests. Figures 9 and 10, the subtraction spectra of Fluid-M and Fluid-PS, show the presence of three carbonyl groups: 1691 cm^{-1}, 1726 cm^{-1}, and 1782 cm^{-1} bands. The thermal oxidation products of a mineral base oil were identified as ketone, acid (1695 cm^{-1}), ester (1734 cm^{-1}), and diester (1778 cm^{-1}) according to Figure 8 of Reference 8. The small

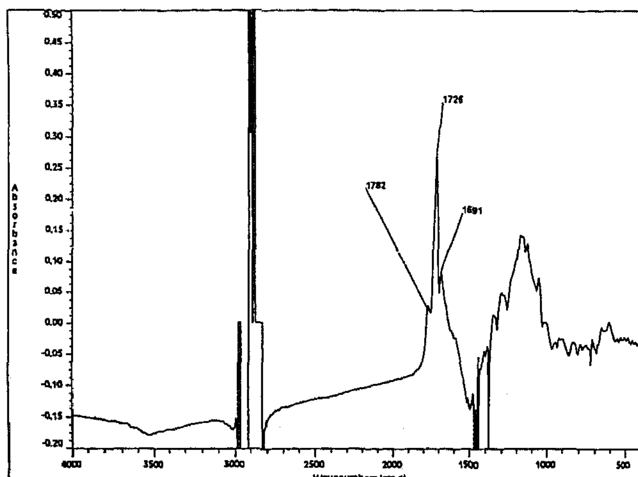


Figure 9. FTIR subtraction spectrum: (Fluid-M after thermal oxidation test) - (Fresh Fluid-M).

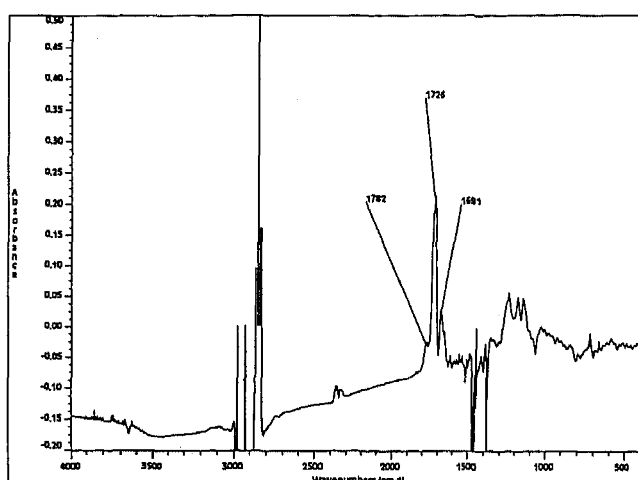


Figure 10. FTIR subtraction spectrum: (Fluid-PS after thermal oxidation test) - (Fresh Fluid-PS).

shifts (4 to 8 cm^{-1}) in peak locations between this study and Reference 8 is negligible, and may be due to the instrumentational differences between the laboratories, and the compositional differences between the fluids (formulated fluids versus base oil). During the thermal oxidation tests, both Fluid-PS and Fluid-M formed acid, ketone and ester products (1691 cm^{-1} and 1726 cm^{-1} bands). However, Fluid-M formed more diester (1782 cm^{-1}) than Fluid-PS did. In catalytic thermal oxidation tests, various commonly used wet clutch resin chemistries such as phenolic, modified phenolic, and silicon resins were added into the fluids to see the effects of various resins on the degradation of the fluids. As an example, Figures 11 and 12 show the effect of phenolic resin on the carbonyl group formation in Fluid-M and Fluid-PS. As compared with Figures 9 and 10, the diester peak intensity in Fluid-M and acid, ketone, ester peak intensities in Fluid-PS increased when the fluids were tested in the presence of the resin. The diester formation in Fluid-M and the acid, ketone, ester formation in Fluid-

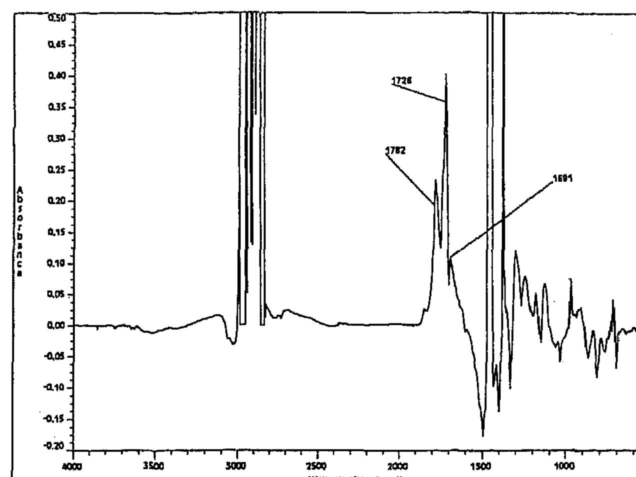


Figure 11. FTIR subtraction spectrum: (Fluid-M after catalytic thermal oxidation test with phenolic resin) - (Fresh Fluid-M).

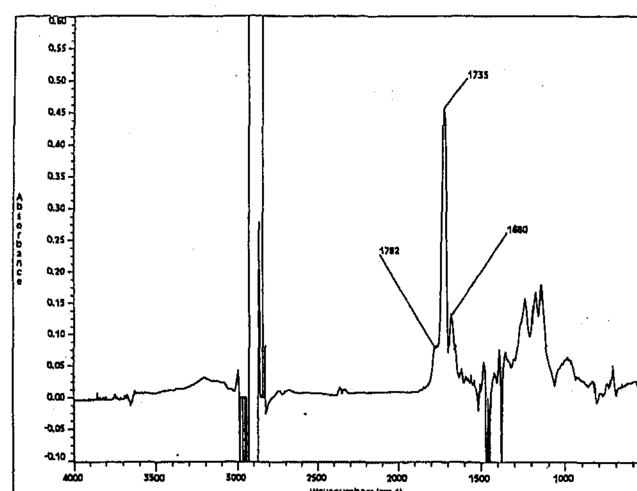


Figure12. FTIR subtraction spectrum: (Fluid-PS after catalytic thermal oxidation test with phenolic resin) - (Fresh Fluid-PS).

PS were increased more or less by any kind of resin tested.

MECHANICAL TESTS IN THE MINERAL AND THE PARTIAL-SYNTHETIC FLUIDS - Eight different friction materials were compression tested in Fluid-M and Fluid-PS. Figure 13 shows the static compression-displacement and permanent compression set behavior of a friction material in Fluid-M. The curve on the right is the displacement under 60 seconds of compression and the curve on the left is the permanent-compression-set measured 60 seconds after the load is reduced to 100 kPa. Figure 14 shows the compression curves of the same material in Fluid-PS. The slope of a compression-displacement curve is defined as the compression modulus. Due to the nonlinear shape of the compression-displacement curves of the friction materials, it is more suitable to define two compression moduli: one for the low pressure region, from 0 to 2 MPa,

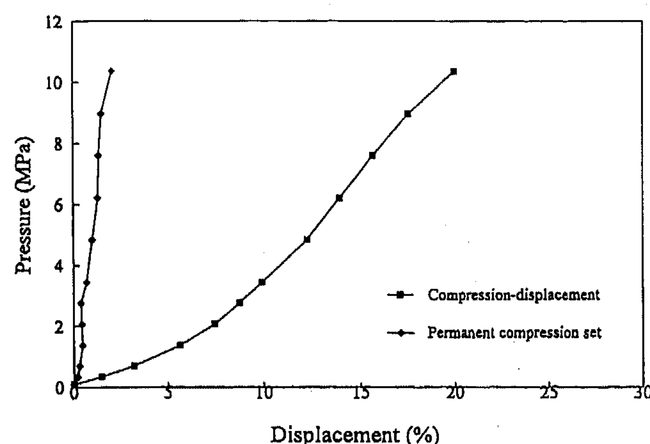


Figure 13. Compression-displacement and permanent compression set curves of Friction Material#5 in Fluid-M.

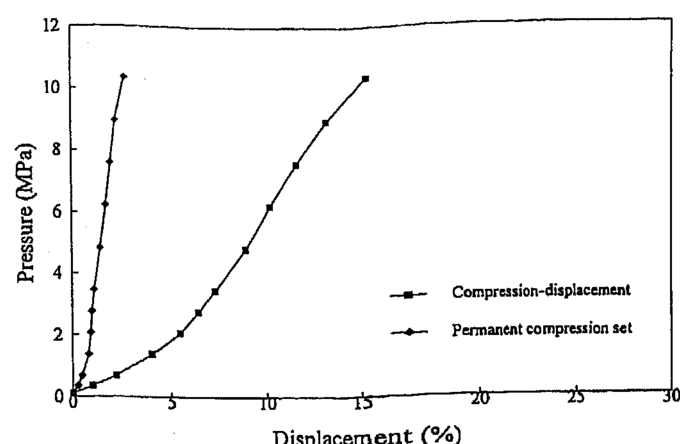


Figure 14. Compression-displacement and permanent compression set curves of Friction Material#5 in Fluid-PS.

Table 3. Low pressure (0-2 MPa) and high pressure (2 to 10.4 MPa) compression moduli of friction materials in Fluid-M and Fluid-PS

	Low pressure modulus (kPa)			High pressure modulus (kPa)		
	in Fluid-M	in Fluid-PS	Difference (%)	in Fluid-M	in Fluid-PS	Difference (%)
Fric. Mat. #1	37570	50480	34.4	65820	68150	3.5
Fric. Mat. #2	56850	61510	8.2	56470	60830	7.7
Fric. Mat. #3	49790	47690	-4.2	86190	83780	-2.8
Fric. Mat. #4	23620	25580	8.3	49050	52950	8.0
Fric. Mat. #5	36750	38550	4.9	88510	103040	16.4
Fric. Mat. #6	38190	44070	15.4	89060	93680	5.5
Fric. Mat. #7	54420	79090	45.3	112940	135530	20.0
Fric. Mat. #8	46770	47220	1.0	136190	137680	1.0
			avg: 14.2			avg: 7.4

and one for the high pressure region from 2 MPa to 10.4 MPa. The compression moduli data presented in Table 3 indicate that, in general, friction materials showed higher rigidity in Fluid-PS than in Fluid-M. The low and high pressure moduli were increased 14.2% and 7.4%, respectively, in Fluid-PS as compared to Fluid-M for all the materials tested. The compression set was also affected by the kind of fluid. In general, friction materials showed less compression set in Fluid-PS than in Fluid-M.

Friction materials with a range of compressibility characteristics were tested for μ-P-v-T characteristics in a SAE No.2 Test Machine. Low and high temperature frictional torque curves for two fluids and two of eight friction materials tested, are shown in Figures 15 and 16. Torque curves of Friction Materials #1 and #8 represent compressible and rigid friction materials, respectively. Each curve was of the last cycle of level C in the μ-P-v-T test procedure (Table 1). The other levels of the test procedure showed similar trends. At level C, the speed of sliding decreased from 2600 rpm to 0 rpm in about 1 second during an engagement, with an applied pressure of 760 kPa. For a smooth engagement, descending torque curve shape with descending speed is desirable. The ascending torque curve shape towards the end of an engagement is commonly referred to as a rooster tail effect. When the coefficient of friction increases with decreasing speed, stick-slip conditions occur between friction plates and separator plates, which may result in vibrational shudder. Hence, the rooster tail effect is not desirable. In μ-P-v-T tests, Fluid-M and Fluid-PS had similar torque curve shapes at low temperatures: a flat or slightly descending initial torque shape with a slight

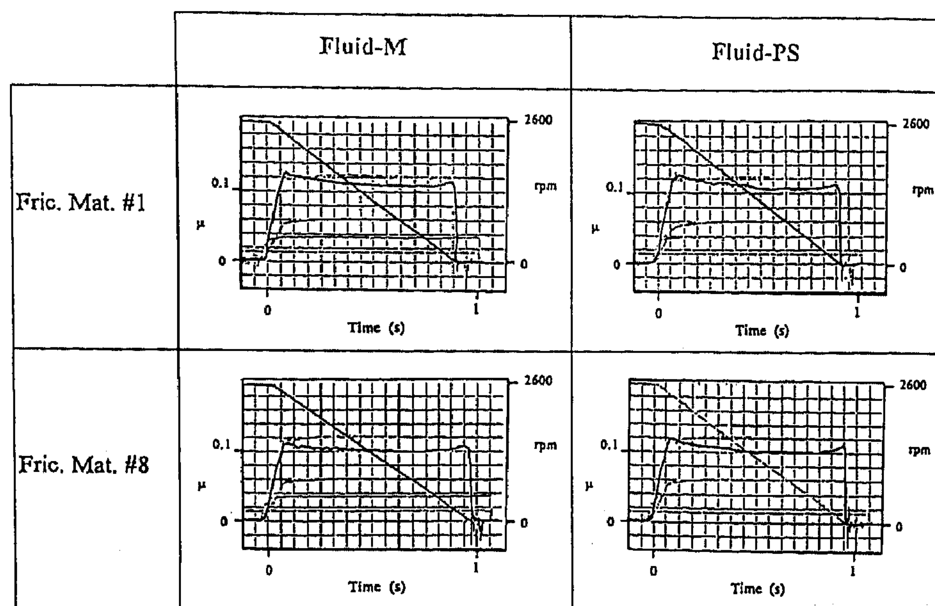


Figure 15. Torque curves of a compressible material (Fric. Mat. #1) and a rigid material (Fric. Mat. #8) in Fluid-M and Fluid-PS at 30°C.

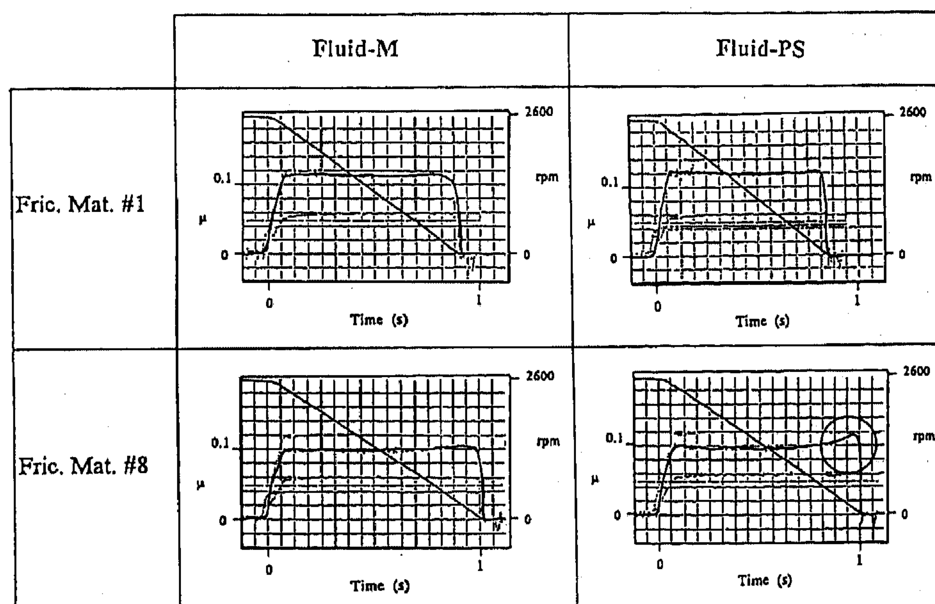


Figure 16. Torque curves of a compressible material (Fric. Mat. #1) and a rigid material (Fric. Mat. #8) in Fluid-M and Fluid-PS at 120°C. The rooster tail effect in Fluid-PS is circled.

rooster tail near the end of engagement (Figure 15). At high temperatures such as 120°C, however, Fluid-PS caused large rooster-tails especially with rigid friction materials such as Fric. Mat. #8, while Fluid-M caused small or no rooster tails (Figure 16). The least negative μ–v relationship (i.e., the smallest rooster tail) in Fluid-PS was obtained with a low modulus (i.e. highly compressible) friction material such as Fric. Mat. #1. The low modulus materials such as Fric. Mat. #1, generated higher friction coefficients than the high modulus materials such as Fric. Mat. #8, in both fluids. High friction coefficients during a clutch engagement are desirable because high coefficients provide high torque transmission capability.

Friction materials with various degrees of compressibility were tested for high-energy-high-pressure durability in Fluid-M and Fluid-PS in an SAE No.2 Test Machine. The durability test conditions are summarized in Table 4. The applied pressure on the friction plates was 1295 kPa. The failure criterion was a 2 mm displacement of the pressure apply piston. At least ten different friction material compositions were subjected to durability tests in both fluids. The number of

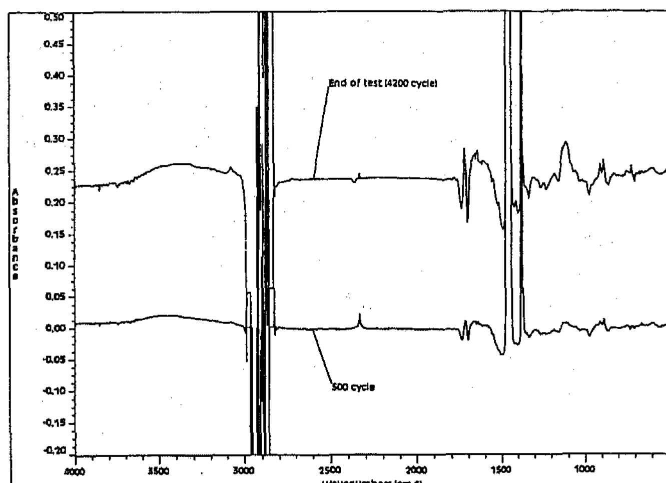


Figure 17. FTIR subtraction spectra: (Residual Fluid-M) - (Fresh Fluid-M). The samples were taken at the cycle 500 and at the end of test.

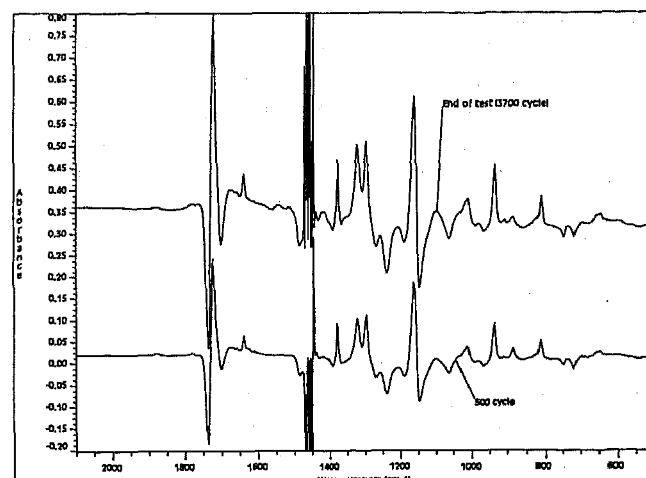


Figure 18. FTIR subtraction spectra: (Residual Fluid-PS) - (Fresh Fluid-PS). The samples were taken at the cycle 500 and at the end of test.

Table 4. SAE No.2 Test Machine durability test procedure

Number of plates in pack	2 FP, 3 SP
Facing OD/ID (mm)	126/106
Oil temperature (°C)	100-110
Oil amount (l)	1
RPM	3600
Inertia (Nms^2)	0.52
Kinetic energy (J)	36946
Pressure (kPa)	1295
Cycle time (s)	1.44
Power density (W/mm^2)	1.66
Failure criterion:	2 mm displacement
Number of cycles	5100, or to failure

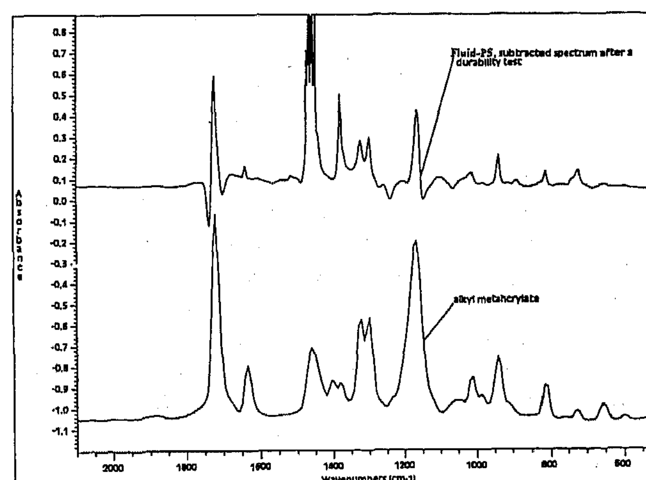


Figure 19. The differential infrared spectrum of a residual Fluid-PS matches the infrared spectrum of an alkylmethacrylate.

cycles to failure ranged from 300 to 5100 depending on the friction material-fluid combinations. Since the number of cycles to failure depends on the friction material-ATF combination, fluid samples were not only collected at the end, but also at 500 cycles of each durability test, in order to do a one-to-one comparison of the fluid degradation. In general, the friction materials with higher compressibility lasted longer in durability tests in both fluids. Another important observation was that each fluid showed different but consistent degradation characteristics, regardless of the type of friction material used in the durability tests. As examples of the fluid degradation characteristics, Figures 17 and 18 show the differential FTIR spectra of Fluid-M and Fluid-PS after the durability testing. An unspecified commercial material used in the tests shown in Figures 17 and 18 lasted to comparable cycles in both fluids: 4200 cycles in Fluid-M and 3700 cycles in Fluid-PS. The subtraction spectra between the tested and fresh fluids indicated that the degradation of Fluid-M is negligible during the SAE tests, however the degradation increases with the number of engagement cycles as evidenced in differential FTIR spectra between 500 cycles and 4200 cycles in Figure 17. The subtraction spectra of tested Fluid-PS indicated an increase in the intensity of peaks located at 1730 cm^{-1}, 1647 cm^{-1}, 1300 cm^{-1}, 1172 cm^{-1}, 1020 cm^{-1}, 944 cm^{-1}, 893 cm^{-1} and 817 cm^{-1} (Figure 18). A library search on these bands resulted in a 85% goodness of fit with the bands of polyalkylmethacrylates (PAMA) which is the VI improver additive (Figure 19). The GPC chromatogram of the tested fluids showed that the VI improver molecules in Fluid-PS were considerably fragmented during the durability tests. Some fragments may oxidize. The fragmentation of the PAMA molecules followed by an oxidation of some of the fragments causes a small shift in the infrared peak locations and results in incomplete subtraction of PAMA peaks in the differential FTIR spectrum of Fluid-PS. The fragmentation of PAMA molecules increased with the increasing number of

engagements in Fluid-PS (Figure 18). Since the fragmentation of VI improver happened with any kind of friction material tested, the cause of fragmentation was physical, namely the mechanical shear rather than a chemical effect. Since the VI improver concentration is considerably higher in Fluid-PS than in Fluid-M, the fragmentation of VI improver was more significant in Fluid-PS than in Fluid-M.

DISCUSSION

The majority of wet clutch friction materials used in automotive applications today is polymeric composite materials saturated with thermosetting resins. Polymeric friction materials are developed with a fundamental understanding of material-friction characteristics and basic ATF-friction material interactions. In order to compare the wet clutch performance of a mineral-based ATF and a partial-synthetic-based ATF, some important characteristics of the fluids were first identified. Then, the fluids were subjected to a series of mechanical tests with various kinds of friction materials. The following paragraphs discuss the important performance differences between Fluid-M and Fluid-PS.

Fluid-PS has higher viscosity index and high-temperature-high-shear viscosity than Fluid-M. Fluid-PS also has much lower viscosity at sub-zero temperatures. High fluidity at low temperatures reduces drag loss in disengaged clutches at sub-zero temperatures, hence, contributing to fuel economy. A high concentration of VI improver, in addition to the 22.4% synthetic base fluid, is responsible for the superior viscosity characteristics of Fluid-PS. Superior viscosity characteristics contribute to a low level of metal-to-metal wear because a high viscosity index and high shear viscosity provide better fluid film formation capabilities for an extended range of temperatures in elastohydrodynamic conditions. When the sliding surfaces can not be separated by a lubricant film due to severe operating conditions such as low speed, high pressure, and high temperature, the asperity interaction is inevitable. When there is appreciable asperity contact, as in the case of the 4-Ball tests, the anti-wear additives react with the exposed nascent metal surfaces and form protective surface films. The protective ability of surface films formed from a phosphate-type anti-wear additive in a boundary lubrication regime, was documented earlier (6,8). Less metallic wear in Fluid-PS as compared to Fluid-M may be an indication of the relative effectiveness of phosphate and phosphonate type antiwear additives, which are used in Fluid-PS and Fluid-M respectively.

High concentrations of calcium and sulfur in Fluid-M suggest that there is probably calcium sulfonate in Fluid-M, added as a detergent. Having a basic character, calcium sulfonate also acts as a buffer against increasing the acidity of degrading oil. Furthermore, it follows from the thermal oxidation tests that in the presence of calcium sulfonate, there is a tendency to polymerize the degradation products to diester and to form sludge. The polymerization and sludge formation is not desirable because they may plug the pores of the friction material. Other evidence showed that SO_3 may form from metal sulfonates upon degradation, and be adsorbed on the sliding surfaces, thereby, reducing coefficient of friction (9). On the other hand, the formation of ketones, acids and esters in Fluid-PS may affect the fiber-resin bond strength in friction materials in the long term. The long-term effects of fluid's degradation products need further investigation. Fragmentation of polyalkylmethacrylate molecules in Fluid-PS is not desirable. The fragmentation of VI improver not only changes the fluid's viscosity characteristics but may also cause formation of alkylmethacrylate free radicals, which help to initiate degradation of the fluid. Furthermore, if the alkyl methacrylate fragments oxidize, they may no longer be chemically inert, but interact with the friction material and with the additives of the fluid.

It is the compressibility of a wet clutch system which helps absorb vibrations and allow weeping of ATF through pores during engagements. A wet clutch system with Fluid-PS behaves more rigidly than a system with Fluid-M. One explanation for this behavior lies both with the branched structure of the polyalphaolefin molecules, and with the differences in concentration and molecular weight of the VI improvers of Fluid-PS and Fluid-M. It was reported that the VI improver molecules may orient themselves under pressure if the molecules are longer than a threshold value. The onset of molecular orientation lies in the range of 10000 to 100000 dalton for most fluids (10). Having long molecular chains heavier than 100000 dalton, the VI improver molecules of Fluid-M may reversibly orient themselves under pressure, thereby, contributing to the compressibility of the friction material-ATF system. On the other hand, having a large amount of shorter chain molecules, the VI improver of the Fluid-PS shows stronger resistance to orientation under pressure, hence, increasing the rigidity of the system. Furthermore, it will often be more difficult for the branched PAO molecules of the partial-synthetic fluid to get in and out of the pores of the friction material than for the linear molecules of the mineral fluid. The fluid penetration becomes more difficult when there is a large concentration of VI improver in the fluid, than when there is not. At this moment, it is not known whether the rigidity is also affected by the rest of the additives of the fluid or not.

The torque curve shape in an engagement is affected by the type of base oil/VI improver combination, and/or by the type and ratios of friction modifier, dispersant, detergent, and anti-wear additives. Since the type and ratio of dispersant additive present in the Fluid-M and Fluid-PS are very similar, the cause of undesirable torque curve shape with Fluid-PS at high temperatures, can be explained by the narrow molecular weight distribution of the partial-synthetic base oil, higher concentration of VI improver, absence of calcium

sulfonate detergent additive, presence of inorganic phosphate additive and type of friction modifier which is not known. During an engagement, the interfacial temperatures easily reach several hundred degrees centigrade. Furthermore, the interfacial temperatures keep increasing from the start of the engagement to the end of engagement. Remember that the rate of evaporation of Fluid-PS is twice as high as that of Fluid-M in the temperature range of 180°C and 210 °C. The rapid evaporation of partial-synthetic base oil with increasing interfacial temperature leaves behind a highly inhomogeneous, high molecular weight fluid, which may cause the higher friction near zero sliding speed. Secondly, Scanning Electron Microscope/Electron Diffraction X-ray analysis of the friction material surfaces indicated the presence of a considerable amount of Ca and S on the surfaces after durability tests in Fluid-M, but not in Fluid-PS. Furthermore, phosphorus concentration on friction material surfaces run in Fluid-PS, was slightly higher than that in Fluid-M. Obviously, the surface species adsorbed from the oil additives, which may affect the torque curve shapes, have a different composition in Fluid-M and in Fluid-PS. The reduction in rooster tail with a compressible friction material may be explained by the weeping of the cooler ATF from the pores of friction material to the interface under compression. Furthermore, a compressible material dampens the shudder effectively, resulting in a smaller rooster tail. The high rigidity in Fluid-PS can be accommodated by designing more compressible friction materials.

CONCLUSIONS

The interactions of a mineral-based ATF and a partial-synthetic based ATF with various kinds of friction materials, were investigated. Friction material/ATF systems showed marked performance differences in mechanical and thermal-oxidation tests. The compressibility and torque response curve were greatly affected by the type of fluid and the type of material. Partial-synthetic fluid has many superior characteristics over mineral fluid such as good metallic wear prevention, excellent low temperature fluidity, and high viscosity index. These superior characteristics were achieved by radically different molecular weight properties and the additive package of the partial-synthetic fluid from those of the commercial mineral-based ATF. Partial-synthetic fluid may increase the performance of those components of an automatic transmission which operate in the elastohydronamic lubrication regime. However, this study has shown that the partial synthetic fluid which also contains a high concentration of VI improver and a detergent-free additive chemistry, requires a careful selection of friction material for wet clutch applications. The major findings of this study are:

1. The partial-synthetic fluid caused less friction and wear than mineral fluid did in a lubrication regime where there is appreciable metal-to-metal contacts because of its superior viscometrics and anti-wear additive characteristics.

2. The viscosity index improver molecules of the partial-synthetic fluid were fragmented during the wet clutch engagements. The thermal oxidation tests indicated acid, ketone, and ester formation in both fluids. However, in Fluid-M the degradation products polymerize to diester or polyester. The different degradation paths of the Fluid-M and Fluid-PS have different implications on the wet clutch performance.

3. Friction materials showed higher rigidity in the partial-synthetic fluid than in the mineral fluid in compression tests. The high rigidity may be caused by the branched structure of the synthetic molecules, a high concentration of VI improver, and the detergent-free additive chemistry of the partial-synthetic fluid.

4. The partial-synthetic fluid caused a much more negative friction-speed relationship near the end of an engagement (the rooster tail effect) than the mineral fluid did at high temperatures. The torque curve shape may be affected by both the detergent free-additive chemistry, and by the base oil/VI improver combination of the partial synthetic fluid. The rooster tail effect of partial-synthetic fluid was reduced considerably with a compressible friction material.

REFERENCES

(1) Ford Engineering Material Specification Nr. M2C202-B, 1995.

(2) Ford Engineering Material Specification Nr. M2C166-H, 1975.

(3) D. McFall, "Will ATF Requirements Mesh or Clash?" Lubes-n-Greases, 2, 1, 13-16, 1996.

(4) Official Method Cd 6-38 in "Official and Tentative Methods of American Oil Chemists Society", published by A.O.C.S., (1938).

(5) Y. Yang, R.C. Lam, Y.F. Chen, H. Yabe, "Modeling of Heat Transfer and Fluid Hydrodynamics for a Multidisc Wet Clutch," SAE Paper Nr. 950898, (1995).

(6) B. Çavdar, K.C Ludema, "Dynamics of Dual Film Formation in Boundary Lubrication of Steels, Part I. Functional Nature and Mechanical Properties", Wear, 148, 305-327, (1991).

(7) B. Çavdar, S.K. Sharma, L.J. Gschwender, "Wear-Reducing Surface Films Formed by a Fluorinated Sulfonamide Additive in a Chlorotrifluoroethylene-based Fluid", Lubrication Engineering, 50, 11, 895-902, (1994).

(8) B. Çavdar, K.C Ludema, "Dynamics of Dual Film Formation in Boundary Lubrication of Steels, Part II. Chemical Analyses", Wear, 148, 329-346, (1991).

(9) T. Miyazaki, R. Toya, T. Matsumoto, "The Effect of Additives in ATF on the Friction Characteristics of a Paper-Based Facing," Pr°C. of Int. Trib. Conf., Yokohama, Paper Nr. 313-1, (1995).

(10) J.W. Sprys, D.R. Vaught, E.L. Stephens, "Shear Viscosities of Automatic Transmission Fluids," SAE Paper Nr. 941885, (1994).

970977

Influence of Paper-Based Friction Material Visco-Elasticity on the Performance of a Wet Clutch

Takayuki Matsumoto
NSK-Warner K.K.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Friction performance of paper-based materials for wet clutches in automotive automatic transmissions is influenced by the entire structure of the material, which is characterized by a combination of porosity and visco-elasticity. In the clutch plate engagement process, the coefficient of friction is higher when the porosity of the material is higher.

In this paper, visco-elasticity of the material was measured by a testing apparatus which was originally designed and built. It has become clear that the compressive visco-elastic deformation increases as the porosity becomes higher.

The relationship between the compressive visco-elastic deformation of the material and the friction torque curve of the clutch plate was also analyzed. It was found that the friction torque curve of a clutch plate is influenced by the visco-elasticity of the material as well as the porosity.

1. Introduction

The friction characteristics and thermal resistance of paper-based wet friction materials are influenced by the material structure formed as a result of the combination of porosity and visco-elasticity (resiliency).

In three previous papers (1), (2), (3) by the author, the effects of porosity on the frictional performance and the thermal resistance were characterized by Stribeck Diagrams based on continuous sliding test data, and were also discussed with reference to torque curves and temperature measurements by an SAE No. 2 friction testing machine. The effects of porosity on the thermal durability of paper-based friction materials were also studied, by discussing material durability in terms of mechanical and physical property changes of the material.

In this study, the effects of visco-elasticity (resiliency) on the friction characteristics of paper-based friction materials were studied. The relationship between the compressive visco-elastic deformation of a paper-based friction material and the friction torque curve of a clutch plate from the initial to the final stage of the engagement process was analyzed.

2. Concept of Visco-Elasticity in the Engagement Process of a Wet Clutch

The entire structure of a paper-based friction material, which is characterized by a combination of porosity and visco-elasticity, has a strong influence on the tribological property existing between the sliding surfaces of the friction material and the mating steel plate. Visco-elasticity and porosity are considered to be the two important factors that control the friction performance of the material. Fig. 1 shows the concept of visco-elasticity and porosity of a paper-based friction material. Assuming that the figure shows the section of a clutch plate, the porous material

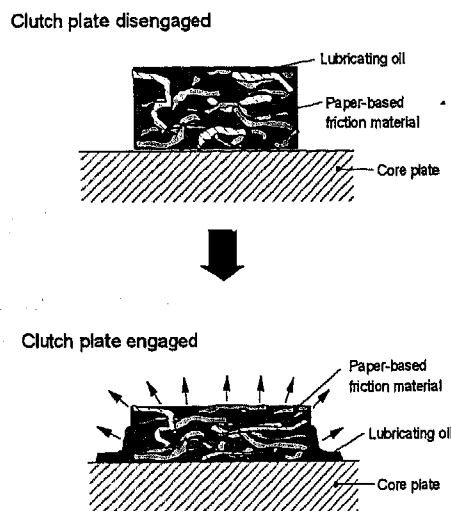


Fig. 1 Concept of Visco-Elasticity and Porosity

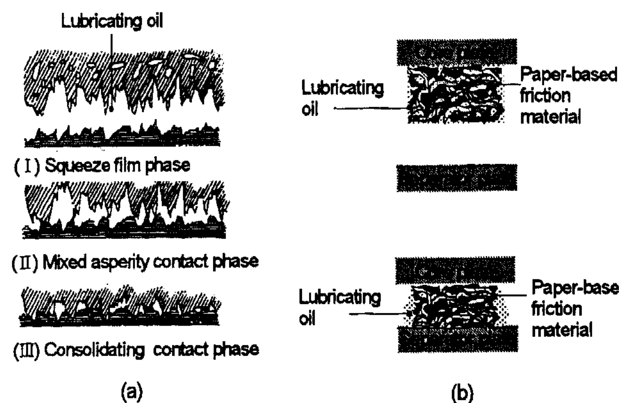


Fig. 2 Concept of Engagement Model

is filled with lubricating oil, when a clutch plate is disengaged. When a clutch plate is engaged, the porous material is visco-elastically deflected (compressive deformation), and the lubricating oil is squeezed out of the material.

The behavior of the lubrication in the engagement process of a wet clutch and a wet brake is considered to be divided into three phases, that is, squeeze film phase, mixed asperity contact phase and consolidating contact phase, as shown in Fig. 2 (a) (4), (5). The porous material becomes filled with lubricating oil, when the clutch is disengaged. When the clutch is engaged, the porous material becomes visco-elastically deflected, and the lubricating oil is squeezed out of the porous material, as shown in Fig. 2 (b).

3. Experiment

3.1 Measuring Apparatus of Compressive Visco-Elastic Deformation

In order to investigate the visco-elastic behavior at the surface of paper-based friction materials during a practical clutch plate engagement, a special measuring apparatus was designed and built, as shown in Fig. 3. Fig 3 (a) shows the schematic outline of the apparatus. The test piece is installed on the spline of the oil cup. The piston and the test piece, which is bonded a paper-based friction material, can

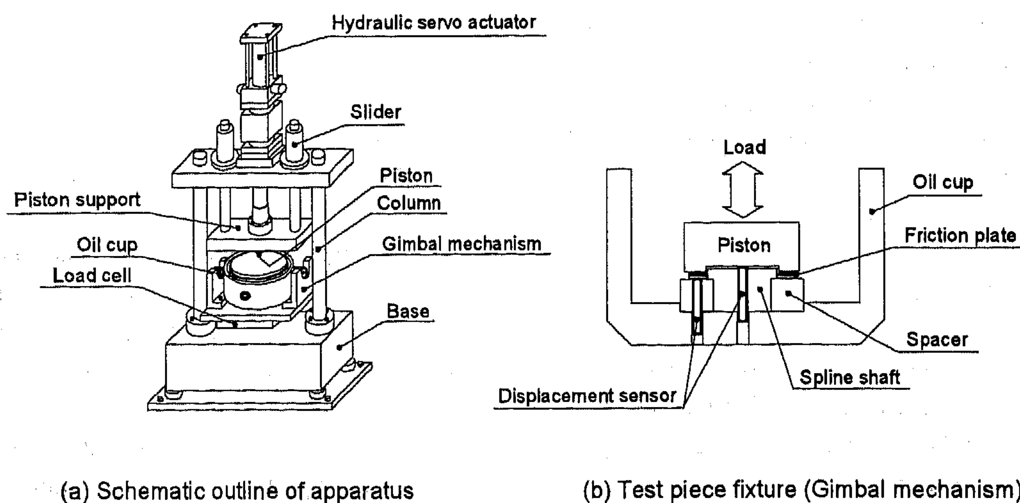
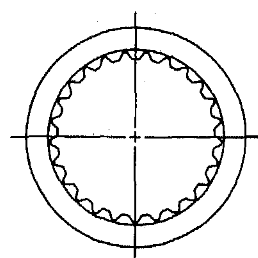


Fig. 3 Measuring Apparatus of Compressive Visco-Elastic Deformation

be made to fully contact each other by a Gimbal mechanism without any unbalance, as shown in Fig. 3 (b). The test piece is a practical clutch plate as shown in Fig. 4. It is a bonded paper-based friction material on one side only. The load is applied by hydraulic servo actuator, and is measured by a load cell equipped under the Gimbal mechanism. Compressive visco-elastic deformation is measured by displacement sensors installed in the oil cup.

In order to eliminate the influence of core plate flatness on the deformation measurement, the displacements between core plate and spacer are measured at and averaged from three places on the circumference of the plate.

The entire deformation is measured by the displacement sensor located at the center of piston. The deformation of a paper-based friction material is computed by subtracting the displacement between core plate and spacer from the entire deformation.



Non - groove

(Facing dimensions)
Outside diameter : 70mm
Inside diameter : 56mm
Thickness : 0.5mm

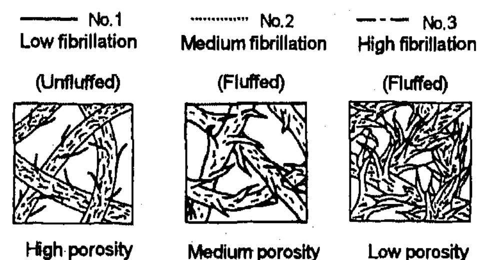
Fig. 4 Friction plate

3.2 Relationship between Porosity and Compressive Visco-Elastic Deformation

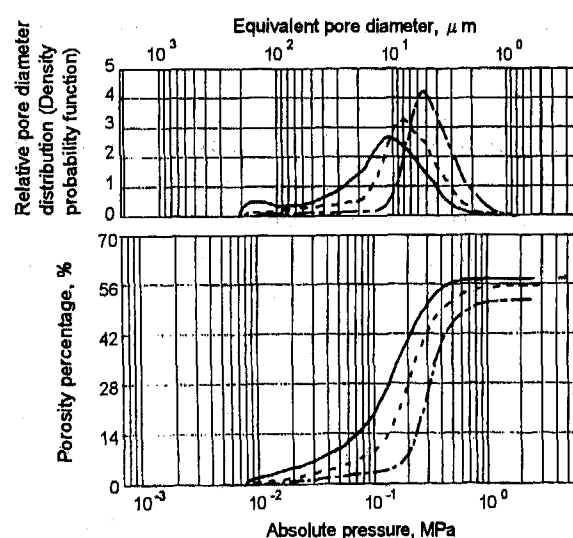
As for the test results based on friction torque curves obtained by an SAE No. 2 friction testing machine using the sample paper-based friction materials (whose porosities were varied), it was found that the coefficient of friction is higher when the porosity of a material is higher (1), (2), (3).

In order to vary only the porosity while keeping the ingredients almost constant in the paper-based friction material, the fibers of the sample materials were supplied with different degrees of fibrillation treatment. Fibrillation is a refining process used in manufacturing paper that makes fibers fluffy. Fig. 5 (a) shows a picture of fluffed fibers (fibrillation) and unfluffed fibers.

The porosities measurement results of materials with the same ingredients are shown in Fig. 5 (b). The pore diameter, pore diameter distribution and porosity percentage were measured by a mercury intrusion porosimetry method (6). Using the porosimeter, mercury penetrates the lining piece as the pressure is increased. As the volume of penetrated mercury is directly read, porosity percentage at each pressure can be continuously plotted, if the volume of a sample is known. The lower graph in Fig. 5 (b) shows the relationship between porosity percentage and absolute pressure (MPa). The upper graph in Fig. 5 (b) is the derivative curve of the lower graph, from which the equivalent pore diameter (μm) and the



(a) Structure of material



(b) Pore diameter and porosity

Fig. 5 Porosity Measurement Result of Sample Material

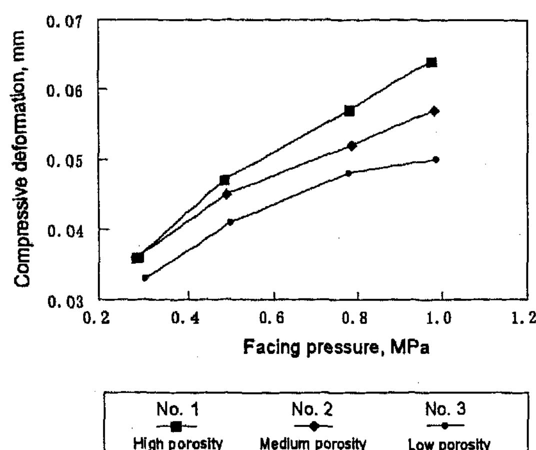


Fig. 6 Relationship between Facing Pressure and Compressive Deformation

relative pore diameter distribution can be determined. The pore diameter as well as the porosity percentage of sample material is arranged in a descending order: No. 1 > No. 2 > No. 3. The sample material No. 1, which received the lowest degree of fibrillation, has the highest porosity, and the sample material No. 3 with the highest degree of fibrillation has the lowest porosity. The fibers in raw papers have a tendency to be oriented to flow in the prescribed direction during the paper manufacturing process. However, the pore size distribution and the orthotropy of the material are not changed circumferentially, because the friction material is stamped out in a round shape from the raw paper sheet.

These sample materials No. 1, No. 2 and No. 3 were bonded to practical clutch plates to make the test pieces, as described in section 3.1. Compressive visco-elastic deformation of each sample material was measured while immersed in ATF (Dexron II Type) by the measuring apparatus shown in Fig. 3.

The relationship between facing pressure and compressive deformation for each sample material is shown in Fig. 6. It is presumed that the compressive deformation in Fig. 6 includes permanent (plastic) deformation, to some extent. However, since the facing pressure is not higher than 1.0 MPa, it can be said that the compressive deformation is mostly elastic. The compressive deformation of a paper-based friction material increases as the porosity of the material becomes higher. It is clear that increased deformation functions effectively to raise the coefficient of friction, which will be discussed later.

3.3 Experimental Procedure

The visco-elasticity of a paper-based friction material is dependent on the basic ingredients in the raw paper and the kind of resins. In order to vary the visco-elasticity while keeping the ingredients almost constant in a paper-based friction material, the sample materials were made by changing the compression rate during the clutch plate bonding process and by changing the volume percentage of resin. As the same kind of resin was used, the ingredients of the sample materials are identical. However, the ingredient percentage of the sample materials are different because the volume percentage of resin was varied.

Five kinds of sample materials, whose visco-elasticity are different, were made as shown in Table 1. As for the volume percentage of resin, Low (a) is about 15%, Medium (b) is about 30%, and High (c) is about 40%. Porosity was measured by a mercury intrusion porosimetry method. The average pore diameter of a porous body is determined by Eq. (1), using the data measured by a mercury intrusion porosimetry method (6).

$$D_{ave} = \frac{2V_{max}}{S_{max}} \quad (1)$$

where:

D_{ave} : Average Pore Diameter (μ m)

V_{max} : Total Micropore Volume (cm^3/g)

S_{max} : Total Micropore Specific Surface Area (m^2/g)

V_{max} is the total volume of the penetrated mercury, and S_{max} is the total specific surface area of the penetrated mercury.

The apparent density is determined by Eq. (2), and the real density is computed by Eq. (3).

$$\rho_a = \frac{W}{V} \quad (2)$$

where:

ρ_a : Apparent Density (g/cm^3)

W : Measured Sample Material Weight (g)

V : Measured Sample Material Volume (cm^3)

Table 1 Properties of Sample Materials

Volume percentage of resin		Compression rate used in the bonding process		
		Low (A)	Medium (B)	High (C)
Low (a)	Sample No.	Aa		Ca
	Apparent density (g/cm^3)	0.592		0.834
	Average pore diameter (μm)	4.29		2.20
	Maximum existing pore diameter (μm)	11.06		3.72
	Porosity percentage (%)	58.3		43.9
	Real density (g/cm^3)	1.420		1.487
Medium (b)	Sample No.		Bb	
	Apparent density (g/cm^3)		0.853	
	Average pore diameter (μm)		4.16	
	Maximum existing pore diameter (μm)		6.52	
	Porosity percentage (%)		41.8	
	Real density (g/cm^3)		1.466	
High (c)	Sample No.	Ac		Cc
	Apparent density (g/cm^3)	0.864		1.151
	Average pore diameter (μm)	7.73		3.17
	Maximum existing pore diameter (μm)	11.09		3.61
	Porosity percentage (%)	39.0		20.2
	Real density (g/cm^3)	1.416		1.442

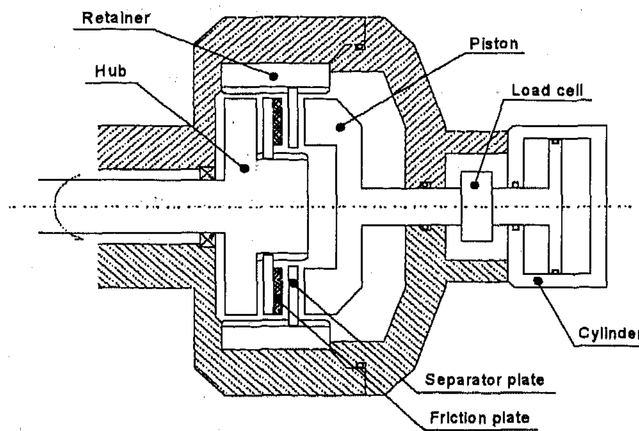


Fig. 7 Test Piece Fixture of a Modified SAE No. 2 Friction Testing Machine

$$\rho_r = \frac{\rho_a}{1 - PP/100} \qquad (3)$$

where:

ρ_r : Real Density (g/cm^3)

PP : Porosity Percentage (%)

Compressive visco-elastic deformation of each sample material is measured by the method as described in section 3.1. Friction torque curve of each sample material is obtained by a modified SAE No. 2 friction testing machine, of which the schematic outline of the test piece fixture is shown in Fig. 7. Standard SAE No. 2 friction testing machine was modified so that the friction torque curve of the 70 mm outside diameter plate bonded a paper-based friction material on one side only, could be obtained. Therefore, the friction torque curve of the very plate, whose compressive visco-elastic deformation was measured, could be obtained by this modified SAE No. 2 friction testing machine. Test condition is shown in Table 2. Friction torque curve is measured after 200 engagement cycles.

Table 2 Test Condition by a Modified SAE No. 2 Friction Testing Machine

Rotating speed	2380 rpm
Inertia	0.02107 kg · m^2
Facing pressure	0.5 MPa, 0.8 MPa
Lubricating oil	ATF (Dexron II Type)
Oil temperature	100 ℃
Oil volume	0.0008 m^3 (0.8 L)
Total energy absorbed	0.6 kJ
Number of friction surface	1

4. Results and Discussion

4.1 Compressive Visco-Elastic Deformation

As described in section 3.2, three identical quality materials (identical ingredients, identical ingredient percentage, but with different porosities) were made, in order to study the influence of paper-based friction material porosity on the performance of a wet clutch (1), (2), (3).

As for the sample paper-based friction material whose porosities were varied, the relationship between compressive visco-elastic deformation and the facing pressure received is shown in Fig. 6. The data were obtained by the sample materials No. 1, No. 2 and No. 3 whose porosities are different. The porosity measurement result is shown in Fig. 5. It can be understood that the compressive deformation of a paper-based friction material increases as the porosity

of the material becomes higher.

Visco-elasticity of a paper-based friction material varies with porosity change. Porosity is a comprehensive term that includes three aspects: porosity percentage, pore diameter, and pore diameter distribution. At present, there is no known method to independently change the visco-elasticity of a paper-based friction material without either changing the porosity or the material composition.

Therefore, the sample materials whose properties are shown in Table 1, were made, by the method described in section 3.3, with a view to making the sample materials having different visco-elasticity property but retaining the porosity percentage approximately constant for Ac, Bb and Ca (but the pore diameters are different). Compressive visco-elastic deformation of those sample materials was measured while immersed in ATF (Dexron II type) by the measuring apparatus shown in Fig. 3. The relationship between facing pressure and compressive deformation for each sample material is shown in Fig. 8. By the use of two test pieces each, the compressive deformation was measured, changing the measurement position three times randomly in the circumferential direction. Finally, the average value of each set of six measurements was plotted.

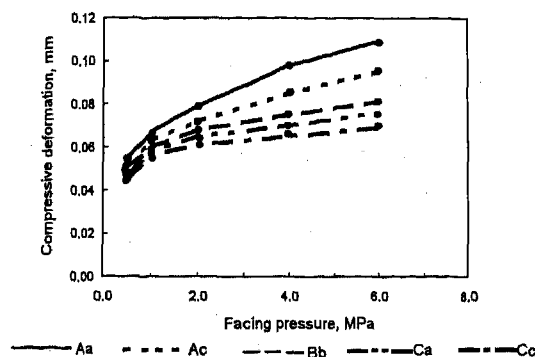


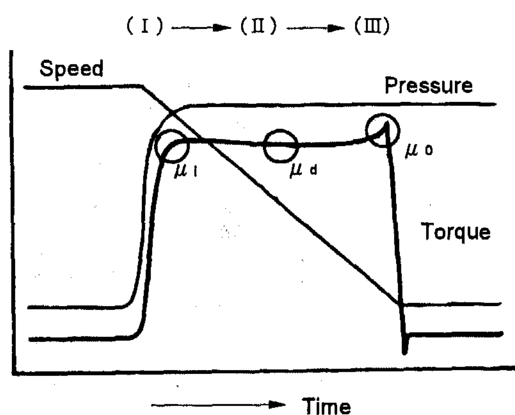
Fig. 8 Relationship between Facing Pressure and Compressive Deformation of Each Sample Material

The compressive visco-elastic deformation of the sample material is arranged in descending order: Aa > Ac > Bb > Ca > Cc. The compressive visco-elastic deformation is a function of density (Apparent density, Real density) and porosity (Pore diameter, Pore diameter distribution, Porosity percentage), provided the ingredients of the paper-based friction material are identical. A more detailed study, however, is necessary to clarify this relationship.

4.2 Friction Torque Curves

The behavior of the lubrication during the engagement process of a wet clutch and a wet brake is discussed based on the concept of three phases, that is, squeeze film phase, mixed asperity contact phase, and consolidating contact phase (4), (5). Moreover, when torque curves obtained by an SAE No. 2 friction testing machine are discussed, three modes of friction coefficient are used, namely, the coefficient of friction near the initial stage of the engagement process: μ_i, the coefficient of friction during sliding: μ_d, and the coefficient of friction in the final stage of the engagement process: μ_0, as shown in Fig. 9 (1), (2).

A comparison of torque curves due to porosity change (sample materials No. 1, No. 2 and No. 3) at the facing pressures of 0.5 MPa and 0.8 MPa is shown in Fig. 10. The sample friction plate is a non-grooved plate having splines on the inner diameter. The outside diameter, inside diameter and thickness of the facing are 127 mm, 102 mm and 0.5 mm, respectively. Test conditions are shown in Table 3. Coefficient of friction was measured after 200 engagement cycles. The coefficients of friction (μ_i and μ_d) are higher when the porosity (the porosity percentage and pore diameter) of a material is higher, and as a result, torque curves are held at higher levels. The same results were obtained at the facing pressures of 0.3 MPa and 1.0 MPa (1), (2). The higher porosity material removes the oil film more effectively in the sliding area where boundary lubrication and hydrodynamic lubrication exist at the same time (mixed-lubrication), and it increases boundary friction, thus resulting in a rise of the coefficient of friction (1), (2), (4), (5), (7), (8), (9).



μ_I : Initial dynamic coefficient of friction
μ_d : Medial dynamic coefficient of friction
μ_o : Final dynamic coefficient of friction

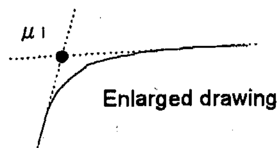


Fig. 9 Definition of μ_I, μ_d and μ_o

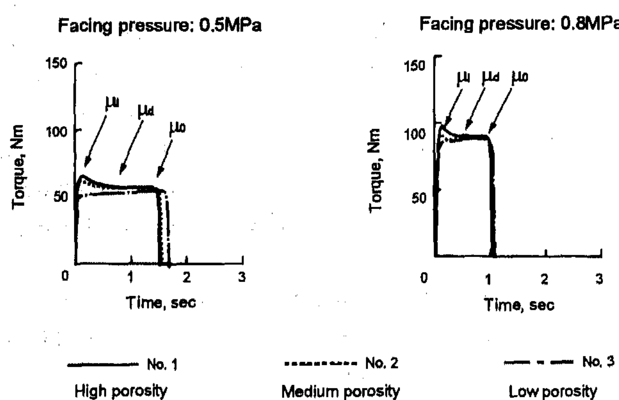


Fig. 10 Torque Curves of Friction Materials with Different Porosity

As described in section 3.2 and 4.1, the compressive visco-elastic deformation of a paper-based friction material increases as the porosity of the material becomes higher. It is clear that this deformation increase functions effectively so as to remove the oil film between the sliding surfaces.

When a wet clutch or a wet brake is engaged, a paper-based friction material actually receives a visco-elastic deformation in the shearing direction as

Table 3 Test Conditions

Tester	SAE No. 2 Friction testing machine
Rotating speed	3478 rpm
Inertia	0.245 kg · m^2
Facing pressure	0.5 MPa, 0.8 MPa
Lubricating oil	ATF (Dexron II Type)
Oil temperature	100 ℃
Oil volume	0.0007 m^3 (0.7 L)
Total energy absorbed	16.26 kJ
Number of friction surface	4

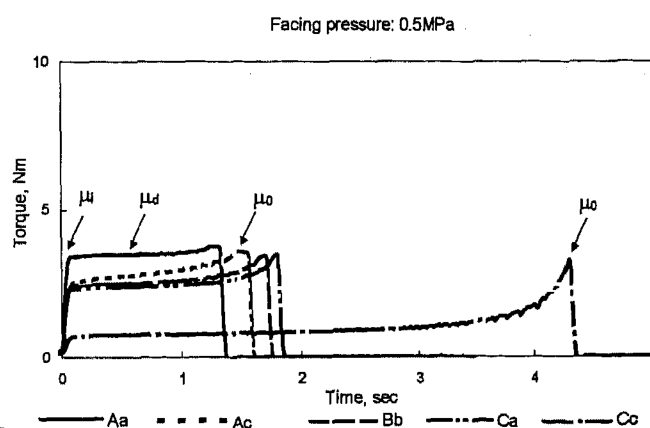


Fig. 11 Torque Curves of Friction Materials with Different Visco-Elasticity - 1

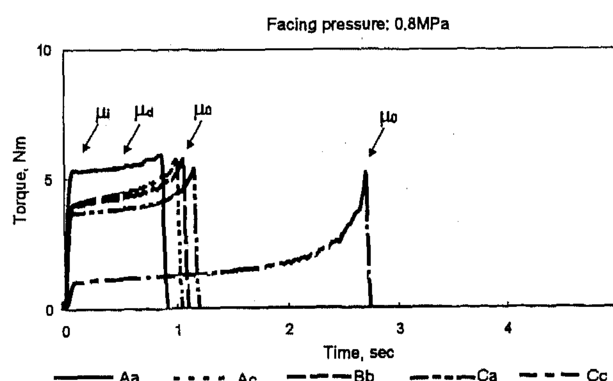


Fig. 12 Torque Curves of Friction Materials with Different Visco-Elasticity - 2

well as undergoes the compressive visco-elastic deformation. In this study, we paid close attention to the compressive visco-elastic deformation.

Friction torque curves of the sample materials, Aa, Ac, Bb, Ca and Cc were obtained by a modified SAE

No. 2 friction testing machine, as described in section 3.3. Comparisons of torque curves due to compressive visco-elastic deformation change at the facing pressures of 0.5 MPa and 0.8 MPa are shown in Fig. 11 and Fig. 12. The coefficients of friction (μ_i and μ_d) are higher when the compressive visco-elastic deformation of a material is higher, and as a result, torque curves are held at higher levels. It can be understood that the material whose compressive visco-elastic deformation is higher, removes the oil film more effectively in the sliding area where boundary and hydrodynamic lubrication co-exist and it increases boundary friction, thus resulting in a rise of the coefficient of friction. Fig. 13 shows the real contact area of the sample materials, Aa, Ac, Bb, Ca and Cc. After 200 engagement cycles, the real contact area of each sample was observed at the facing pressure of 0.8 MPa by the special testing machine (10), shown in Fig. 14. Fig. 13 is a photograph of the real contact area under stationary state. The real contact area does not change when sliding exists between the two surfaces (< about 0.5 m/sec) (10).

Facing pressure: 0.8 Mpa

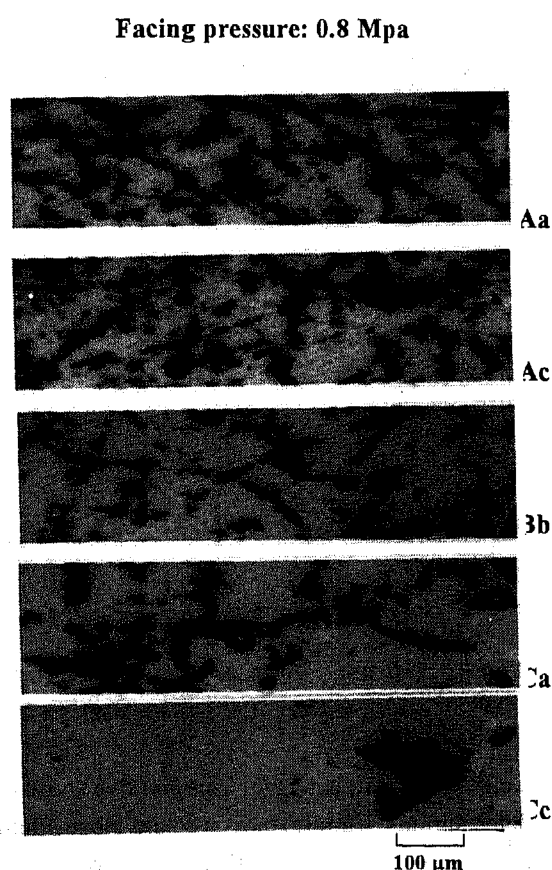


Fig. 13 Photograph of Real Contact Area

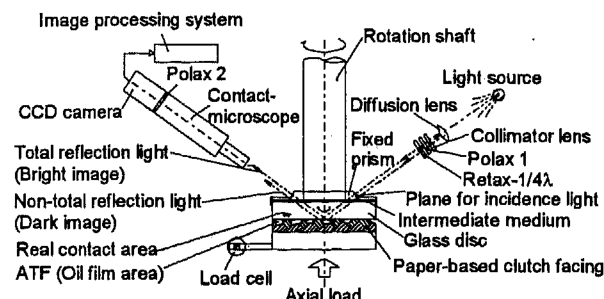


Fig. 14 Scheme of the Visualizing Apparatus

The real contact area percentage of the sample material, Aa was measured to be about 38 %, the percentage decreased in the order of Aa > Ac > Bb > Ca > Cc, and the real contact area percentage of Cc was about 16 %. The number of the real contact points decreased in the following order: Aa > Ac > Bb > Ca > Cc. Moreover, the uniformity of the contact was lost in the same order. These facts are considered to be important evidence especially in mixed asperity contact phase and consolidating contact phase, that the material having higher compressive visco-elastic deformation removes the oil film more effectively and increases boundary friction, thus resulting in a raise of the coefficient of friction.

At present, there is no known method to independently change the visco-elasticity of a paper-based friction material without either changing the porosity or the material composition. Therefore, it is difficult to determine which influence is higher, porosity or compressive visco-elastic deformation, in raising the coefficient of friction. The porosity percentages of the sample materials, Ac, Bb and Ca are at the same level, and the coefficients of friction, μ_i and μ_d of the sample materials are arranged in a descending order: Ac > Bb > Ca (The compressive visco-elastic deformation is in the same order: Ac > Bb > Ca). Therefore, it can be said that these phenomena are due to the pure influence of the visco-elasticity, but the phenomena might be due to the influence of pore size difference, because the average pore diameter and maximum existing pore diameter of the sample materials are arranged in a descending order: Ac > Bb > Ca. Further study, however, is

necessary to clarify this subject.

5. Concluding Remarks

Visco-elasticity and porosity are considered to be the two important factors that control the friction performance of a paper-based friction material. In the clutch plate engagement process, the coefficients of friction (μ_i and μ_d) are higher when the porosity of a material is higher. The higher porosity material removes the oil film more effectively in the sliding area where boundary and hydrodynamic lubrication co-exist, and it increases boundary friction, thus resulting in a rise of the coefficient of friction. It is presumed that the oil on the sliding surface penetrates into the material more effectively when the porosity of a material is higher. This fact is based on various kinds of data obtained by testing with a continuous sliding type test machine and an SAE No. 2 friction test machine.

As for the test results, the compressive visco-elastic deformation of a paper-based friction material increases as the porosity of the material becomes higher. This increase in deformation functions effectively to raise the coefficient of friction.

Using sample materials that consist of the same ingredients but different visco-elasticities, a comparison of torque curves due to compressive visco-elastic deformation change at the facing pressures of 0.5 MPa and 0.8 MPa was made. The coefficients of friction (μ_i and μ_d) are higher when the compressive visco-elastic deformation of a material is higher, and as a result, torque curves are retained at higher levels. The material whose compressive visco-elastic deformation is higher, removes the oil film more effectively in the sliding area where boundary and hydrodynamic lubrication co-exist and it increases boundary friction, thus resulting in a rise of the coefficient of friction. These phenomena were confirmed by the observation of the real contact area of the sample materials.

References

(1) T. Matsumoto, A Study of the Influence of Porosity and Resiliency of a Paper-Based Friction Material on the Friction Characteristics and Heat Resistance of the Material, SAE Transactions, Journal of Passenger Cars - Section 6, Vol. 102 (1993) p. 2417 - 2424.

(2) T. Matsumoto, The Influence of Paper-Based Friction Material Porosity on the Performance of a Wet Clutch, SAE Paper 941032 (1994).

(3) T. Matsumoto, A Study of the Durability of a Paper-Based Friction Material Influenced by Porosity, Transactions of the ASME Journal of Tribology, Vol. 117 (1995) p. 272 - 278.

(4) A. E. Anderson, Friction and Wear of Paper Type Wet Friction Elements, SAE Paper 720521 (1972).

(5) L. L. Ting, Engagement Behavior of Lubricated Porous Annular Disks (Part 1, Part 2), Wear, Vol. 34, No. 2 (1975) p. 159 - 182.

(6) A. Kanzawa, M. Hasetani, Porous Material Handbook, IPC Inc. (1988) p. 28 - 31.

(7) H. Wu, Squeeze-Film Behavior for Porous Annular Disks, ASME Journal of Lubrication Technology, Vol. 92 (1970) p. 593 - 596.

(8) L. L. Ting, A Mathematical Analog for Determination of Porous Annular Disk Squeeze Film Behavior Including the Fluid Inertia Effect, ASME Journal of Basic Engineering, Vol. 94 (1972) p. 417 - 421.

(9) H. Ito, K. Fujimoto, M. Eguchi, T. Yamamoto, Friction Characteristics of a Paper-Based Facing for a Wet Clutch under a Variety of Sliding Conditions, STLE Tribology Transactions, Vol. 36 (1993) p. 134 - 138.

(10) Z. Q. Huang, S. Aihara, S. Umezawa, T. Matsumoto, A Study on Visualizing Contact Surface of Wet Clutch and Measurement of Real Contact Area, Proceedings of the 73rd JSME Fall Annual Meeting, Vol.4, No. 95 - 10 (1995 - 9) p. 152 - 153.

970978

Theoretical and Experimental Studies on the Thermal Degradation of Wet Friction Materials

Yubo Yang†, Paul S. Twaddell, Yih-Fang Chen, and Robert C. Lam
Borg-Warner Automotive

Copyright 1997 Society of Automotive Engineers, Inc.

Abstract

A theoretical model for predicting the life cycle of a friction material used in wet friction clutches has been developed and verified. For a given friction material, the degradation mechanism can be identified by performing a Thermal Gravimetric Analysis (TGA) on the samples of worn friction materials. The samples are taken from the friction plates after they undertake various periods of the continuous slip experiments on the full-pack test machine. The degradation rate constants are obtained by performing the TGA experiments on the samples from the continuous slip experiments with different input powers and interface temperatures. The degradation for a dynamic engagement cycle is calculated by integrating the degradation rate with the temperature history near the friction interface as a function of time. The temperature history is predicted by the Borg-Warner computer model for the engagement of a wet clutch [1], which has been verified experimentally. The predicted thermal degradations agree with the experimental measurements on the worn friction samples from the dynamic engagements. This work illustrates that the degradation of wet friction materials in a wet friction clutch can be predicted by a theoretical model using temperature history near the friction interface.

Introduction

The wet disc clutch used in automatic transmissions consists of multiple separators, friction plates, and other components as illustrated in Figure 1. The key element in the clutch pack is the friction plate, made of a low carbon steel core plate with friction materials bonded on each side. A typical clutch pack configuration as shown in Figure 1 contains five separator plates and four friction plates. The friction plates are splined to the input shaft while the separator plates are lugged to the clutch housing. The function of a clutch pack is to transmit torque from a driving to a driven member when shifting from one gear to another. During engagement (shifting) the separator plates and friction plates are forced together by pressure, and the kinetic energy is converted to heat at the interfaces. The time for the relative rotating speed to reach zero is called the stop time, which is typically less than a second.

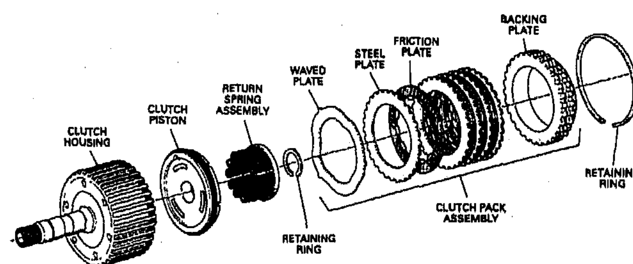


Figure 1: Schematic of Automatic Transmission Clutch Pack

During the engagement of wet clutches, a large amount of friction heat is generated at the separator-friction plate interface in a short time.

† To whom all correspondence should be addressed to.

The separator absorbs most of the heat and the friction plate (friction material and core plate) receives much less heat, due to the low conductivity of friction material. This results in a non-uniform temperature distribution in the axial direction in the friction material, i.e. large temperature gradient across the thickness of friction materials. The temperature rise at the interface during engagement is determined by the input energy, and physical and processing parameters. The temperature history near the friction interface is important, since it determines the performance characteristics, and especially, the thermal degradation rate of friction material.

The degradation of friction materials has been the subject of several investigations [2, 3, 4]. However, all the previous works to authors' knowledge were experimental in nature, and theoretical work is lacking. In order to predict the life cycle of friction materials for a given application, it is necessary to understand theoretically the mechanism and rate of material degradation. The life of a paper-based friction material is affected by both the thermal degradation (carbonization) and mechanical degradation (wear). The thermal degradation of a friction material is the major factor for determining the life cycle under the normal engagement of a wet clutch due to the high temperature environment at the friction interface, caused by the friction heat generated during clutch engagements. After the friction material is thermally degraded to a certain extent, the friction characteristics will be changed dramatically and its performance will suffer. Generally, the least heat resistant composition in the friction material will be degraded first at elevated temperatures.

The objective for this work is to predict the thermal degradation of friction materials during the dynamic engagement of wet clutches for any given condition. To achieve this goal, the mechanism of thermal degradation and its rate constants need to be determined. The temperature history near the friction interface during the dynamic engagement, as computed by the thermal model [1], has to be accounted for when calculating the cumulative degradation. In this paper a model that predicts the thermal degradation of friction material is presented, and the comparisons between experimental measurements and model predictions of thermal degradation are illustrated.

Degradation Mechanism

Borg-Warner Automotive produces many types of friction materials, consisting of different constituents for various applications. For this study a production friction material with low heat resistance was selected for the sake of saving experimental time, since it may take over 10,000 engagement cycles for a friction material with high heat resistance to degrade a measurable amount. A TGA experiment of the raw friction material under a high resolution temperature ramp condition can be utilized to determine the percentages of cellulose fibers and inert components. The inert components are those that start to degrade after all cellulose fibers are carbonized. Figure 2 illustrates the sample weight loss as a function of temperature (the temperature ramp rate was set to 100°C/min for up to 300°C, and 200^oC/min for temperatures between 300°C and 950°C). The solid curve is the percentage of sample weight loss, and the dotted curve is the negative gradient of weight loss percentage. The gradient peak at lower temperatures corresponds to the maximum rate of cellulose degradation, and the second peak corresponds to that of inert contents. This figure indicates that cellulose fibers start to carbonize near 200^oC and finish the degradation at 420^oC under the TGA testing condition. It also reveals that the weight percentage of cellulose fibers for this material is 31.2%.

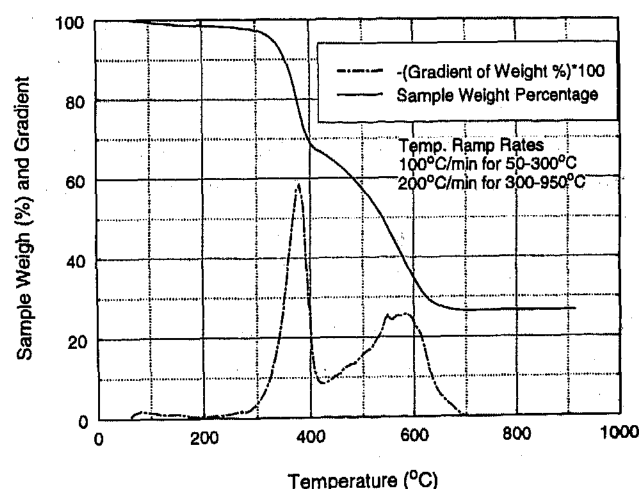


Figure 2: Weight Loss of Friction Material by TGA Analysis at Temperature Ramp Mode

Cellulose fiber is the least heat resistant component in this friction material. As cellulose fibers

encounter the high interface temperature during the repeated clutch engagements, they carbonize. The cellulose fibers undergo a chemical reaction and the interwoven fibers become brittle pieces of carbon after carbonization. After the carbonization reaches a certain degree, the friction material loses strength and small chunks of material at the surface may be pulled away.

The degradation amount/life cycles depends on the degradation rate and temperature history. The degradation mechanism and degradation rate need to be determined first. The degradation of friction material is a simple chemical reaction. We can first assume that the degradation is a n-th order reaction, and then determine the n value.

For a n-th order degradation, its reaction rate can be expressed as:

$$\frac{dW}{dt} = -kW^n \tag{1}$$

where W is the concentration of cellulose fibers (weight per unit volume), t is time, and k is the rate constant, which is a function of temperature and activation energy:

$$k = k_0 \exp\left(-\frac{E}{RT}\right) \tag{2}$$

where k_0 is the reaction rate constant, E is the activation energy of reaction, R is the Boltzmann constant, and T is temperature in °K. For example, for a first order reaction $n = 1$, and for a zeroth order reaction $n = 0$.

For a zeroth order degradation at the isothermal condition (constant temperature), by solving EQ (1) we obtain:

$$\frac{W}{W_0} = 1 - \frac{kt}{W_0} \tag{3}$$

where W_0 is the initial cellulose concentration, and t is the time duration of degradation at the isothermal condition.

Similarly, for the first order degradation at the isothermal condition, we obtain:

$$\ln\left(\frac{W}{W_0}\right) = -kt \tag{4}$$

Similar expressions can be derived for other orders of reaction ($n < 0$ and $n > 1$). However, there are more fundamental reasons for the zeroth and first order reactions.

If the degradation mechanism is a zeroth order, the plot of (W/W_0) vs t at various durations of experiments should be linear, where W is the cellulose weight per unit volume at time t. If the mechanism is a first order, the plot of $\ln(W/W_0)$ vs various t should be linear. Figure 3 shows the weight loss ratio W/W_0 as a function of time for the zeroth and first order degradations. The rate of zeroth order degradation is independent of cellulose concentration, and the rate of its weight change is constant. For a first order degradation, the rate of degradation depends on the concentration of cellulose fibers, and the rate of degradation decreases as the cellulose concentration decreases.

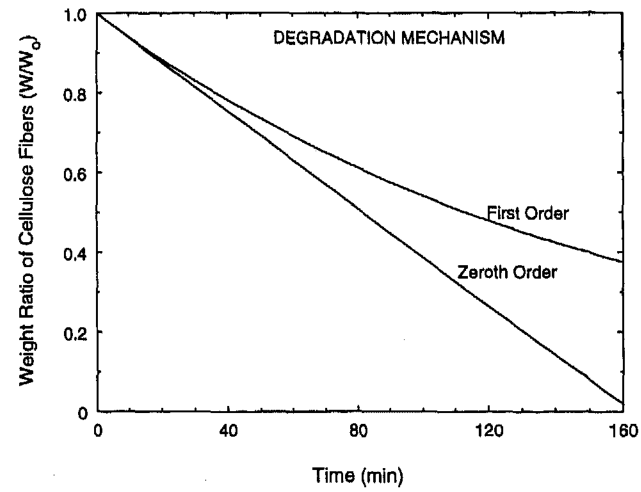


Figure 3: Degradation Mechanism of Friction Materials: Zeroth and First Order

The weight loss of cellulose fiber as a function of time under the isothermal condition can be measured from TGA experiments of new samples of friction material in a nitrogen or oxygen environment, or by measuring the cellulose content of the friction material after it undergoes the continuous slip test on the full-pack test machine. Under the continuous slip test, a constant power is supplied to the clutch pack and the temperature field reaches a pseudo-steady value in a short period of time (typically less than 1 minute) after the test starts. The tests typically run for more than 10 minutes. The temperature gradient across the thickness of the friction material is small at the pseudo-steady state. Thus, the temperature field exposed to the friction mate-

rial at the interface can be considered isothermal.

The weight loss of cellulose fiber under the continuous slip test best imitates the normal operating conditions of friction material in a clutch engagement. The detailed procedure and the experimental results of the continuous slip tests can be found in Twaddell's work [5]. Since the cellulose-to-inert ratio on the new friction material is not perfectly uniform, the average cellulose-to-inert ratio needs to be determined first. Ten samples were measured for the baseline cellulose-to-inert ratio with an average of 1.243±0.04, which is needed to determine the cellulose ratio W/W_0 on the worn materials. For each continuous slip test, eight samples of friction material were made by slicing a very thin layer (0.15mm) of friction material next to the friction interface. These samples were analyzed by the TGA to determine the cellulose-to-inert ratio, and the average ratio was calculated to obtain a mean value. From the baseline average cellulose-to-inert ratio and the average degraded cellulose-to-inert ratio, the cellulose weight ratio W/W_0 after degradation of time t is obtained. The cellulose ratio W/W_0 at various durations was obtained by measuring the samples from each corresponding continuous slip test.

Figure 4 shows the experimental results of $\ln(W/W_0)$ vs slip time for interface temperature of $260°C$. The temperature was measured by inserting a miniature thermocouple into the middle point between the separator ID and OD at its half thickness location. The temperature at this location should be very close to the temperature at the friction interface under the continuous slip condition because of the large thermal conductivity of steel and the pseudo-steady temperature distribution. If the degradation is a first order reaction, the experimental results (symbols in Figure 4) from various slip times should appear as a straight line in a logarithm scale. The best fit for the data (solid line) is reasonable except for the slip time of 60 minutes (relative error of W/W_0 between the fit and data is 9.9%). The W/W_0 ratios, from the same experimental data but plotted in a linear scale are illustrated in Figure 5 with the maximum relative error being 3.5%. If the degradation is a zeroth order reaction, all the experimental data should regress to a straight line. It is clear from Figures 4 and 5 that a zeroth order reaction represents the degradation mechanism better than the first order reaction. The degradation data obtained from the samples of continuous slip test under several other interface temperatures (230, 245, 275, 290, and $305°C$) also indicate that a zeroth order reaction describes the degradation mechanism better.

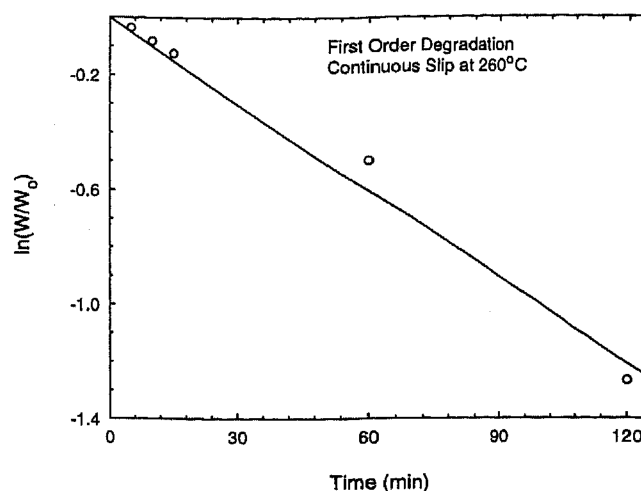


Figure 4: First Order Degradation Fit to the Experimental Weight Loss of Cellulose Fiber

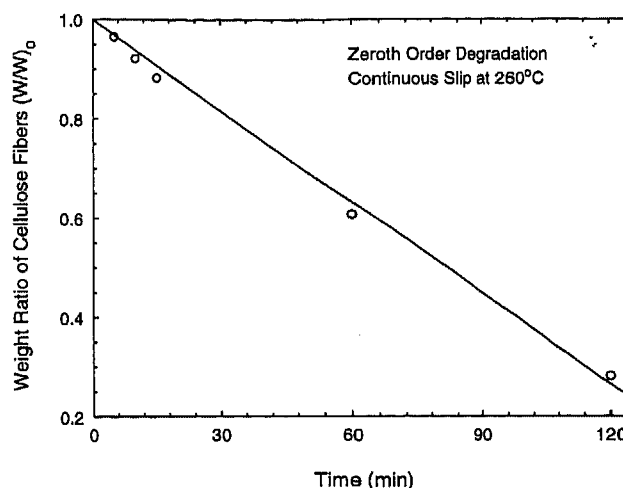


Figure 5: Zeroth Order Degradation Fit to the Experimental Weight Loss of Cellulose Fiber

For a zeroth order degradation, the slope of the W/W_0 vs time from one isothermal condition (one temperature) is the k/W_0 value for that temperature as described in EQ (3), while for the first order degradation the slope of $\ln(W/W_0)$ vs time is the k value from EQ (4). The reaction rate constant k_0 and the activation energy E of reaction can then be obtained by the least-square fitting of EQ (2) from several k values at different temperatures T.

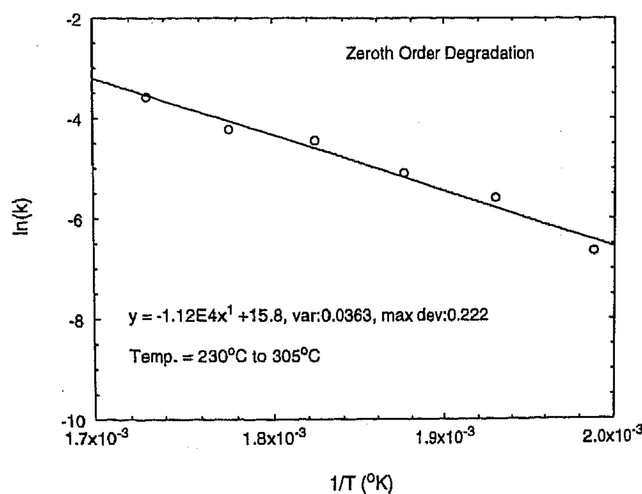


Figure 6: Determination of Rate Constant k_0 and Activation Energy E, Zeroth Order Model

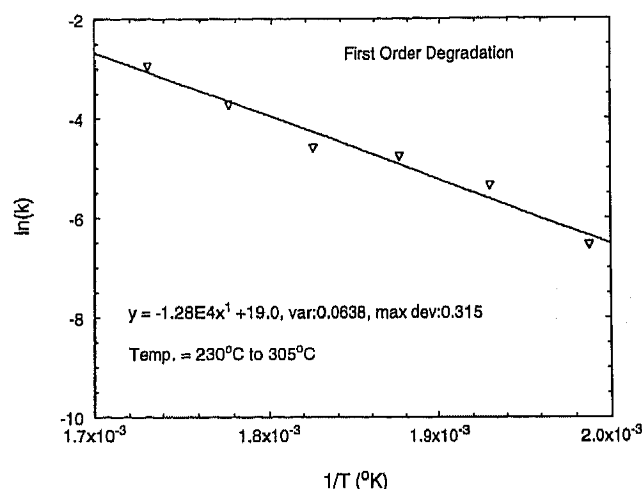


Figure 7: Determination of Rate Constant k_0 and Activation Energy E, First Order Model

Figures 6 and 7 depict the $\ln(k)$ vs $1/T$ for the zeroth order degradation and the first order degradation, respectively. The symbols are the experimental data and solid lines are the least-square fits to the experimental data. It is evident from these two figures that the least square fit for the zeroth order degradation yields smaller variance and deviation (0.0363 and 0.222, respectively) than those for the first order degradation (0.0638 and 0.315, respectively). It is then concluded that the degradation of this friction material is a zeroth order reaction with $k_0 = 1.203 \times 10^5 W_0$ (1/sec) and $E/R = 1.118 \times 10^4$ (°K).

Temperature History of Friction Material at Interface

Since the degradation rate is a strong function of temperature, the entire temperature history of friction material near the friction interface has to be accounted for when calculating the cumulative degradation during dynamic engagements. The temperature history at the interface can be measured by thermocouple (to be exact, the temperature at the separator plate can be measured). However, the temperature history of friction material below the interface cannot be measured easily since the friction plates are typically rotating. Therefore, we have to use a validated computer model [1] to obtain the necessary temperature information for evaluating the thermal degradation in the dynamic engagement situation.

Borg-Warner Automotive has developed a comprehensive and detailed mathematical model and computer software for calculating the temperature at the interface, as well as the temperature distribution as a function of time and spacial variables during the engagement cycle of a wet clutch. Many improvements and enhancements have been made since we presented this work in 1995 [1]. The model equations consist of heat balance for the separator, friction material, core plate, and Automatic Transmission Fluid (ATF) in the clearance and in the grooves. The momentum balance for ATF flow in the friction lining (porous medium) during the engagement was also considered. It has been verified by the experiments that the model can predict the temperature well during the engagement of a wet clutch [1]. Many important factors are included in the model, such as physical properties of steel and ATF as a function of temperature, non-uniform contact between separators and friction material, ATF flow into the friction material due to liquid permeability, compressibility of friction material, effects of separator splines, effects of rotational speeds, etc.

To further illustrate the capability of the thermal model, another comparison was made, as shown in Figure 8, between the model predictions and experimental measurements on the temperature history of the separator during a stable cycle of dynamic engagement. The experimental values were measured by inserting a miniature thermocouple into the middle point between the separator ID and OD

at the location of half separator thickness. The predicted values are also for the same location. It is demonstrated again that the model prediction of temperature agrees well with the experimental measurement. The physical parameters and operating conditions for the clutch configuration are shown in Table 1 with the initial rotating speed of 3000 RPM (low energy), which corresponds to one of the degradation studies for dynamic engagements. The temperature increases rapidly during the engagement stage, reaching the maximum temperature near the end of stop time. The temperature then decreases due to the heat conduction from the interface to the other parts of the clutch, and due to the heat transfer to the test machine housing at the outer and inner diameters. During the stabilization period (motor-on stage, from $t = 7.25$ sec to the end of cycle), ATF flows from the inner to outer diameter through the clearance and grooves on the friction plate, which rapidly lowers the temperature. At the end of a stable engagement cycle, temperatures reach the initial values of the previous cycle.

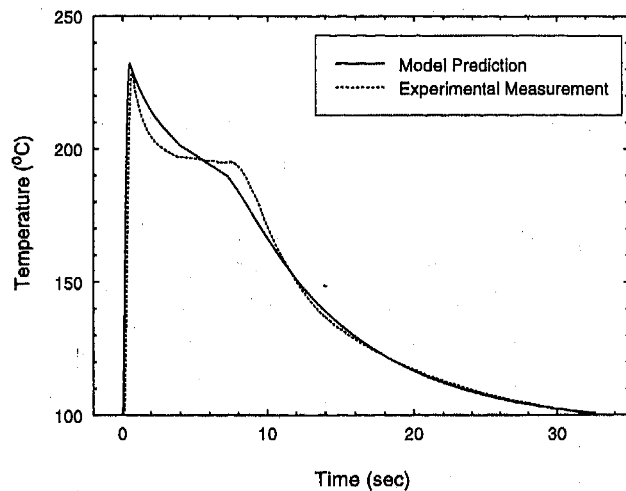


Figure 8: Model Prediction and Experimental Measurement on Interface Temperature

Figure 9 illustrates the temperature distribution across the axial thickness at the middle point of radial direction during the engagement stage (from t=0 to 0.52 sec). τ is the dimensionless engagement time ($\tau = t/t_e$, where t_e is the stop time, and $\tau = 1$ is at the end of engagement stage). The dimensionless thickness $\xi = 0$ is at the symmetric line of the separator, and $\xi = 1$ is at the symmetric line of the core plate. This figure illustrates that a large temperature gradient exists across the friction material

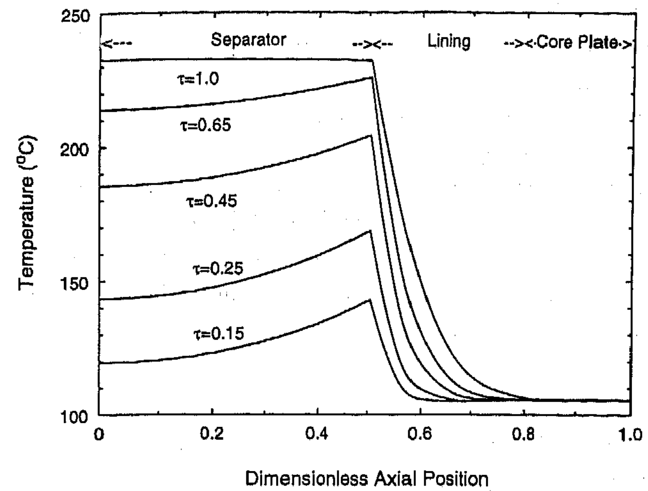


Figure 9: Temperature Distribution in Thickness Direction During Clutch Engagement

Table 1. Summary of Parameters for Dynamic Engagements

Energy Level	Low	High
Cycles	1000 4000 8000	1500 2500 4000
Speed (RPM)	3000	3400
Inertia ($kg.m^2$)	0.8466	0.8466
Groove Type	double parallel	double parallel
No. of Grooves	52	52
Stop Time (sec)	0.52	0.5
Soak Period (sec)	3.48	3.5
Dwell Period (sec)	3.25	3.25
Stabilization P.(sec)	24	24
Inlet ATF Temp.(^oC)	87±2	87±2
Oil Flow Rate (lpm)	1.514	1.514
No. of Friction Faces	8	8
Lining ID (mm)	117.5	117.5
Lining OD (mm)	149.7	149.7
Separator Thick. (mm)	2.36	2.36
Lining Thickness (mm)	0.76	0.76
Core Thickness (mm)	0.81	0.81
Pack Clearance (mm)	1.3	1.3

(lining) between the interface and the bonding plane due to the low conductivity of friction material. The temperature profile of friction material between the interface and 0.15 mm depth from the interface is important, since the degradation amount of experiment for dynamic engagement is performed by slic-

ing a 0.15 mm thickness of friction material from the interface. The temperature history of this 0.15 mm thick friction material next to the interface needs to be accounted for when computing the cumulative degradation by the degradation model. By doing so, meaningful comparisons can be made between the experimental data and model predictions. It should be emphasized that the temperature history is a function of time, radial position, axial position, and other parameters [1].

Degradation Under Dynamic Engagement

In order to predict the degradation amount for a dynamic engagement cycle, we have to account for the temperature history of friction material as a function of time and axial locations. For a small slice of friction material at a certain radial location near the friction interface, the degradation amount for a zeroth order reaction from time t to time $t+\Delta t$ (Δt: time step from the thermal model) can be evaluated by:

$$\left(\frac{W_{t+\Delta t}}{W_t}\right) = 1 - \frac{1}{W_t}\int_0^{\Delta z}\frac{dz}{\Delta z}\int_t^{t+\Delta t} k_0 \exp\left[-\frac{E}{RT(t,z)}\right] dt \tag{5}$$

where Δz is the small thickness of friction material from the interface (slicing thickness), and $T(t,z)$ is the temperature, a function of time and axial depth from the interface. The degradation amount for this small time step Δt is:

$$\Delta W_i = W_{t+\Delta t} - W_t \tag{6}$$

The cumulative degradation for a dynamic engagement cycle is obtained by summation of all the time steps, i.e.

$$\Delta W_{cycle} = \int_0^{t_{cycle}} \frac{dW}{dt} dt = \sum_{i=1}^{N_{step}} \Delta W_i \tag{7}$$

where N_{step} is the number of time steps.

The double integration in EQ (5) needs to be evaluated efficiently and accurately, considering that there are over 200 time steps for each engagement cycle, and more than 10 divisions from the inner radius to outer radius for the degradation calculation. Several integration schemes have been tried and the Gaussian quadrature is chosen for this work because of its numerical accuracy and fast computation.

For any given two-dimensional function $f(x,y)$, the double integration by the Gaussian quadrature can be expressed by the following formula, and the numerical value of integration is exact if $f(x,y)$ can be represented by a polynomial degree of $(2n_g - 1)$:

$$\begin{aligned} P &= \int_a^b\int_c^d f(x,y)dxdy \\ &= \frac{(b-a)}{2}\frac{(d-c)}{2}\int_{-1}^{+1}\int_{-1}^{+1} f(\xi,\eta)d\xi d\eta \qquad (8) \\ &= \frac{(b-a)}{2}\frac{(d-c)}{2}\sum_{i=1}^{n_g}\sum_{j=1}^{n_g} f(\xi_i,\eta_j)w(i)w(j) \end{aligned}$$

$$\xi = \frac{2x-(a+b)}{b-a} \qquad \eta = \frac{2y-(c+d)}{d-c}$$

where n_g is the number of Gaussian points, $w(i)$ and $w(j)$ are the Gaussian weights, ξ_i, and η_j are the Guassian values. Table 2 shows the Gaussian values and Gaussian weights for $n_g = 11$.

Table 2: Gaussian Points and Weights ($n_g = 11$)

Gaussian Points	Gaussian Weights
-0.97822866	0.05566856
-0.88706260	0.12558037
-0.73015201	0.18629021
-0.51909613	0.23319376
-0.26954316	0.26280454
-0.00000000	0.27292509
0.26954316	0.26280454
0.51909613	0.23319376
0.73015201	0.18629021
0.88706260	0.12558037
0.97822866	0.05566856

The cumulative degradation amounts from the dynamic engagement were measured on the samples of friction materials after they have been engaged for 1000, 4000, and 8000 cycles for the low energy engagement and for 1500, 2500, and 4000 cycles for the high energy engagement. The testing parameters are listed in Table 1 and the samples are taken near the inner diameter and outer diameter of the

friction plate. The sample measurements were made in a similar manner as those for the samples of the continuous slip as described earlier for determining the degradation rate. A total of sixteen samples (eight near the inner diameter and eight near the outer diameter) of friction material were taken and measured for each set of engagement cycles in order to obtain an average value.

The comparison between the experimental measurements and model predictions on the cumulative degradation is illustrated in Figures 10 and 11 for the low and high energy engagements, respectively.

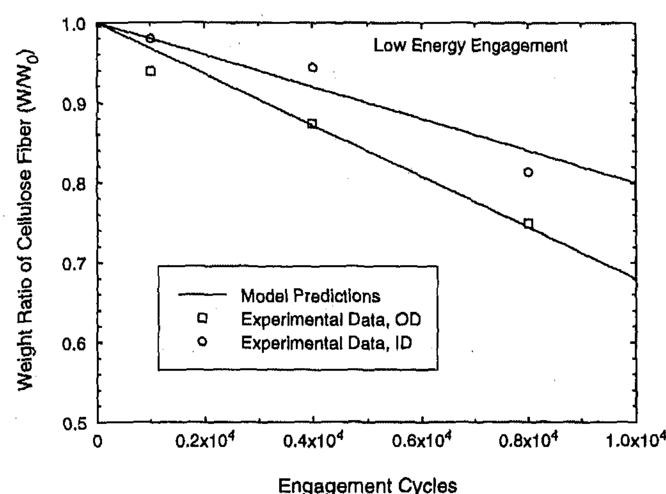


Figure 10: Model Predictions and Experimental Measurements on Cumulative Degradation (Low Energy Level)

The symbols in these two figures are the experimental measurements, with squares for the samples taken near the outer diameter and circles near the inner diameter (each symbol is the average of eight measurements). Figures 10 and 11 show that the predicted degradations of friction material agree fairly well with the experimental measurements of thermal degradation for both the low and high energy engagements, and for the samples taken from both the inner and outer diameters of friction material. The friction material close to the outer diameter degrades more than that close to the inner diameter. This is due to the fact that the temperature near the outer diameter is higher than that near the inner diameter, since more friction heat is generated near the outer diameter, and cold ATF flows from the inner diameter to the outer diameter of the friction plate. This results in the temperature near the outer diameter being higher than that near

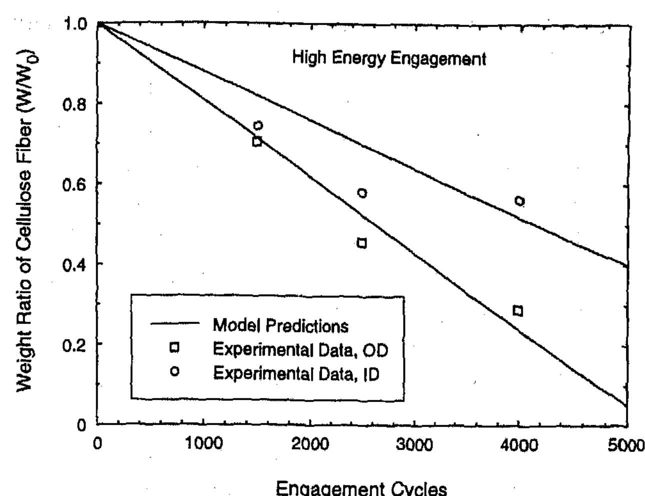


Figure 11: Model Predictions and Experimental Measurements on Cumulative Degradation (High Energy Level)

the inner diameter during engagement.

Since the prediction of the thermal degradation of friction materials involves many uncertainties, it is not expected that the model prediction will be perfect. The non-perfect agreements between the theoretical predictions and the experimental measurements shown in Figures 10 and 11 (especially Figure 11) may be due to one or more of the following factors. The degradation rate constant k_0 and the activation energy E as shown in Figure 6 are not perfectly accurate. The prediction of temperature distribution by the thermal model does not agree exactly with the experimental values. The contact between the friction plate and the separator may not be perfectly uniform, and non-uniform contact may generate locally high temperature. All these factors contribute to the slight disagreements between the model predictions and measurements. The key factor is the temperature history that friction materials encounter during the dynamic engagement. The degradation rate is an exponential function of temperature, and any inaccuracy in temperature predictions will be propagated to the degradation predictions.

Summary

A model and computer software have been developed to predict the thermal degradation of friction material after it undergoes a certain number of dy-

namic engagement cycles. The degradation mechanism for the selected friction material obeys the zeroth order reaction well. The degradation rate as a function of temperature was determined by the TGA measurement of the samples from the continuous slip test on the full-pack machine. The whole temperature profile of friction material in a thin layer close to the interface was considered when evaluating the degradation. The temperature profile was obtained by the thermal model. The experimentally measured cumulative degradations from dynamic engagements were compared favorably with the predictions of the degradation model. This model will provide useful information for engineers, and assist them to optimize the design of a clutch pack for a given application. Furthermore, the model is flexible and can be used to project material life under variable cycle conditions at any defined set of application conditions.

Acknowledgments

We would like to express our appreciation and gratitude to Donn Fairbank and Bruce Clay for their valuable help and assistance in this work.

References

[1] Yang, Y.B., Lam, R.C., Chen, Y.F. and Yabe, H., "Modeling of Heat Transfer and Fluid Hydrodynamics for a Multidisc Wet Clutch", SAE 950898 (1995).

[2] Osanai, H., Ikeda, K., and Kato, K. " Relations between Temperature in Friction Surface and Degradation of Friction Materials During Engagement of Friction Paper", SAE 900553 (1990)

[3] Ohnuma, H. and Kato, K. "The Effect of Groove Pattern of Paper Friction Plate on Its Life", SAE 910804 (1991)

[4] Takezaki, K. and Kubota, M. "Thermal and Mechanical Damage of Paper Friction Material Induced by Non-uniform Contact", SAE 922095 (1992)

[5] Twaddell, P.S., "Development of a Thermal Distress Model for the Life of Wet Clutch Friction Material", B.S. Thesis, GMI Engineering & Management Institute, Detroit, MI (1996)

970979

Friction and Wear Properties of Integrated Composite Copper-Based Friction Materials

Katsuyoshi Kondoh, Yoshishige Takano, and Yoshinobu Takeda
Sumitomo Electric Inds., Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

A new sintered composite friction material consisting of mechanically-alloyed copper-based composite powders has been developed. It has a unique microstructure of fine hard particles that are embedded in the matrix of the copper-based primary particles. This friction material reduces abrasive wear and/or seizures that are often caused by hard particles which become detached from the matrix, because the hard particles in this material are bonded tightly to the matrix even under harsh service conditions. The fine hard particles are also very useful for improving high friction force when contacting a surface material. Therefore, this new friction material provides a higher friction coefficient than the conventional material containing coarse hard particles. Furthermore, this friction material is less abrasive to the surface material than the conventional one. The new sintered composite friction material shows outstanding performance and reliability when used as a friction component in some friction devices, such as magnetic clutches in automotive air-conditioners and brake systems. The friction characteristics are bound to be superior to those of the conventional sintered copper alloy or organic material currently used for such friction components.

INTRODUCTION

In friction devices, such as magnetic clutches and brakes, good wear resistance, high friction force and high durability are required to make a more compact or smaller friction device system with high performance. As shown in equation (1), Friction force F is relative to the coefficient of friction μ of friction material and the material with higher coefficient friction has the ability to produce a friction device system with greater braking force. Furthermore, when the friction coefficient μ increases, the number of contact components n will decrease. This means that it is possible to design a more compact friction device system by using a friction material with a higher friction coefficient.

$$F=\Sigma(\mu P_i)=n \mu P \qquad (1)$$

P_i ; Pressure on each contact component

The friction material usually also requires the heat resistance because the temperature at the surface sliding against a surface material rises due to the increased friction heat when the friction coefficient increases. Furthermore, metal-based friction material is more available than a paper based or organic one when developing an advanced friction material with a higher friction coefficient.

Sintered friction material used for friction devices usually consists of hard particles such as ceramics and intermetallic compounds, and self-lubricant particles. Figure 1 shows a typical microstructure of a sintered copper-based friction material. Copper, particularly a Cu-Sn brass alloy, is often used for the base metal of sintered friction materials. Elements

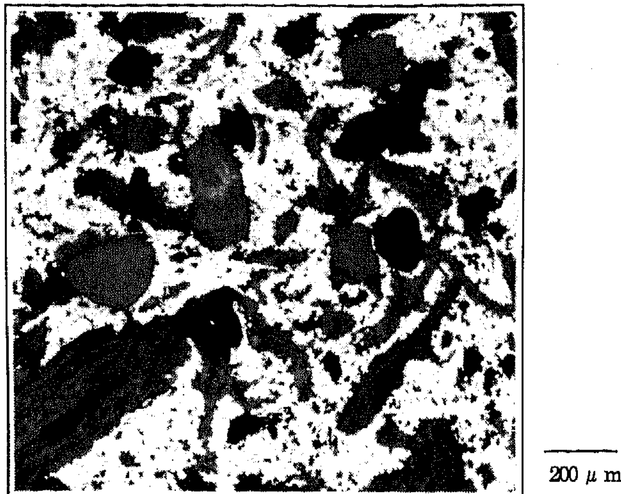


Figure 1. Microstructure of the conventional sintered friction material.

added to a copper-based alloy, such as Ni, Fe, Zn improve the mechanical and wear properties of the sintered copper alloy. A self-lubricant, such as graphite, MoS_2 ,or BN, creates lubricant films along the interfaces between the friction material and the surface, so it is easy to maintain a constant friction coefficient. Hard particles, such as SiO_2 or Al_2O_3 dispersed in the sintered material, also have an important role in improving the wear resistance and high friction coefficient when contacting a surface material. In particular, sintered friction components have poor wear resistance when the content of contained hard particles is less than the optimum range, 10~30mass%.

WEAR PHENOMENA OF THE CONVENTIONAL SINTERED FRICTION MATERIAL

Firstly, the mechanism of wear phenomenon, particularly a seizure, was analyzed in order to develop a new sintered friction material with excellent anti-wear properties. Figure 2 shows a change in the friction coefficient of the conventional sintered copper alloy using the ring-on-disk wear test under dry conditions as shown in Figure 3, using a low carbon steel (S15C) as the surface material. The sintered Cu-9%Sn+FeMo alloys, whose microstructures are shown in Figure 4, are applied to contact/friction materials in this work. FeMo intermetallics with average diameters of 60~70 μ m and

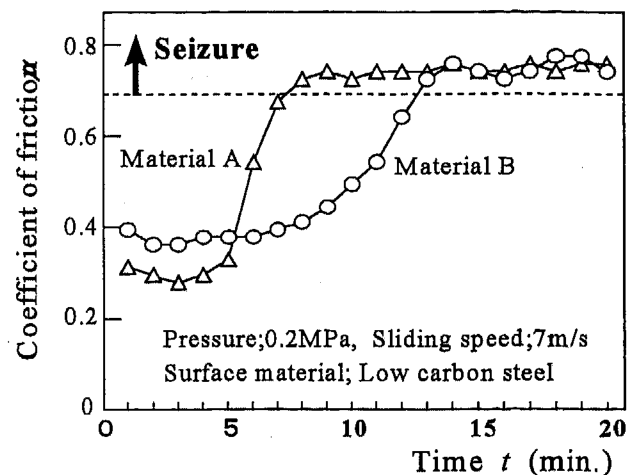


Figure 2. A change in the friction coefficient of conventional sintered copper-based materials.
Material A; Cu-9%Sn+10%FeMo (Mass%)
Material B; Cu-9%Sn+25%FeMo (Mass%)

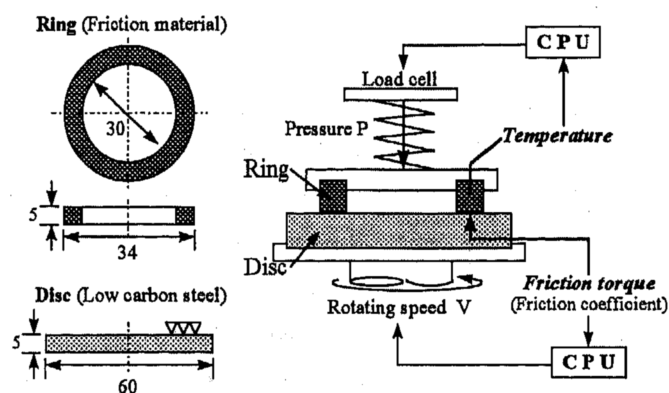


Figure 3. Ring-on-disc sliding wear test equipment used in this work.

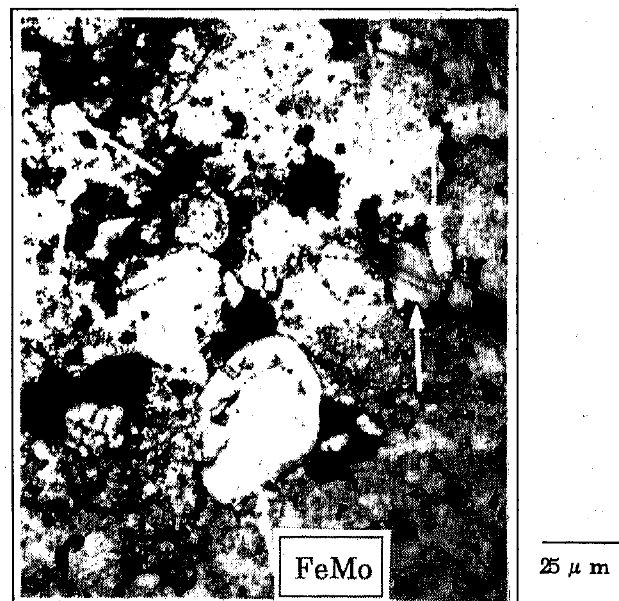


Figure 4. Microstructures of the conventional sintered composite friction material.
(Chemical compositions ; Cu-9%Sn+20%FeMo)

average hardness of 830mHv (micro-vickers hardness) are added as hard particles to each sintered copper alloy. The initial friction coefficient of each sintered copper alloy is a constant 0.3~0.4. The friction coefficient, however, gradually increases and finally each alloy has a high friction coefficient over 0.7 because of the seizure phenomena at the surface in which it contacts a surface material, independent of the content of FeMo intermetallics. The results of the microscopic and Scanning Electron Microstructure (SEM) analysis show that FeMo hard particles are dispersed not only at the primary particle boundaries (PPB) of the sintered friction material, but also at the contacting surface of the surface material. Furthermore, pores are observed at the contacting surfaces of each sintered material. Therefore, it is quite certain that these pores at the PPB correspond to the traces where FeMo hard particles detached from the matrix of the sintered friction material. The hard particles which detach from and cause seizure phenomena when contacting a surface material do so because of gaps between FeMo hard particles and the matrix of the primary copper particles as shown in Figure 4. Based on these results, the mechanism of seizure phenomena of the conventional sintered friction material is considered as follows. Raw powders, which consist of copper-based powders and hard particles, are consolidated into the green compact by pressing as shown in Figure 5.

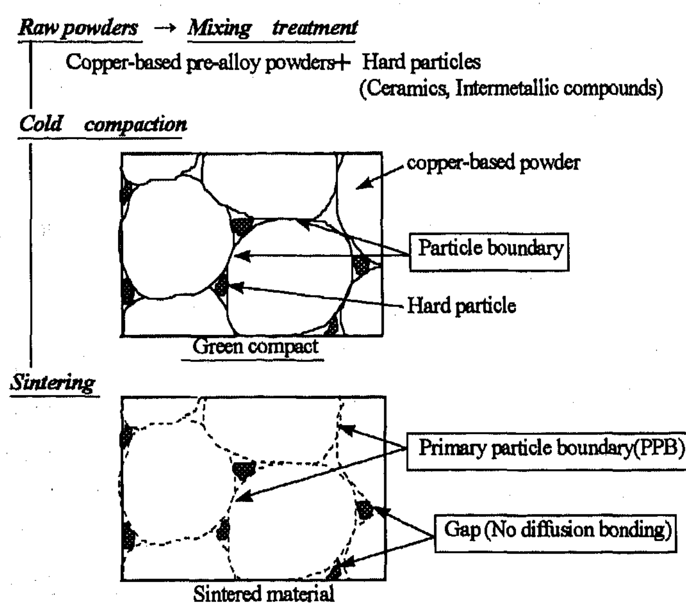


Figure 5. Procedure of the conventional sintered composite friction material.

The hard particles are dispersed at the primary copper-based powder boundaries in the compact. After sintering, however, it is impossible to provide a diffusion layer between the hard particles, such as ceramics particles and/or intermetallic compounds, and the copper-based powders because metallic diffusion bonding does not occur between them even in sintering. Therefore, as shown in Figure 6, FeMo hard particles easily detach from the matrix of the sintered friction material under severe dry conditions, such as high load or high sliding speed, because of the lack of a diffusion layer with the matrix.

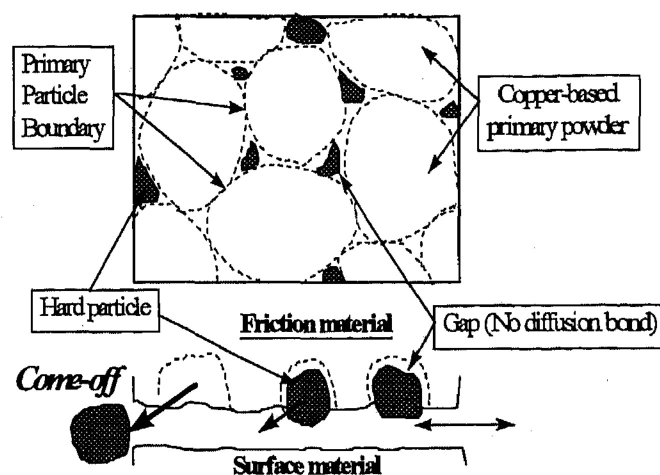


Figure 6. Schematic microstructure of a typical sintered friction material and the cross section when contacting a surface material.

That is, gaps between the hard particles and the matrix cause a poor bonding at the interface therebetween, so the hard particles easily detach from the matrix when contacting the surface material causing seizure or wear phenomena. In addition, these hard particles which detach from the matrix attack the surface material cause wear phenomena to occur at the contacting surface of the surface material as well.

ADVANCED SINTERED COMPOSITE COPPER-BASED FRICTION MATERIAL

MATERIAL DESIGN BY MECHANICAL ALLOYING PROCESS — Based on the above results, it is essential to

bond the hard particles to the matrix of the sintered friction material in order to improve the anti-seizure property, wear resistance and the stability of the friction coefficient when contacting the surface material. An ideal schematic microstructure of the sintered friction material, as shown in Figure 7 compared with the conventional one, has essentially a reaction/diffusion layer at the interfaces between hard particles and the matrix. This layer serves to bond well the hard particles to the matrix.

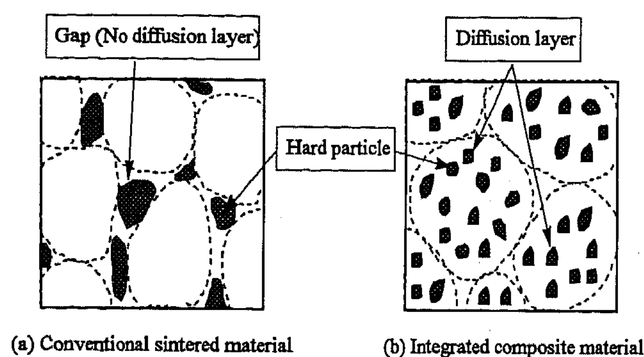


Figure 7. Ideal schematic microstructures of the integrated sintered composite friction material compared to the conventional sintered one.

There are methods to realize this microstructure with a powder metallurgy process, such as reaction sintering or diffusion bonding, etc. In this work, a mechanical alloying (MA) process is selected to provide a reaction/diffusion layer between the hard particles and the matrix. Figure 8 shows a schematic of the MA process and the microstructure of the metal matrix composite (MMC) powders made by the MA process. Coarse hard particles ground into fine hard particles by ball milling, are dispersed in the matrix of the copper alloy based powders. They also have a diffusion layer with the matrix due to the high energy given to powders by ball milling. The new sintered friction material consists of these mechanically-alloyed MMC powders used as raw powders. Therefore, it has good ability to control seizure phenomena because FeMo hard particles have become sufficiently embedded in the matrix to resist friction with a surface material. In addition, it is possible to consider that this material has not only good anti-seizure properties but also good wear resistance, because the matrix of base copper alloy which has fine crystal microstructures produced by the MA process, is harder and stronger than that of the conventional sintered copper-based friction material.

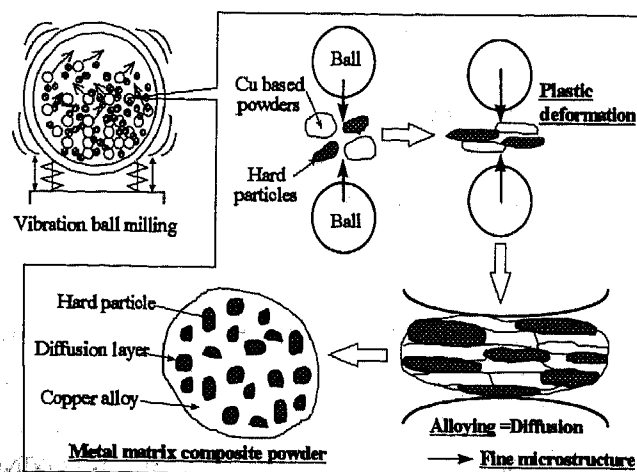


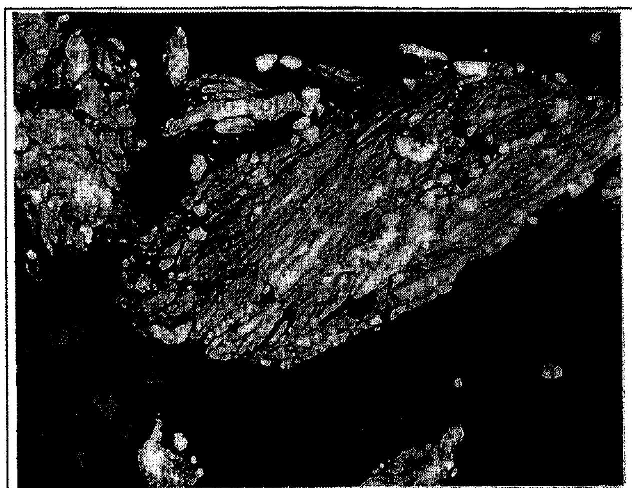
Figure 8. Schematic of the mechanical alloying(MA) process and the microstructure of a metal matrix composite(MMC) powder made by the MA process.

MICROSTRUCTURES OF METAL MATRIX COMPOSITE POWDERS AND SINTERED MATERIAL— Cu-9mass%Sn mixed powders are used for base metal alloys. FeMo intermetallics whose average diameter is 70 μ m and average hardness is 830mHv are selected as hard particles in this work. The MA conditions with a vibration ball milling device are shown in Table 1 and the milling time is used as a parameter in this work.

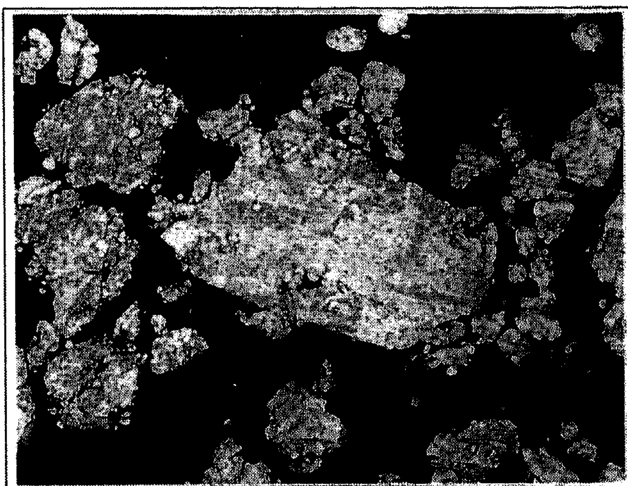
Table 1. Test conditions in mechanical alloying with vibration ball milling device.

Vibration conditions		Milling time	Powder weight	Ball weight
Frequency	Width	t	W_P	W_B
1200 rpm	8 mm	1~20 hr	800 g	8 kg

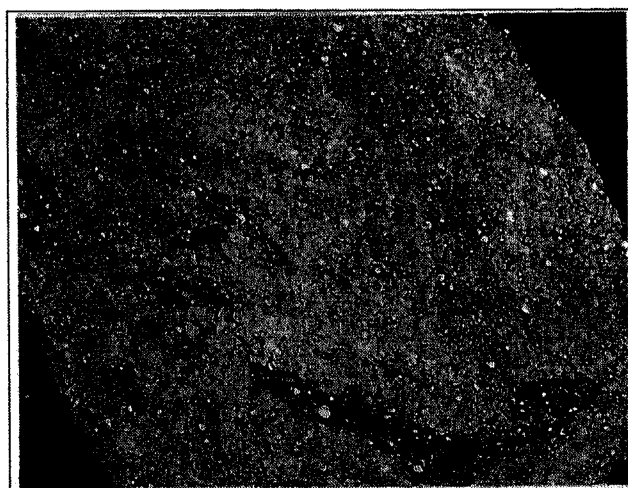
Figure 9 reveals the microstructures of MMC powders produced by the MA process at each milling time. For a milling time period under 3hrs., FeMo hard particles of a size less than about 10 μ m in a diameter, are uniformly dispersed in the matrix, although layer-like microstructures of the copper alloy are clearly observed in the matrix of powders. On the other hand, with a milling time of over 5hrs., very fine FeMo hard particles of less than 5 μ m in diameter are uniformly dispersed in the matrix, and the uniform microstructure of the copper alloy is also observed in the matrix. Figure 10 shows the



(a) After 3 hrs. for milling



(b) After 5 hrs. for milling



(c) After 8 hrs. for milling 50 μ m

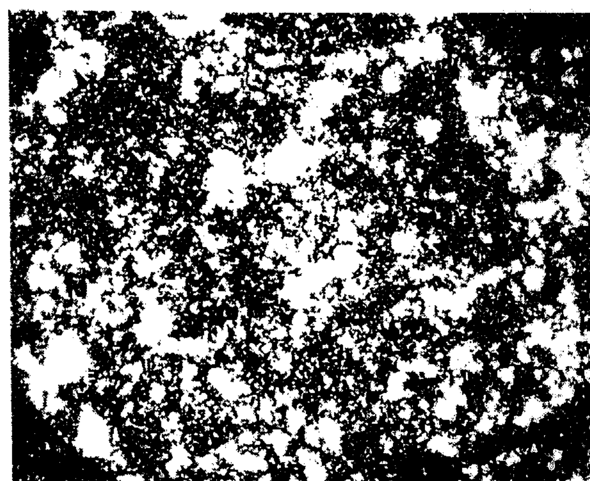
Figure 9. Microstructures of MMC powders produced by MA process. (Chemical composition; Cu-9%Sn+20%FeMo).

result of an Electron Probe Micro-Analyzer (EPMA) on the mechanically-alloyed MMC powders after 5hrs. for milling.

This means that FeMo hard particles, which are shown by white dots, are fine and uniformly dispersed in the matrix of the mechanically-alloyed Cu-Sn powder.



(a) Microstructure (SEM observation) 10 μ m

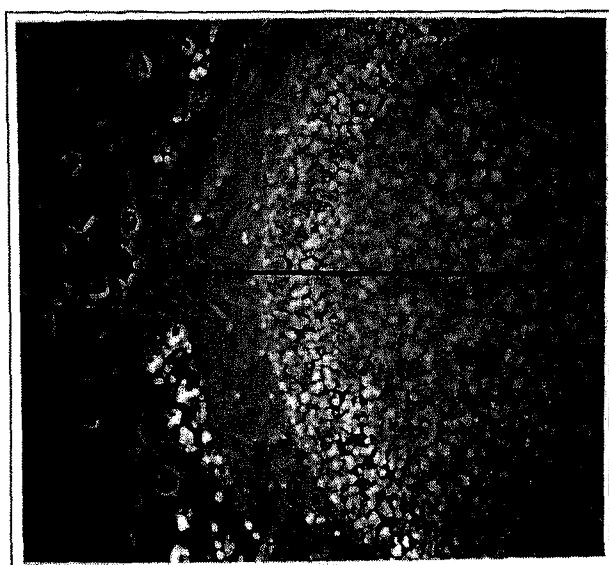


(b) Distribution of Fe element (EPMA result)

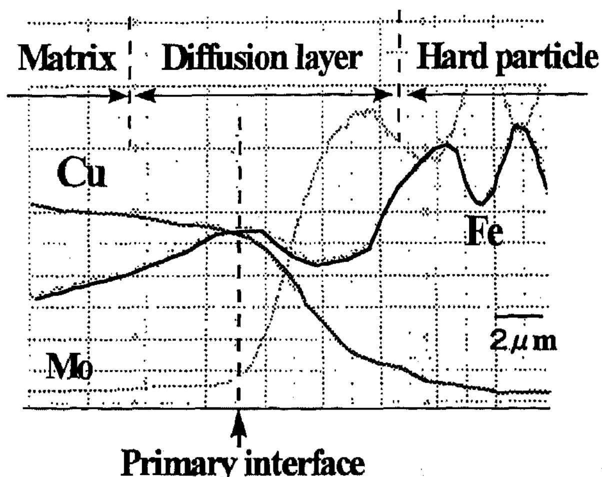
Figure 10. A result of EPMA on the mechanically alloyed MMC powders after 5hrs. for milling.

Furthermore, Figure 11 reveals that the diffusion layer of over 10 μ m width is clearly observed at the interface between the hard particle and the matrix by EPMA equipment.

This means that FeMo hard particles, which are embedded in the matrix of the mechanically alloyed MMC powders, are bonded well to the matrix. Furthermore, no gap is observed at the interface between the hard particle and the matrix.



(a) Microstructure (SEM observation) 5 μ m



(b) Distribution of Fe, Mo and Cu elements

Figure 11. A result of X-ray analysis at the interface between the hard particle and the matrix of MMC powders after 5hrs. for milling.

These mechanically-alloyed powders are consolidated into the sintered composite friction material based on the experimental procedure as shown in Figure 12, and the microstructures of the sintered material are evaluated by a microscopic analysis.

Figure 13 shows the microstructure of this advanced sintered friction material using mechanically-alloyed MMC raw powders, compared to that of the conventional sintered one with the same chemical composition. Very fine FeMo hard particles are uniformly dispersed in the matrix of the mechanically-alloyed Cu-9%Sn based powders.

Raw powder

Cu-9%Sn mixed powders, FeMo hard particles

Mechanical Alloying (Vibration ball milling)

Powders; 800g (Cu-9%Sn+20%FeMo)
Ball; 8.0kg (Φ 9.4mm SUJ II steel ball)
Milling time; 8hrs

Cold compaction

Relative density of green compact; 70~75%

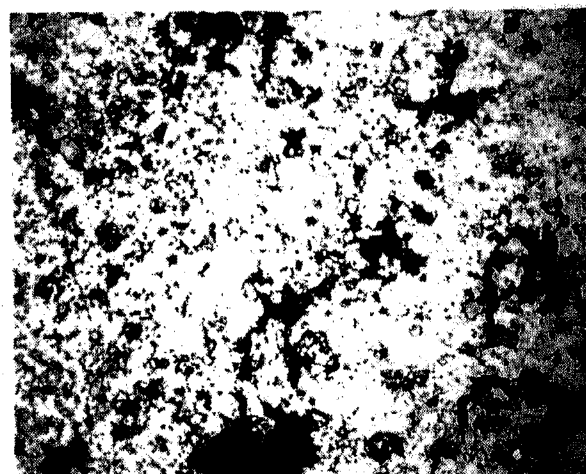
Sintering

Temperature; 800℃, Sintering time; 1hr
Atmosphere; N_2

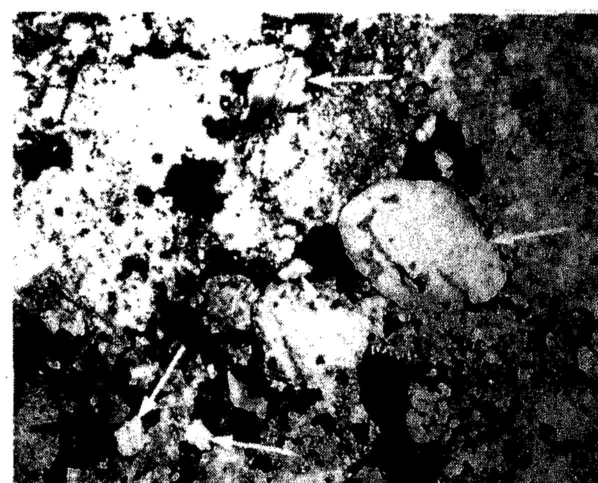
Evaluation

Mechanical and tribological properties

Figure 12. Procedure of the mechanically-alloyed sintered composite friction material.



(a) Mechanically alloyed sintered friction material



(b) Conventional sintered friction material 25 μ m

Figure 13. Microstructures of the mechanically-alloyed MMC sintered friction material compared to the conventional one.
(Chemical compositions of both materials ; Cu-9%Sn+20%FeMo)

On the other hand, coarse FeMo hard particles are dispersed in the primary particle boundaries of the conventional sintered friction material and there are gaps at the interfaces between hard particles and the matrix. This is why in producing the conventional sintered friction material, the mixing treatment of raw powders makes it impossible to embed the FeMo hard particles in the matrix of copper alloy powders. Thus, this means that the MA process has the ability to realize the ideal microstructures of a sintered friction material which has excellent anti-seizure properties. Therefore, it is expected that this advanced sintered composite friction material will satisfy the requirements of good wear resistance and a constant friction coefficient under severe friction conditions.

MECHANICAL AND TRIBOLOGICAL PROPERTIES OF ADVANCED SINTERED MMC FRICTION MATERIAL — Table 2 shows the physical and mechanical properties of this new sintered composite friction material which consists of mechanically-alloyed MMC powders as raw powders, compared to the conventional one with the same chemical composition. The strength and hardness of the new sintered friction material are superior to those of the conventional one because FeMo hard particles are bonded well to the copper alloy matrix, and the matrix is strengthened by fine copper crystals.

Table 2. Physical and mechanical properties of sintered composite friction materials.

	Density	Bending strength	HardnessHRB
MA composite	5.96 g/cm³	207 MPa	56
Conventional	5.91 g/cm³	118 MPa	42

(Chemical compositions. Cu-9%Sn-25%FeMo)

Figure 14 shows a change in the friction coefficient of both friction materials, using a ring-on-disk wear test equipment as shown in Figure 3 under dry conditions (Pressure;0.2MPa, Velocity;10m/sec.). This advanced friction material has a higher friction coefficient with a constant 0.4~0.45, and no seizure phenomenon occurs at the surface contacting the surface material. On the other hand, the friction coefficient of the conventional material gradually increases, and finally, it reaches 0.7 with seizure phenomena. The above result reveals that the friction coefficient of the new friction material is higher than that of the conventional one in which seizure phenomena has not occurred yet. Usually, the surface area of hard particles which contact the surface material determines the friction force, and the surface area of fine hard particles contacting a surface material is larger than that of coarse ones. Therefore, this new friction material provides a higher friction coefficient than that from a conventional material made of coarse hard particles, because the finer hard particles which raw coarse particles are ground by the MA process, are dispersed in the matrix of this new friction material. Figure 15 shows the wear thickness of each sintered friction material and surface materials. In the case of this new sintered friction material, wear of both the friction material itself and its surface is very low, so this result means that a new integrated friction material has excellent anti-wear property. It is also found that no FeMo hard particles detach from the matrix at the contacting surface, using microscopic and SEM analysis. That is, hard particles are seen to be perfectly embedded in the matrix.

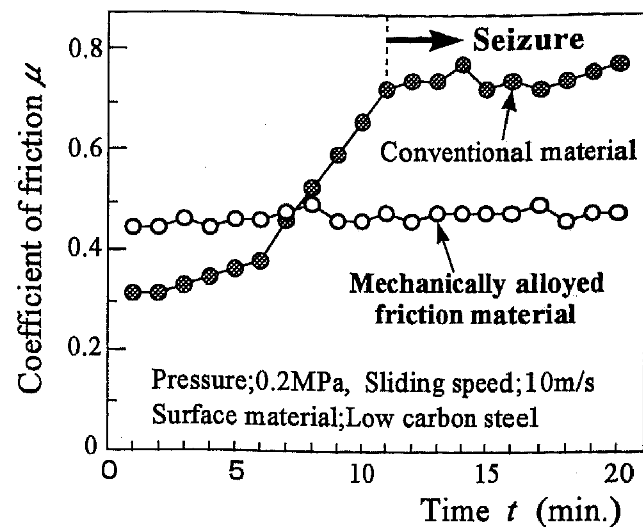


Figure 14. A change in the friction coefficient of the mechanically-alloyed sintered composite friction material and the conventional one. (Chemical composition of both materials ; Cu-9%Sn+25mass%FeMo)

APPLICATION TO A MAGNETIC CLUTCH MATERIAL IN AUTOMOTIVE AIR-CONDITIONERS

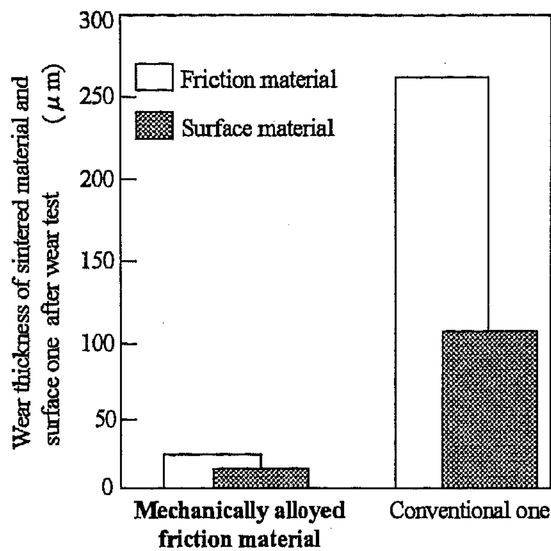


Figure 15. Wear properties of mechanically alloyed copper-based friction material compared to the conventional one.

In a magnetic clutch system, the ring-shaped friction material which is fixed on the surface of either a rotor or an armacher as shown in Figure 16, has an important role in stopping the rotation of the armacher instantaneously and synchronizing the armacher with the rotor. Therefore, friction material with a higher friction coefficient is required to produce a new magnetic clutch system with higher friction force and outstanding performance. It must also contribute to weight reduction on the magnetic clutch system.

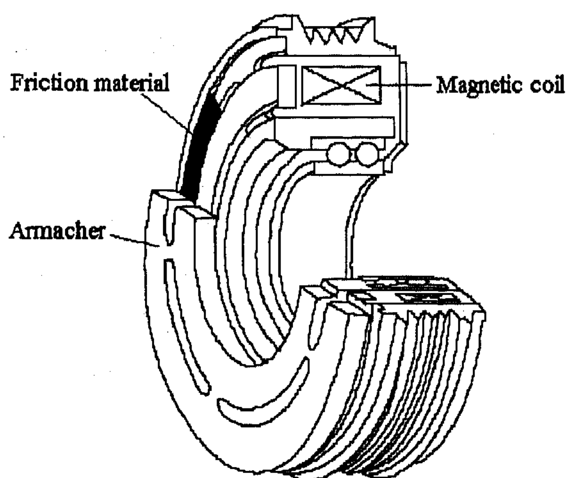


Figure 16. Schematic of a magnetic clutch system used in an automotive air-conditionaer.

That is, as mentioned in the above equation (1), a higher friction clutch material makes it possible to realize a more compact system with the same friction force because the radius of the clutch material R is smaller due to the higher friction coefficient. Good wear resistance is also necessary under more severe sliding conditions. Thus, this integrated composite sintered material with both higher friction coefficient and excellent wear resistance has good potential to realize the new magnetic clutch system with high performance.

Figure 17 shows a change in the friction coefficient of each friction material, the mechanically-alloyed composite sintered material, the conventional sintered copper alloy and an organic one in use, evaluated by a cyclic wear test simulating a magnetic clutch in an automotive air-conditioner by using Ring-on-disc wear test equipment (See Figure 3) under sliding conditions as shown in Figure 18.

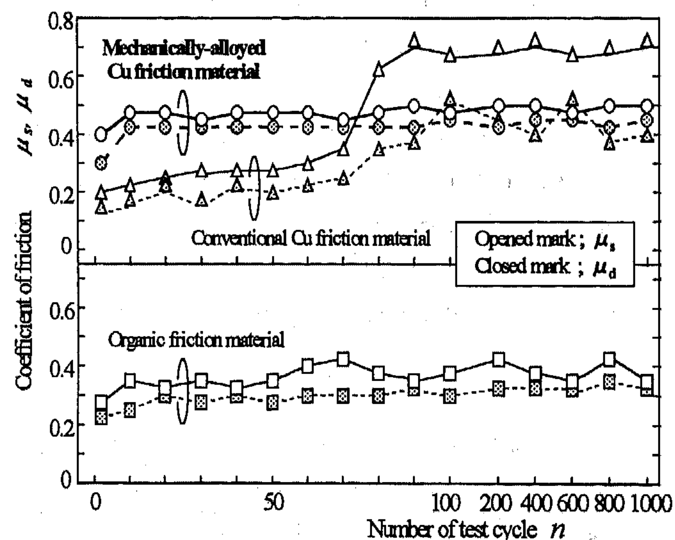


Figure 17. A change in the friction coefficient of each friction material evaluated by cyclic wear test simulating a magnetic clutch used in an automotive air-conditioner. (Chemical composition of both sintered materials ; Cu-9%Sn +15mass%FeMo)

In particular, μ_s corresponds to the maximum friction coefficient at the moment of starting or stopping, meaning the statistic one, and μ_d means the dynamic friction coefficient in this measurement. This mechanically-alloyed sintered friction material reveals a higher constant friction coefficient than both the conventional sintered friction material and the organic one, and no seizure or wear phenomenon at the contacting surface is observed. The difference particularly of μ_s and μ_d, which

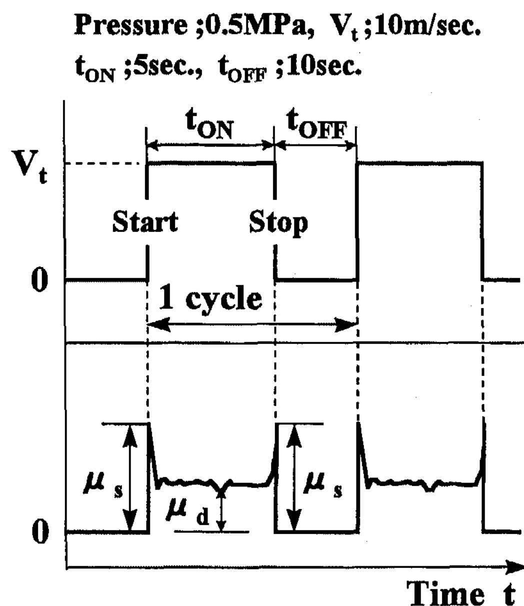


Figure 18. Ring-on-disc cyclic wear test conditions used in this work.

is relative to the vibration at the moment of stopping and starting the magnetic clutch, is smaller than that of the conventional sintered material with the same composition. Therefore, this friction material makes very little noise during the cyclic wear test as does the organic one, particularly at the moment of stopping and starting the magnetic clutch.

Figure 19 shows the contacting surfaces of each friction material and that of the surface (S15C low carbon steel) after a wear test in which the surface underwent 1000 cycles. No seizure or wear phenomenon is observed at the contacting surface of not only the friction material but also the surface material using the mechanically-alloyed sintered friction material. On the other hand, the conventional sintered friction material does have seizures and damaged areas at the contacting surface, and the thermally damaged areas are observed at the surface of the organic friction material. The same results, of course, are fully recognized in various durability and reliability tests by our customers and it is found that this mechanically-alloyed sintered friction material has the ability to realize a more compact magnetic clutch system with high performance.

CONCLUSIONS

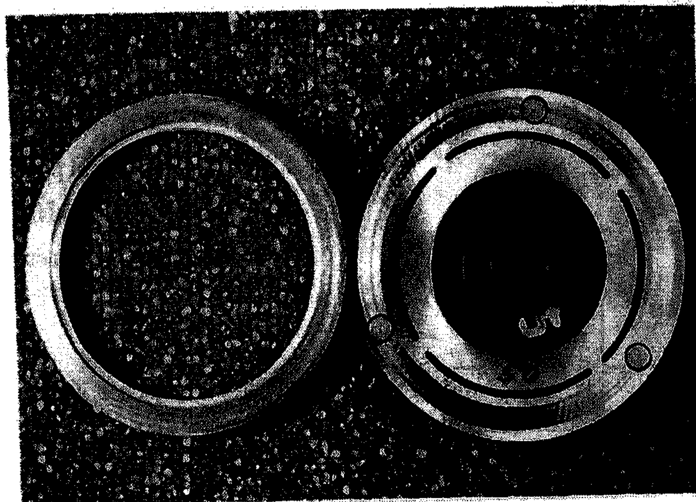
1.Wear and/or seizure phenomena of the conventional typical sintered friction material including some hard particles usually occur when contacting a surface material under severe friction conditions.

2. There are gaps at interfaces between the hard particles and the matrix, and such hard particles are poorly bonded to the matrix. The hard particles therefore, detach from the matrix and the seizure phenomena occur.

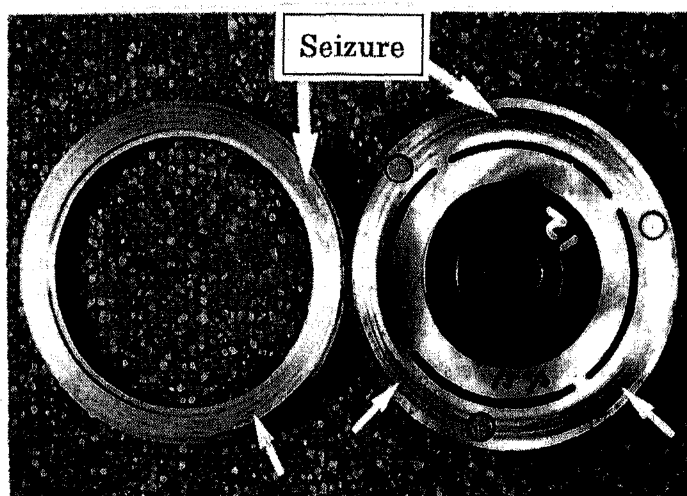
3. The integrated composite sintered material consists of the mechanically-alloyed copper-based powders which include the hard particles dispersed uniformly in the matrix. There is a diffusion layer at the interface between hard particles and the matrix in the mechanically-alloyed powders, so that finer hard particles with less 10 μ m in a diameter are securely bonded to the matrix.

4. This new sintered friction material shows superior anti-seizure properties, wear resistance and mechanical properties to the conventional sintered one, and it is also less abrasive to the surface material because no fine hard particle detach from the matrix when contacting the surface material. Furthermore, this new friction material shows a higher friction coefficient with a constant 0.4~0.45 than that of the conventional one.

5. In the application of this advanced sintered friction material to a magnetic clutch for automotive air-conditioners, it also shows a higher friction coefficient and excellent anti-seizure properties. It is, therefore, expected that this friction material will provide a more compact magnetic clutch system with high reliability.



(a) MA sintered friction material



(b) Conventional sintered friction material



(c) Organic composite friction material 20mm

Figure 19. Contacting surfaces of each friction material and that of surface material (Low carbon steel/S15C) after a wear test with undergoing 1000 cycles. (Left; Friction material, Right; Surface material)

971003

Evaluation of the Corrosion Rate of Zn Coated Steel Sheets for Automotive Body Use

Shigeo Kurokawa, Morishige Uchida, Chiaki Kato, Kazuo Mochizuki, and Toshiyuki Kato
Kawasaki Steel Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The corrosion mechanism of zinc coated steel sheets in automotive bodies was studied in field vehicle tests and several types of accelerated tests. Perforation corrosion starts in unpainted areas of lapped parts, and proceeds in the following steps: i) galvanic protection by the Zn coating, ii) protection by corrosion products, and iii) corrosion of the steel substrate and perforation. Although the corrosion processes were the same in all the cases tested, the corrosion rate depended significantly on the environment, such as atmospheric exposure conditions and the part of the automotive body. In accelerated corrosion environments, Zn coating is largely ineffective against perforation corrosion because galvanic protection and protection by corrosion products cannot be maintained over the long term. Conversely, coating is highly effective under actual atmospheric conditions which produce slow corrosion rates because galvanic protection and protection by corrosion products continue to protect the steel substrate over an extended period. Accelerated tests may be an accurate predictor of corrosion in cold rolled steel, but may not be representative of corrosion in coated products. On the other hand, cold rolled steel can be used to extrapolate the rate of corrosion in coated products. Therefore to evaluate the effects of Zn coatings quantitatively, the authors suggested an index of for corrosion resistance relative to cold rolled steel sheets under each atmospheric conditions. It is considered possible to estimate actual perforation corrosion life by using this index.

INTRODUCTION

Corrosion of automotive bodies became a problem after the widespread use of de-icing salts on roads was adopted in North America and Europe. In the late 1970s, Zn coated steel sheets were introduced to protect automotive bodies from corrosion. However, corrosion can be divided into two significantly different types. One may be characterized as outside-in corrosion, caused by paint chipping damage to outer panel surfaces, resulting in under-the-film corrosion and / or scab corrosion. This type is generally called cosmetic corrosion. The second is inside-out corrosion caused by salty or muddy water penetration into places that do not dry readily, such as panel joints and the inner surfaces of closed structures (perforation corrosion).

Progress of cosmetic corrosion is often discussed in terms of pH changes of the under paint space (1,2). As for perforation corrosion, extensive analytical studies have been conducted in the lapped portions of dismantled field vehicles (3,4) to elucidate the corrosion mechanisms. Other studies have provided information about the effects of corrosive environments (5,6) and the effects of gap in the lapped parts (7), which was obtained from cyclic corrosion tests. These studies were usually qualitative, and information on the corrosion rate of Zn coated steel was not available in the literature.

In this work, corrosion rate measurements were conducted in various corrosive environments using Zn-Ni alloy coated steel and cold rolled steel. Based on the results, the authors suggested an index of the corrosion resistance of coated steel extrapolated from cold rolled steel sheets under various atmospheric conditions as a means of evaluating the effects of Zn coatings quantitatively. It is considered possible to estimate actual perforation corrosion life using this index.

EXPERIMENT

MATERIALS- Zn-Ni alloy electroplated steel sheets and cold rolled steel sheets were prepared for the tests. The alloy contained 11.2% Ni, and the coating weight was 21g/m². Front doors of a 4-door sedan type vehicle were manufactured from these two kinds of steel sheets. The door was chosen because the door hem flange is categorized as a part which is subjected to severe corrosive conditions and is also an exchangeable part. Based on a domestic Japanese specification, the door samples were phosphated, painted and sealed then exposed to the following corrosion tests.

FIELD VEHICLE TEST-In the field test vehicle, the doors were installed in a monitor vehicle of the four door sedan type which was driven for 8 years, from October, 1985 to November, 1993 in Naha-City, Okinawa Prefecture.

Okinawa Prefecture is classified as a subtropical marine environment, with high temperature and high humidity, and a large amount of dispersed sea salt particles. The mileage driven was approximately 80,000km. The vehicle was used for business purposes and was parked in a garage. The details of this test have been described in our previous paper (8).

SEA SHORE EXPOSURE TEST-Another set of doors was exposed to the atmosphere at a coastal exposure test site located 10m from the ocean for 5 years from May, 1985 to July, 1990. The site was located on the coast in southeast Okinawa.

MODIFIED VOLVO TEST-Another set of doors was exposed at Chiba Works of Kawasaki Steel Corp. The test site is located 10m from the ocean. In this test, sea water was sprayed for 10min. twice a week (Modified Volvo Test). The test was conducted from 1985 to 1988.

CYCLIC CORROSION TEST-For cyclic corrosion tests, bottom portions of the door (hem) were cut away and placed in a cyclic corrosion test chamber. The cut edge was mended with a seal. The following two cycles were used in the test.

CCT1: Salt spray (35 ℃, 5%NaCl, 5hrs.) / Dry (70 ℃, 7hrs.) / Wet (50℃, 95%RH, 12hrs.) 1cycle/day

CCT2: Salt spray (35 ℃,0.1%NaCl, 6hrs.) / Dry (50 ℃, 3hrs.) / Wet(50 ℃, 95%RH, 15hrs.) 1cycle/day

PERFORATION CORROSION EVALUATION-The inner and outer steel sheets were taken from the door hem. Paint remover (alkaline type) and ammonium citrate were used to remove the paint and corrosion products, respectively. Various area of the sheets were measured to determine the maximum thickness reduction by corrosion. From some areas of the door hem, cross sectional samples were cut and mounted in plastic holder to study the progress of corrosion.

CORROSION PRODUCT ANALYSIS-The corrosion products removed from the door hem flanges were analyzed by X-ray diffraction. The micro X-ray diffraction method was applied for the analysis of the inner and outer steel sheets taken from the door hem flange within the range of parts of 100μm in diameter.

EXPERIMENTAL RESULTS

CORROSION PROGRESS (sheet thickness reduction) AND CHANGES IN CORROSION PRODUCTS- It is generally believed that perforation corrosion starts from closed structures, such as lapped and hemmed portions, where little or no paint can be applied. Many automobile manufacturers employ sealers, adhesives and wax injection to reduce penetration by corrosive water. In this experiment, only the sealer was applied. This was done to simulate defects of the adhesive and imperfect wax injection. The sealer was necessary to avoid the corrosion at the cut edge. In the field vehicle test, obvious perforation corrosion (the sheet thickness was 0.8mm) was recognized after 2.5 years of the test on the cold rolled steel sheet. On the Zn-Ni steel sheet, the maximum thickness reduction was only 0.18mm after 8 years. Figure 1 shows a schematic diagram of the corrosion products identified by the micro X-ray diffraction technique. In the figure, Locations 1-4 correspond to the order of corrosion progress, which was suggested by the amount of thickness reduction (figure 2) and kinds of corrosion products. Figure 2 was a cross section of an outer panel facing the inside of the hem inside. The sample was prepared after taking the flange apart. The numbers correspond to the portions indicated in Figure 1. Location 1 showed only white Zn corrosion products ($ZnCl_2 \cdot 4(OH)_2$), suggesting the beginning of corrosion. Location 2 showed some Fe corrosion products with less amount of Zn corrosion product, although the thickness reduction was less than 0.01mm. Location 3 was covered with β- and γ-FeOOH, Fe_3O_4 and ZnO. By this time, the Zn-Ni coating had disappeared and the thickness reduction had become measurable. At Location 4, significant thickness reduction was measured, and no Zn corrosion products were found. It is reported that Fe corrosion products, such as β- and γ-FeOOH, and amorphous, tend to change to stable α-FeOOH (9), which was found at Location 4. From these results, it can be said that the corrosion started at Location 4, where corrosive water was obtained from inside the door. Corrosion then proceeded along with the gap between the inner and outer steels toward Locations 3, 2, and 1, in that order.

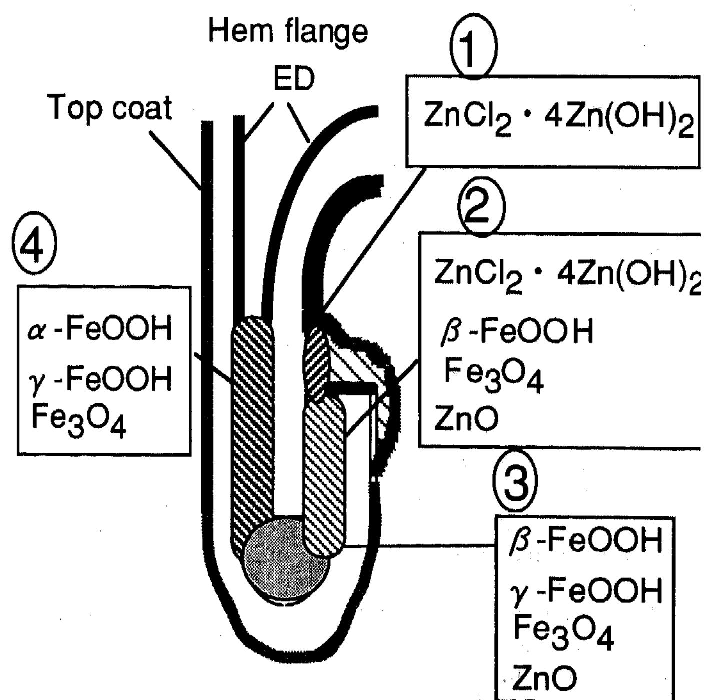


Figure 1 Schematic of corrosion products in door hem flange of Zn-Ni alloy coated steel after field test for 8 years

EFFECTS OF CORROSION TEST CONDITION ON THE CORROSION PRODUCTS-Similarly to the studies conducted in Figure 2, thickness reductions and the identification of corrosion products were conducted on the hem flanges tested in all environments. The corrosion products were the same as those found in the door hem of the field monitor vehicle. Moreover, the cross-sectional studies indicated similar corrosion behaviors to those shown in Figure 1. This suggests similar corrosion mechanisms under all the test conditions.

EFFECTS OF CORROSION TEST ENVIRONMENTS AND Zn COSTING ON CORROSION RATE-Figures 3 and 4 show the variation of thickness reduction with time in the various tests. The thickness reductions of the cold rolled steel are linear with time in the all tests, and the extrapolations of the linear line pass through the origin. The thickness reductions on the Zn-Ni alloy are also linear with time; however, the lines do not pass through the origin, indicating the existence of an incubation period. In the incubation period, the thickness was reduced within the Zn coating, and only Zn corrosion products were identified by X-ray diffraction analysis. After the incubation period, the thickness reduction rate was rather low up to a 0.02mm thickness reduction. In this period, the corrosion products consisted of both the Zn and Fe types. Above 0.02mm, the reduction rate became high and only Fe corrosion products were detected. This corrosion behavior of the Zn-Ni alloy is summarized in Figure 4 for the case of the field vehicle test. These findings indicate that perforation corrosion proceeds in the following steps:

1) The steel is completely protected by the galvanic reaction of the Zn coating and by Zn corrosion products.
2) The steel thickness reduces at a low rate. The steel surface is covered with fewer protective Zn and Fe corrosion products.
3) The steel thickness reduces at a high rate. Only Fe corrosion products are present on the steel surface.

Table 1 shows the corrosion rate in each step. It is obvious that the corrosion rate increases and the effect of the Zn coating decreases with an acceleration of the corrosive environment. Table 2 shows the corrosion acceleration ratios when the corrosion rate of the Zn-Ni alloy in the field vehicle test is considered to be 1. The ratio is obtained for each step of the corrosion process, although the cold rolled steel has only one step. These data indicate that corrosion acceleration is more effective in increasing the corrosion rate with the Zn-Ni alloy than with the cold rolled steel. For instance, CCT 1 accelerates the corrosion rate of the Zn-Ni

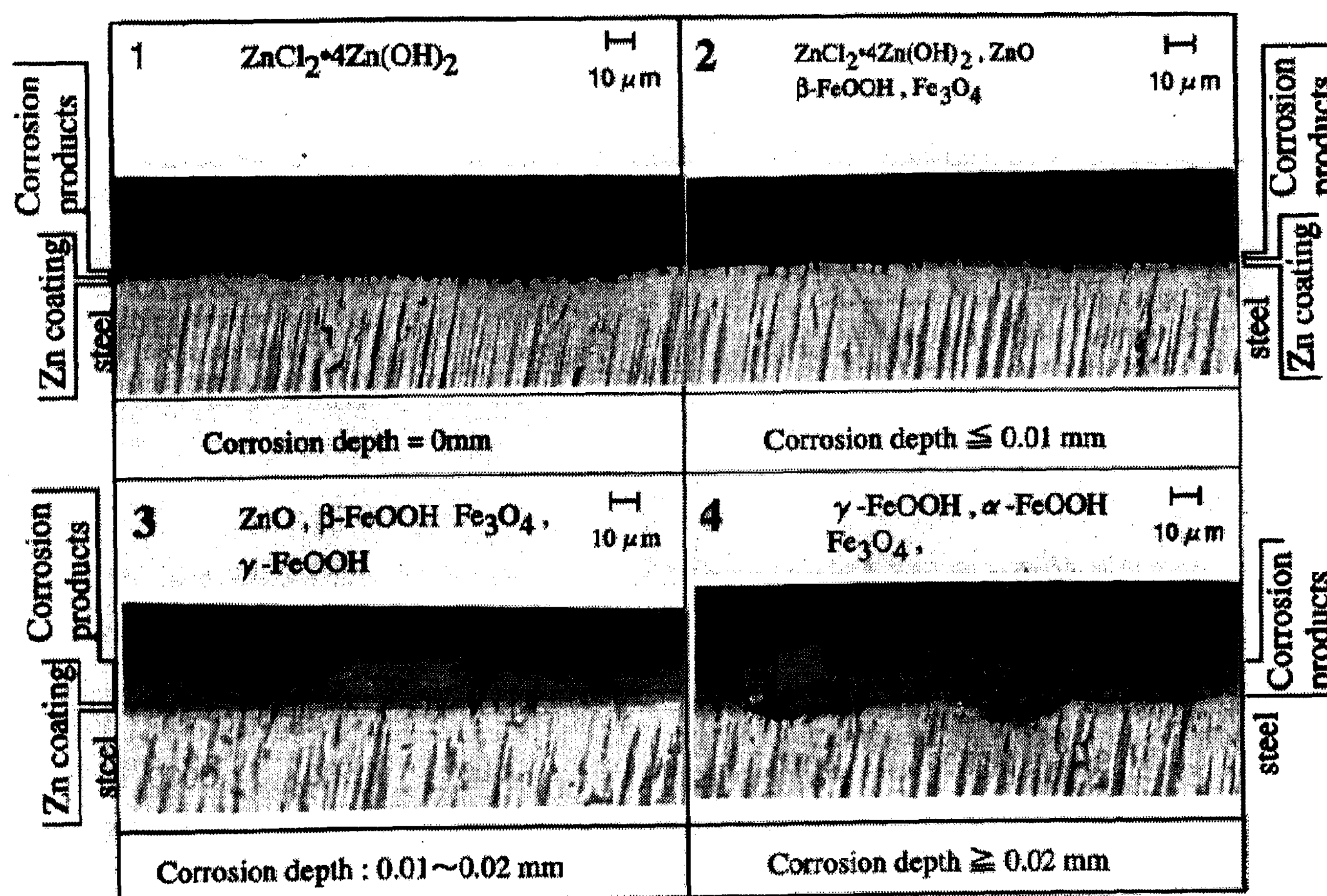


Figure 2 Correlation between corrosion products and corrosion depth at several stages of corrosion on Zn-Ni alloy coated steel.

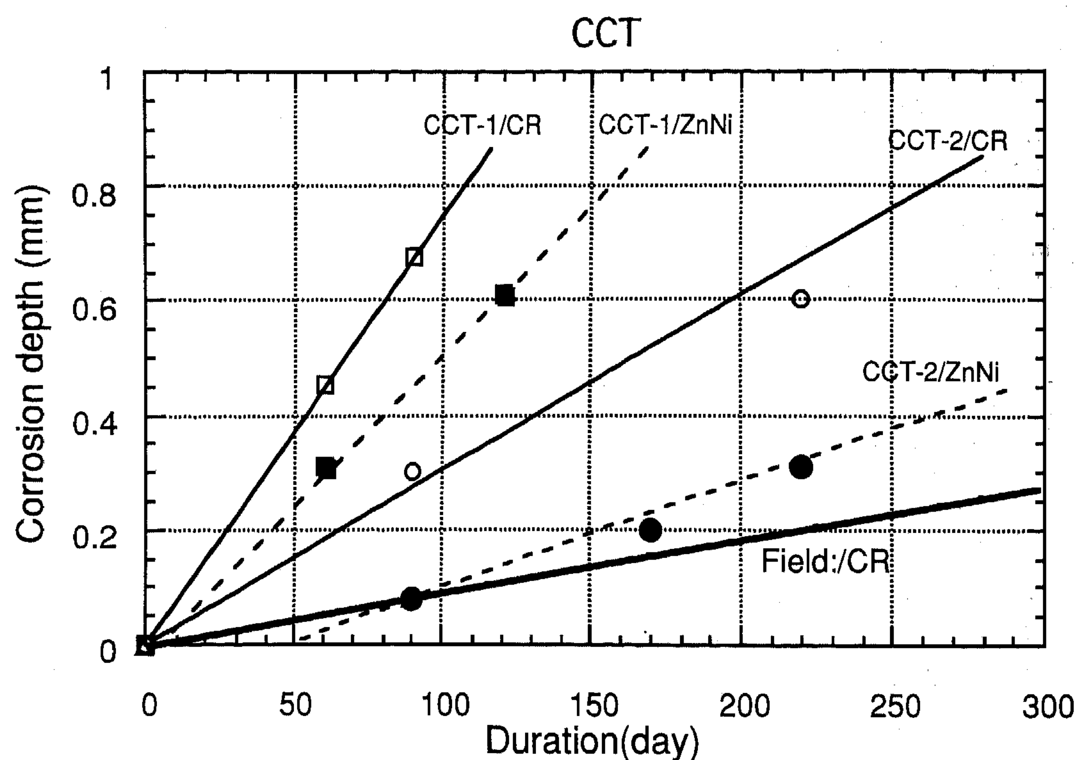


Fig.3 Comparison of corrosion behavior of Zn-Ni alloy coated steels (Zn-Ni) and cold rolled steels (CR) in cyclic corrosion tests

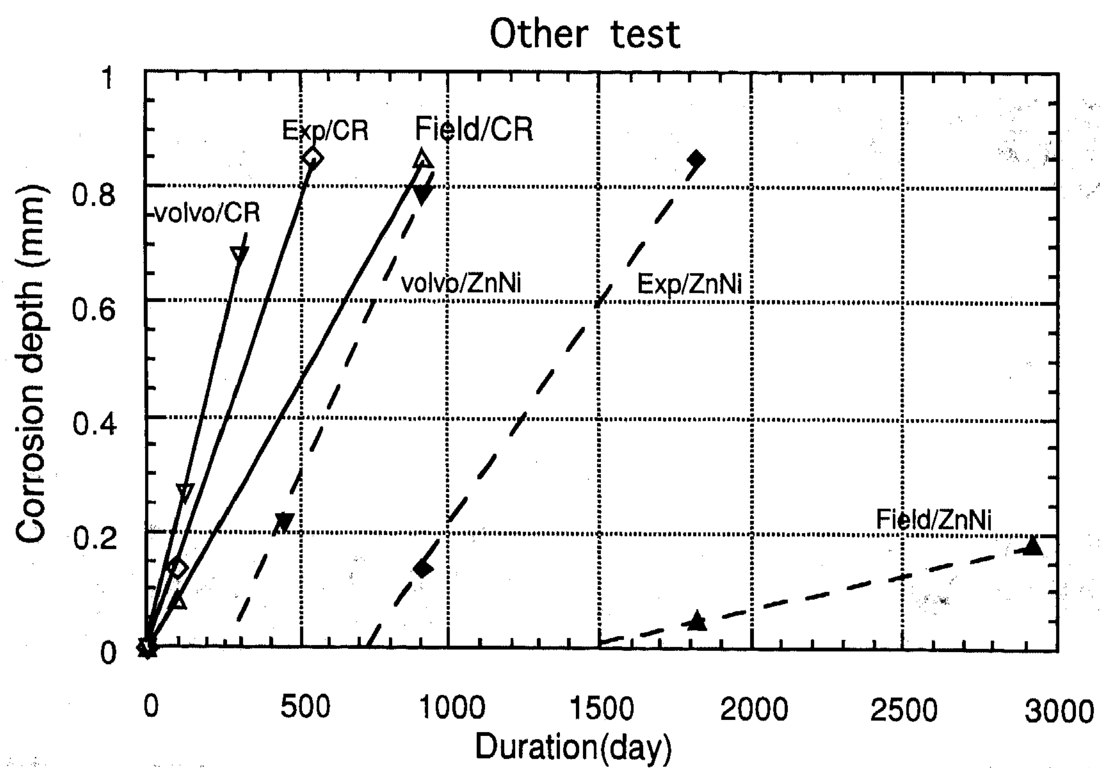


Fig.4 Comparison of corrosion behavior of Zn-Ni alloy coated steels (Zn-Ni) and cold rolled steels (CR) in various corrosion tests

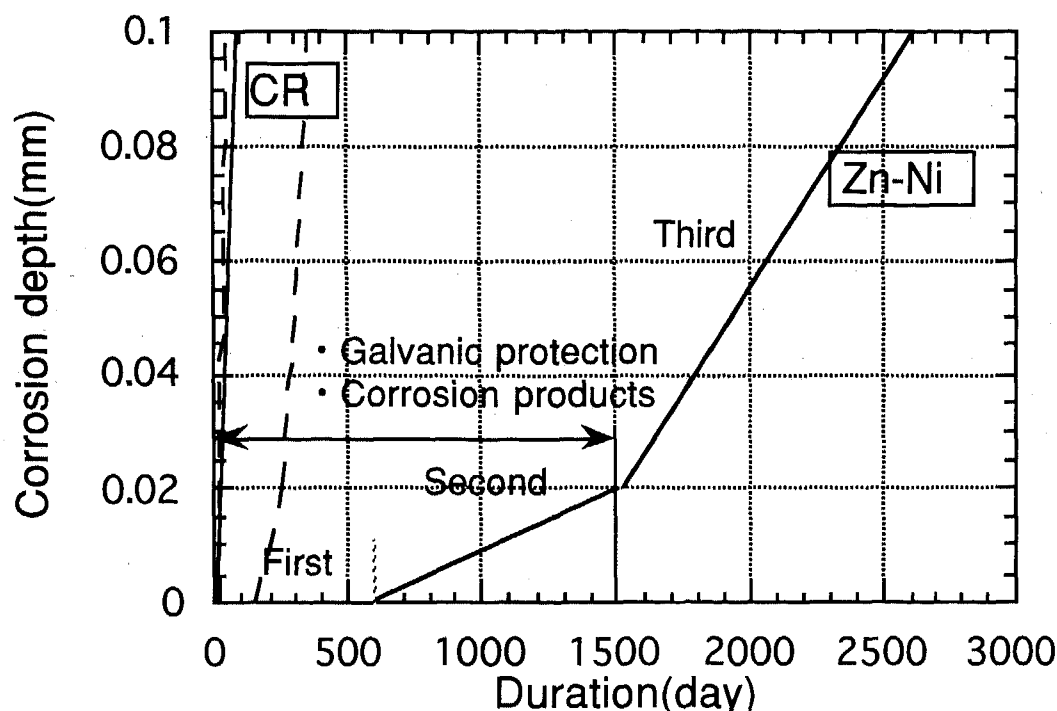


Fig.4. Estimated corrosion behavior of Zn-Ni alloy coated steel applied to door hem

alloy in the field vehicle test 60 to 110 times, where as the acceleration is only 8 times with the cold rolled steel. Table 3 shows the corrosion rate of the Zn-Ni alloy when the rate on the cold rolled is 1. This clearly shows the effect of the Zn coating, which is most obvious in the first corrosion step with the field vehicle test.

DISCUSSION

PERFORATION CORROSION MECHANISMS OF Zn COATING-As shown in Figure 4, the significance of the corrosion mechanism of the Zn coated steel is the existence of an incubation period (first step) and initial slow corrosion (second step) due to the presence of protective Zn corrosion products. In these steps, the corrosion rate is influenced by the Zn coating weight and alloying element. In the third step, corrosion is governed by only the differential aeration cell, and the effect of the Zn coating is minimal. In this step, the Zn coated and uncoated steel were expected to show the same corrosion rates. However, the rate was slower with the Zn coated steel. This may be explained by the existence of a small amount of Zn in the corrosion products, which would lower the corrosion reaction rate or retard the movement of corrosion products (10).

INFLUENCE OF CORROSION TEST CONDITIONS OF THE CORROSION RATE-In order to discuss perforation corrosion more quantitatively, the corrosion rate, in Section 3.3 was converted to a corrosion rate ratio, which is defined as follows:
Corrosion resistance ratio
={Time to reach a given corrosion depth in coated steel sheet}/{Time to reach the same corrosion depth in cold rolled sheet}

The calculations were conducted at the corrosion depths of 0.02mm (effect of Zn coating) and of 0.8mm (perforation depth). The results are shown in Table 4. The values are small in the accelerated test. The effect of the Zn coating was evaluated to be 72 in the field vehicle, but it was only 2.2 in the CCT 1. Since the corrosion products, and thus the corrosion mechanism, are the same in both tests, the difference may be attributed to the time duration of the effective corrosion current (i.e. wet time period) and/or by the intensity of the corrosion current, influenced by the resistance of corrosion current flow through the corrosion products. Moreover, the high NaCl concentration used in the accelerated test increases ionic conductivity, and thus the corrosion rate.

The corrosion resistance ratios evaluated at perforation were 9.2 in the field vehicle test and 1.5 in the CCT-1, and were smaller than those at the depth of 0.02mm. The small values are caused by the fast perforation corrosion rate of the steel substrate, which reduces the effect of the Zn coating evaluated at 0.02mm. The effect of acceleration was also seen in the ratios evaluated at perforation.

In general, cosmetic corrosion is evaluated under relatively dry condition (6). In such dry environments, slight variations in the test conditions do not influence the corrosion rate. Thus, the corrosion rate in the field condition

Table 1 Corrosion rate in the various environments (m/year)

	Zn-Ni20			CR
	First	Second	Third	
Field test	1.83×10^{-6}	8.12×10^{-6}	2.65×10^{-5}	3.60×10^{-4}
Exposure test	2.74×10^{-6}	1.94×10^{-5}	2.81×10^{-4}	5.17×10^{-4}
Volvo test	7.30×10^{-6}	5.64×10^{-5}	4.40×10^{-4}	7.60×10^{-4}
CCT-2	3.65×10^{-5}	2.95×10^{-4}	6.72×10^{-4}	1.11×10^{-3}
CCT-1	1.10×10^{-4}	8.86×10^{-4}	2.03×10^{-3}	2.81×10^{-3}

Table 2 Degree of corrosion acceleration

	Zn-Ni20			CR
	First	Second	Third	
Field test	1	1	1	1
Exposure test	1.5	2.4	10.6	1.4
Volvo test	4.0	6.9	16.6	2.1
CCT-2	20.0	36.3	25.4	3.1
CCT-1	60.1	109.1	76.6	7.8

Table 3 Ratio of corrosion rate relative to cold rolled steel

	Zn-Ni20			CR
	First	Second	Third	
Field test	1/197	1/44	1/14	1
Exposure test	1/189	1/27	1/1.8	1
Volvo test	1/104	1/13	1/1.7	1
CCT-2	1/30	1/3.8	1/1.6	1
CCT-1	1/26	1/3.2	1/1.4	1

can be predicted easily by exposure tests and cyclic corrosion tests. However, the predictions are not as easy for the perforation corrosion rate, because the corrosion rate is significantly influenced by the corrosion condition. The results obtained in this study provide some information for the prediction of the perforation corrosion resistance.

An approximate time to reach the perforation corrosion depth can be calculated from the time required for perforation of cold rolled steel and the corrosion resistance ratio of the specific Zn coated steel sheet. In other words, the perforation corrosion rate of the cold rolled steel obtained under the specific corrosion conditions (the rate varying with the location and shape of the installed part, amount of corrosive water, amount of paint available, and others) is compared with the rate shown in Table 2 to estimate the severity of the corrosion condition. The door hem used in

Table 4 Comparison of corrosion resistance ratio of 0.02mm depth corrosion and perforation corrosion in various environments

	Corrossion resistance ratio*	
	Depth=0.02 mm	Perforation=0.8mm
Field Vehicle	72	9.2
Exposure test	58	3.4
Volvo test (modified)	35	2.6
CCT-2	7.9	1.7
CCT-1	2.2	1.5

* Corrosion resistance ratio

$$= \frac{\text{[Time to reach a given corrosion depth in coated steel sheet]}}{\text{[Time to reach the same corrosion depth in cold rolled sheet]}}$$

this test is classified as a rather mild corrosion condition (4). The quarter rear and hood outer may be severer than the hem tested in this work and may be of the same severity as the Okinawa exposure condition. If the Zn-Ni alloy is applied at a coating weight of 20 g/m^2 in a specific part, one can easily predict the corrosion depth by using the corrosion resistance ratio shown in Table 4. Other Zn coatings and coating weights are applied in these and other parts, and information on these effects must be obtained. In the future, such information can be utilized for the selection of suitable Zn coatings and coating weights to obtain sufficient corrosion protection.

SUMMARY

Perforation corrosion starts from closed structures, such as lapped and hemmed portions where paint is difficult to apply. Thus, the perforation corrosion rate is strongly influenced by the degree of corrosion severity, which is difficult to characterize or define. In this work, corrosion tests were conducted under various conditions with Zn-Ni alloy coated steel and cold rolled steel. Based on the results, a corrosion rate prediction method was suggested.

Through the course of these studies, it was found that the significance of Zn coating was to provide an incubation period and reduce the corrosion rate at the beginning of corrosion. This Zn coating effect is minimized under accelerated (severe) corrosion. The corrosion rate of the Zn coating can be estimated from the above relation and the corrosion rate of the cold rolled steel in the same corrosion condition.

REFERENCES

1) K. Hayashi, "Under-film Corrosion of Precoated Steel Sheets for Automobiles", J. of The Adhesion Soc. of Japan, 26, 5, (1990) 194-201
2) K. Hayashi, Y. Ito, C. Kato, and Y. Miyoshi, "Under-film Corrosion Behavior of Zinc Alloy Coated Steel Sheets for Automobiles", Testu-to-Hagane, 76, 8 (1990) 115-122
3) Y. Miyoshi, Y.I to, L.E.Soreide, D.D.Davidson, "Corrosion Protection Function of Zinc Coating on Steel Sheet, as Revealed by the Investigation of Field Cars and Laboratory Test", 2nd International Conference on Zn and Zinc Alloy Coated Steel Sheet GALVATECH '92, 528-534
4) A. W. Bryant et al. ,"U.S.Automotive Corrosion Trends at 5&6 years", SAE Technical Paper Series 892578 (1989)
5) S. Kurokawa, T. Ban, K. Yamato, and T. Ichida ,"A Study on Cosmetic and Perforation Corrosion Test for Automotive Steel Sheets", Testu-to-Hagane, 72, 8 (1986) 223- 230
6) S. Kurokawa, K. Yamato, T. Ichida, "A Study on Cosmetic and Perforation Corrosion Test Procedures for Automotive Steel Sheets", paper no.396, The NACE Annual Conference and Corrosion Show, Cincinnati, Ohio, USA, CORROSION/91
7) S. Wakano, and M. Nishimura, "Corrosion Resistance of Several Coated Steels through Simulation Tests", Sumitomo Metals, 41, 2 (1989) 11-18
8) M. Uchida, et al. ,"Corrosion Behavior of a Vehicle Submitted in Service in Subtropical Marine Environment

of Okinawa Island for Eight Years", SAE TechnicalPaper Series 960021(1996)
9) T. Misawa, "Research Status and Unsolved Problems in Rusting of Iron and Steel", Corrosion Engineering, 32, 11 (1983) 657-667

10)E. Johansson and J. Gullman, "Corrosion Study of Carbon Steel and Zinc-Comparison Between Field Exposure and Accelerated Tests", ASTM Spec Tech Publ (Am Soc Test Mater) No.1239, Page240-256 (1995)

971005

Stabilization Requirements for T409 (UNS S40900) Ferritic Stainless Steel

Ivan A. Franson and J. D. Fritz
Allegheny Ludlum Corp.

Copyright 1997 Societv of Automotive Engineers, Inc.

ABSTRACT

Type 409 (UNS S40900) ferritic stainless steel was developed more than 30 years ago specifically for automotive exhaust systems where it has subsequently found extensive use. Performance of this 11 Cr- Ti alloy for the most part has been good. However, recent failures of exhaust system components demonstrate that the T409 alloy can become sensitized in weld heat-affected zones and suffer intergranular corrosion. Results of a broad study of commercial heats illustrate that weld heat-affected zones of T409 are stabilized when titanium plus columbium or titanium alone meets the criterion:

$$Ti + Cb \geq 0.08 + 8(C + N)$$

Experimental procedures used to establish this criterion are described, including development of a test to detect sensitization. Also the benefits of stabilizing with titanium in combination with columbium are discussed.

INTRODUCTION

The T409 alloy (UNS S40900) is a low chromium ferritic stainless steel which is used quite extensively in automotive exhaust systems. The ASTM compositional requirements[1] for this alloy are presented in Table 1.

Recently, it has been demonstrated that applications which expose welded T409 alloy to aggressive environments can result in intergranular corrosion (IGC) of the weld heat-affected zones. This form of corrosion which involves a rapid attack of the grain boundaries, can be initiated by exhaust system condensate. Figure 1 shows the cross section of an automobile muffler weldment which joins the muffler's cap, nipple, and tube. As shown, this cross section suffered IGC in the weld heat-affected zone (HAZ) of the tube, which resulted in the muffler prematurely falling off the automobile.

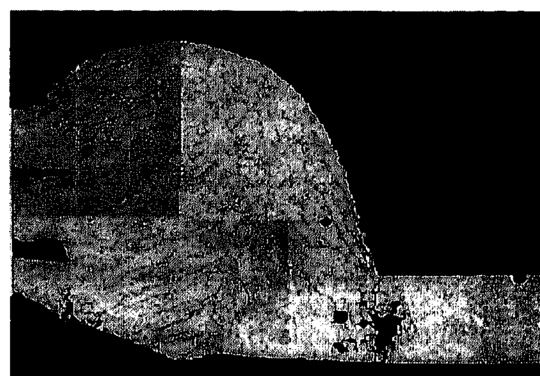


Figure 1. Weldment from failed muffler (8 X Magnification)

The mechanism which is responsible for producing a susceptibility to IGC in ferritic stainless steels is very similar to that which is operative in the austenitic alloys[2]. The precipitation of chromium

Table 1.
ASTM A240 Specification
For T409 stainless Steel

C max.	Mn max.	P max.	S max	Si max.	Cr	Ni max.	Other
0.08	1.00	0.045	0.030	1.00	10.5 - 11.75	0.50	Ti 6 x C min; 0.75 max

carbides (and chromium nitrides in the case of the ferritic alloys) at the grain boundaries of the alloy results in the depletion of chromium from adjacent regions. Because of this loss of chromium, the depleted grain boundaries become sites for rapid preferential attack when exposed to certain acidic environments. This process of producing chromium-depleted regions due to the precipitation of chromium-rich second phases is termed "sensitization".

For the ferritic stainless steels sensitization will typically occur after being heated to temperatures of 925°C or higher. At these higher temperatures there is a substantial increase in the solubility of carbon and nitrogen in the ferrite crystal structure. Upon cooling the solubility is strongly reduced resulting in the precipitation of chromium carbides ($M_{23}C_6$) and/or chromium nitrides (Cr_2N). In the temperature range of 925 to 700°C the rate of Cr diffusion is quite rapid which tends to replenish any Cr depleted regions. However, in the temperature range of 700 to 400°C the precipitation of carbides and nitrides will produce chromium depletion at grain boundaries. Hence any thermal treatment , such as welding, which takes the alloy to temperatures of 925°C or higher followed by cooling through the sensitization temperature range of 700 to 400°C could potentially produce sensitization.

In order to avoid the loss of corrosion resistance due to sensitization, commercial ferritic stainless steels are usually stabilized with additions of Ti, Cb, or Ti + Cb and alloy producers strive to keep the C and N levels to a minimum. The amounts of stabilizing elements required to prevent the sensitization phenomenon depends on how much C and N are present in the alloy. The more C and N which are present the more Ti, Cb, or Ti + Cb is required to prevent susceptibility to IGC.

As shown in Table 1, the ASTM stabilization requirement for T409 mandates the Ti% to be equal to (6 x C%) with a maximum Ti content of 0.75 wt%. The obvious shortcoming of this stabilization requirement is that it ignores the detrimental contribution of N which in a ferritic alloy is isomorphous with C. This results in the undesirable situation of being able to meet the ASTM compositional requirements and still be susceptible to IGC. This is demonstrated in Figure 1 where the muffler tube suffered IGC in the weld HAZ during routine automotive service even though the Ti%/C% ratio of 10 for this tube substantially exceeds the ASTM minimum of 6.

Because of this potential problem with the existing alloy specification, Allegheny Ludlum decided to undertake a study to determine what amount of stabilization is required to prevent IGC of welded T409 sheet material. In order to do this it was necessary to have a test method which would detect the susceptibility to IGC in the T409 alloy. Currently, no standard test methods are defined for low Cr ferritic stainless alloys. The ASTM A763-93 Standard Practices for Detecting Susceptibility to Intergranular Attack in Ferritic Stainless Steels does specify various test methods for evaluating higher Cr-bearing (>16%) ferritic stainless steels but these test methods are too aggressive for use with lower Cr alloys.

Although no standard test method exists for detecting sensitization of T409 material, various methods applicable to low Cr ferritic stainless steels have been reported in the literature[3,4,5]. These test methods are summarized in Table 2. The current investigation evaluated methods I and IV in Table 2 for use with commercial T409 alloys. Method IV was then used to determine the stabilization requirements of welded T409 sheet material.

EXPERIMENTAL PROCEDURES

Two of the reported test methods (I & IV) outlined in Table 2 were chosen for further investigation

Table 2. Test Methods Used For Detecting IGC Susceptibility of Low Chromium Ferritic Stainless Steels

Test ID	Test Solution	Test Temp.	Exposure Time	Reference
I	65% HNO_3	60°C	48 Hours	R. A. Lula & J. A. Davis[3]
II	50% H_2SO_4 6% $CuSO_4$ Cu shot	60°C	2 Hours	R. A. Lula & J. A. Davis[3]
III	$Fe_2(SO_4)_3$ 50% H_2SO_4	Boiling	0.5 Hours	R. A. Lula & J. A. Davis[3]
IV	0.5% H_2SO_4 6% $CuSO_4$ Cu shot	Boiling	20 Hours	T. Divine & B. Drummond[4]
V	0.1 M H_2SO_4 0.4M $CuSO_4$ 1000 ppm KSCN	20°C	Electrochemical Potentiokinetic Reactivation Method	S. Frangini & A. Mignone[5]

to determine if they are appropriate methods for testing T409 production heats. Since the mechanism of IGC involves preferential attack of the Cr-depleted regions at grain boundaries, any test solution which will detect IGC must show a general corrosion rate which is a strong function of the Cr content of the alloy. A series of experimental Fe-Cr alloys was prepared in order to look at how the corrosion rate varies with Cr content using test methods I and IV. The Cr contents of the experimental heats and three production heats used in this experiment are summarized in Table 3. Samples of these compositions in the solution annealed (nonsensitized) condition were exposed to test methods I and IV.

Table 3. Composition of Samples Used to Evaluate IGC Test Method

Sample ID	%Cr
RV279A*	4.67
RV279B*	5.60
RV279C*	6.54
RV280A*	7.82
RV280B*	8.81
RV280C*	9.80
856218**	11.53
848228**	11.54
804004N**	11.43

* Experimental Fe-Cr Heats;
** Production T409 Heats

The results of this testing, summarized in Figure 2, show that the Cu/ $CuSO_4$/H_2SO_4 test (method IV) produces a much larger increase in the corrosion rate with decreasing Cr content than the HNO_3 test. Because of this finding all subsequent work was done using method IV.

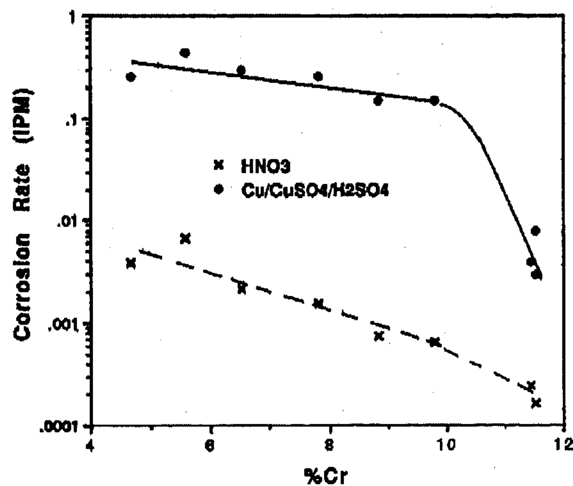


Figure 2. Plot of the General Corrosion Rate vs % Cr.

In order to verify that method IV does detect sensitization of commercial T409 alloys an experiment was conducted using three additional production heats which were stabilized to different levels based on the ratio, Ti/ (C + N). Samples from these three heats were first heat treated to 1200°C to put all C and N into solution and then aged for various times at 600°C. After a specified aging time each sample was tested using method IV. Samples were weighed before and after the test and the corrosion rates determined in inches per month (IPM). The results of this testing are summarized in Table 4. A graph of the corrosion rate versus aging time for heats A and B is presented in Figure 3.

Table 4. Cu/$CuSO_4$/H_2SO_4 Test T409 Production Heats Samples Heated to 1200°C/1Hr/WQ and Aged at 600°C for Various Times

HEAT ID	Aging Time	Ti/(C+N)	Corr. Rate (IPM)
A	0.0 hrs	10.5	0.0060
B	0.0 hrs	5.3	0.0055
C	0.0 hrs	3.8	0.0605
A	0.5 hrs	10.5	0.0088
B	0.5 hrs	5.3	0.0191
C	0.5 hrs	3.8	Samples Disintegrated
A	1.0 hrs	10.5	0.0062
B	1.0 hrs	5.3	0.0098
C	1.0 hrs	3.8	Samples Disintegrated
A	10.0 hrs	10.5	0.0060
B	10.0 hrs	5.3	0.0064
C	10.0 hrs	3.8	0.0300

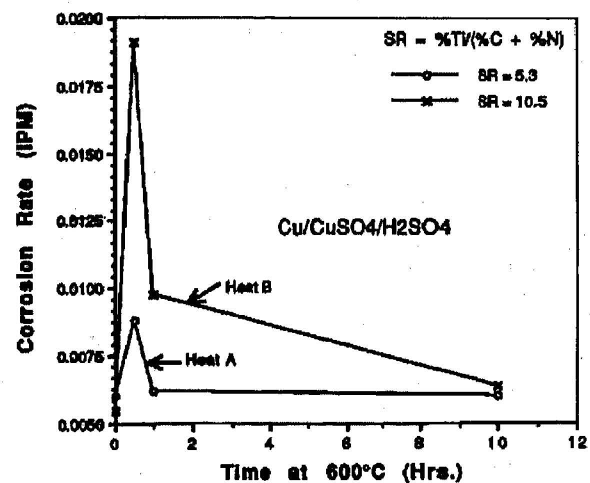


Figure 3. T409 samples held at 1200°C/1hr./WQ and aged at 600°C for various times.

In order to determine if the observed increase in corrosion rate for the short aging times was in fact the result of intergranular attack, the cross sections of the exposed samples were examined. The cross sections from Heat B are shown in Figure 4. These cross sections demonstrate that intergranular attack does occur after the 30 minute aging cycle and the resistance to intergranular attack is restored by the 10 hour aging cycle. As discussed below these observations indicate

that the test method IV does indeed detect susceptibility to IGC with T409 production heats.

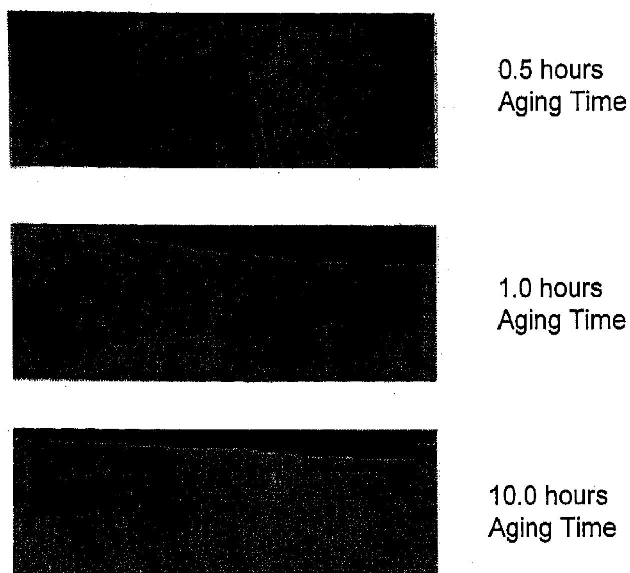


Figure 4. Cross Sections of Heat B after exposure to the $Cu/CuSO_4/H_2SO_4$ test solution. (95X mag)

A sample of the same kind of weldment as that shown in Figure 1 was cut from a muffler which was never put into service. This muffler was chosen because the degree of stabilization of the material used to fabricate the muffler was rather low with a Ti/(C + N) ratio of 5.3, 6.6, and 5.7 for the cap nipple and tube respectively. This sample was tested using method IV to see if the test could detect IGC. The cross section of this muffler weldment after the test exposure is shown in Figure 5.

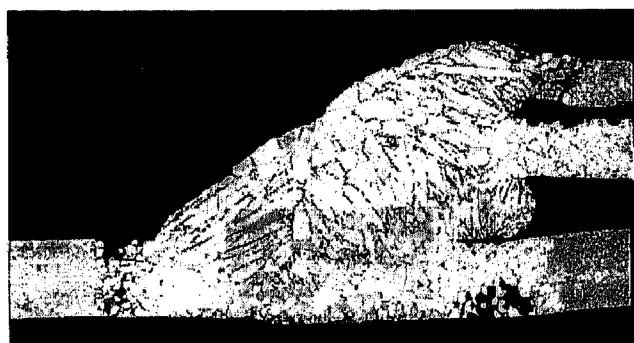


Figure 5. Muffler Weldment with no service. Microstructure Following Exposure to $Cu/CuSO_4/H_2SO_4$ (8X Magnification)

After having qualified a test method for determining susceptibility to IGC, a program was initiated to use this test method on various production heats in order to define the stabilization requirements for the T409 alloy. Samples from thirty-six different production heats of T409 sheet material were welded using an autogenous gas tungsten arc welding process. This group of heats included material which was only stabilized with Ti as well as material stabilized with additions of both Ti and Cb (dual stabilized). Care was taken to avoid any full penetration during welding. After welding the samples were pickled in a 10% HNO_3/2% HF solution at 60°C for 2 minutes and were then tested using method IV.

After testing the samples were mounted and polished to expose the cross section of the weld area and microscopic examination was used to determine if the specimen suffered IGC at the weld or HAZ. The cross sections were classified as "no attack" if no evidence of IGC could be found. Samples which suffered obvious intergranular attack were classified as "IG attack" and samples which showed shallow grain boundary grooving were designated "grooved GB". Typical examples of these three designation are shown in Figures 6-8. The results of this investigation are summarized in Table 5. It should be noted that all the heats which were tested in this investigation meet the current ASTM A240 stabilization requirements of Ti=(6xC).



Figure 6. Cross section of weld showing no intergranular attack.

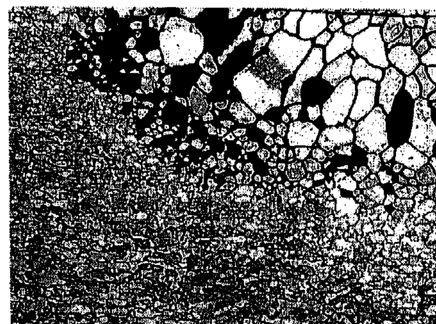


Figure 7. Cross section of weld showing severe intergranular attack of both weld metal and HAZ.

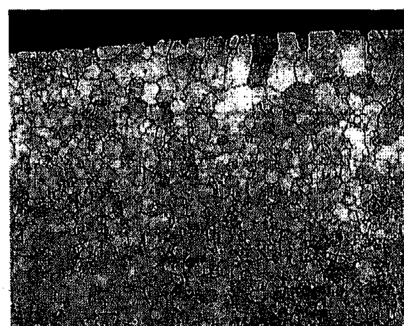


Figure 8. Cross section of weld showing grain boundary grooving of both weld metal and HAZ.

Table 5. Test Results for the Stabilization of Type 409 Alloys.

Sample ID	%C	%N	%C+%N	%Ti	%Cb	%Ti+%Cb	Remarks
			Ti Stabilized				
1	.006	.010	.016	.23	.014	.024	No Attack
2	.005	.010	.015	.17	.050	.220	No Attack
3	.007	.015	.022	.30	.003	.303	No Attack
4	.011	.013	.024	.31	---	.310	No Attack
5	.041	.005	.046	.50	.003	.503	No Attack
6	.049	.006	.055	.56	.003	.563	No Attack
7	.014	.019	.033	.12	.005	.125	IG Attack
8	.029	.013	.042	.40	.003	.403	IG Attack
9	.015	.017	.032	.18	.004	.184	IG Attack
10	.014	.013	.027	.12	.020	.140	IG Attack
11	.013	.018	.031	.13	.010	.140	IG Attack
12	.008	.021	.029	.11	.003	.113	IG Attack
13	.015	.012	.027	.17	.007	.177	IG Attack
14	.006	.013	.019	.12	.003	.123	IG Attack
15	.009	.014	.023	.15	.005	.155	IG Attack
16	.012	.015	.027	.20	.005	.205	IG Attack
17	.007	.009	.016	.17	.009	.179	IG Attack
18	.011	.014	.025	.14	.002	.142	IG Attack
19	.012	.015	.027	.16	.003	.163	IG Attack
20	.060	.005	.065	.47	.005	.475	IG Attack
21	.005	.008	.013	.16	.004	.164	Grooved GB
22	.004	.010	.014	.15	.006	.156	Grooved GB
23	.005	.010	.015	.18	.003	.183	Grooved GB
24	.005	.008	.013	.15	.008	.158	Grooved GB
25	.007	.012	.019	.17	.013	.183	Grooved GB
			Ti + Cb Stabilized				
26	.007	.017	.024	.16	.140	.300	No Attack
27	.009	.012	.021	.14	.140	.280	No Attack
28	.009	.010	.019	.16	.270	.430	No Attack
29	.008	.014	.022	.10	.240	.340	No Attack
30	.011	.012	.023	.11	.260	.370	No Attack
31	.012	.012	.024	.11	.270	.380	No Attack
32	.011	.012	.023	.12	.200	.320	No Attack
33	.010	.018	.028	.17	.360	.530	No Attack
34	.009	.012	.021	.14	.060	.200	IG Attack
35	.011	.014	.025	.09	.100	.190	IG Attack
36	.013	.013	.026	.12	.140	.260	IG Attack

RESULTS AND DISCUSSION

The results summarized in Table 4 and Figure 3 show that the corrosion rates of the test heats increase substantially following the 0.5 hour treatment at 600°C. Also, the corrosion rate after this aging period depends on the stabilization ratio, Ti/(C + N). For example, the heat with the 3.8 stabilization ratio completely disintegrated while the heats stabilized to 5.3 and 10.5 had corrosion rates of 0.0191 and 0.0088 IPM respectively. The high corrosion rates measured after short aging times are restored by longer aging periods. As shown in Figure 4 these high corrosion rates are the result of IGC. All these observations are consistent with the formation of Cr depleted regions during the short aging times which then heal by the back diffusion of Cr during the longer aging times. These results demonstrate that the IV test method does detect Cr depleted regions with commercial T409 alloys and can be used to detect susceptibility to IGC.

Figure 5 demonstrates that the IV test method does detect susceptibility to IGC in HAZ of actual muffler weldments. Note that the attack in Figure 5 is very similar in severity to the attack which occurred during service (see Figure 1). These observations suggest that the test results measured with method IV do parallel service performance.

The data which are summarized in Table 5 were plotted as (%Ti + %Cb) versus (%C + %N) to determine if a relationship could be found which would separate the samples which suffered IGC form those that pass the test. The plot of these data, which is presented in Figure 9, reveals that there is a well defined separation between the samples which passed the test and those which suffered IGC and/or grain boundary grooving. The dual stabilized heats (circled data points in Figure 9) also show a well defined separation between the heats which suffered IGC and those which pass the test. The boundary between the "pass" and "IG attack" heats is independent of whether the heats are dual stabilized with Ti + Cb (circled data points in Fig. 9) or just stabilized with Ti. Based on the straight line which separates the "passes" from the "IG attack", the stabilization requirement for T409 Ferritic stainless steel can be given as;

$$\%Ti = 0.08 + 8(\%C + \%N) \quad (1)$$

(for Ti stabilized heats)

$$(\%Ti + \%Cb) = 0.08 + 8(\%C + \%N) \quad (2)$$

(for dual stabilized heats)

Recent work by Sunomiya et al [6] gives close confirmation to this stabilization requirement for T409 alloy.

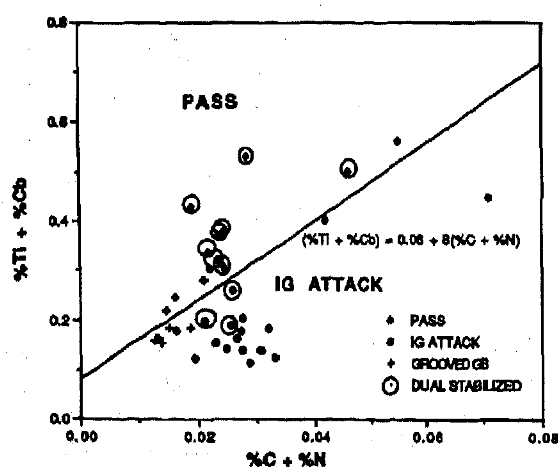


Figure 9. Welded Production Heats

Within the range of chemistries which were tested in this investigation, use of Ti + Cb provides an alternate means of achieving stabilization. Dual stabilization is currently employed for several ferritic stainless steels included in ASTM A240[1]. Dual stabilization provides the alloy producer with an alternative means of achieving a high degree of stabilization without having to increase the Ti content of the alloy. This can be a big advantage to the user because minimizing the Ti concentration of the alloy will reduce the level of Ti precipitates which can degrade surface quality and add to the processing costs. It has also been shown[7] that the dual stabilized T409 product offers improved formability over product which has been stabilized solely with Ti.

CONCLUSION

Based on this work the following conclusions can be drawn about IGC of the T409 alloy and test methods for evaluating susceptibility to IGC.

1. The existing stabilization requirement given in the ASTM A240 specification for T409 alloy does not consider the detrimental effects of N.

2. Muffler weldments fabricated from T409 sheet material which meets the existing ASTM specifications can be susceptible to IGC if the degree of stabilization is inadequate and a corrosive environment is present.

3. Although no standard test method currently exists for low chromium ferritic stainless steels, the boiling Cu/0.5% H_2SO_4/6% $CuSO_4$ test conducted for 20 hours does detect sensitized grain boundaries with T409 production heats.

4. Based on the results from the Cu/0.5% H_2SO_4/6% $CuSO_4$ test, the minimum stabilization requirements for Ti or (Ti + Cb) to prevent susceptibility to IGC in T409 alloys is:

$$\%Ti = 0.08 + 8(\%C + \%N) \quad (1)$$

(for Ti stabilized heats)

$$(\%Ti + \%Cb) = 0.08 + 8(\%C + \%N) \quad (2)$$

(for dual stabilized heats)

5. The use of dual stabilization provides an alternate and highly effective means for attaining a proper degree of stabilization without the detrimental effects of increased Ti levels.

REFERENCES

1. ASTM A240/A240M - 94a: Standard specification for Heat-Resisting chromium and Chromium-Nickel stainless Steel Plate, Sheet, and Strip for Pressure Vessels.

2. A. P. Bond, Transactions Of The Metallurgical Society Of AIME, V245 (1969): p. 2127.

3. R. A. Lula and J. A. Davis, Intergranular Corrosion in 12% Chromium Ferritic Stainless Steels, Intergranular Corrosion of Stainless Alloys, ASTM STP 656, R. F. Steigerwald, Ed., American Society for Testing and Materials, P. 233 1978.

4. T. M. Devine and B. J. Drummond, Corrosion, Vol. 38, No. 6, P. 327, 1982.

5. S. Frangini and A. Mignone, Corrosion, Vol. 48, No. 9, P 715, 1992.

6. T. Sunomiya, H. Furuki, and T. Adachi, Nisshin Seiko Giho, Vol. 71, No3, P53, 1995.

7. Y. Xu, C. I. Garcia, I. A. Franson, and A. J. DeArdo, International Symposium on Low-Carbon Steels Additional

971033

Ventilated Brake Rotor Air Flow Investigation

Michael D. Hudson and Roland L. Ruhl
University of Illinois at Urbana-Champaign

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Air flow through the passages of a Chrysler LH platform ventilated brake rotor is measured. Modifications to the production rotor's vent inlet geometry are prototyped and measured in addition to the production rotor. Vent passage air flow is compared to existing correlations. The inlet modifications show significantly improved vent air flow, over the production rotor. The result improvement in heat transfer and rotor cooling is reported. These benefits in performance should be attainable at very low increases in production cost.

1.0 INTRODUCTION

One of the basic functions of a brake system is to slow a vehicle. In doing so, vehicle kinetic energy is transferred to thermal energy via the friction between brake pad and brake disc (or drum). Brake discs absorb such a large amount of thermal energy in a short period of time (braking) that they effectively serve as energy storage devices. While this large heat input must be dissipated quickly to allow for additional braking when necessary, the effective time constants for each process are far from equal. Thus, brake discs are designed with two major factors in mind:

(1) Thermal Capacitance: Energy Storage
(2) Heat Transfer: Dissipation of heat primarily by forced convection.

In addition, modern cars are designed to minimize weight, including brake rotor weight. Although smaller rotors weight less and typically cost less to manufacture, their thermal capacity is reduced. Furthermore, vehicle down sizing and front wheel drive components can compromise available air flow to the brake discs. Thus, the cooling characteristics of brake systems are increasingly important, and should be maximized with minimal cost and weight penalties.

For years, ventilated brake rotors have been used for their weight savings and additional convective heat transfer from the air passages between the rotor rub cheeks (passages lacking in solid rotors). However, the amount of additional cooling due to this internal air flow is not well defined and depends on the particular brake rotor's geometry and the cooling air flow conditions around the brake assembly.

2.0 PROBLEM STATEMENT

With the aid of Chrysler Corporation, investigation of ventilated brake rotor vane air flow is undertaken herein. The goal of this investigation is to measure the current vane air flow for the given test rotor at various speeds (simulating 30, 45, 60 mph), and to improve this vane air flow to increase brake disc cooling.

Test Rotor

Brake rotor vane air flow is examined for the Chrysler LH-series front brake disc (see Figures 2.1 and 2.2). At the inlet, air flow into the vane is drawn only from the inboard side of the rotor. In addition, the air must flow by the knuckle and turn approximately 90 degrees into the entrance of the rotor intake. Since the bearing hub and knuckle are not axisymmetric, the inlet clearance between the inner rub cheek and the bearing to knuckle mounting varies around the circumference. This inlet restriction due to the mounting/suspension components is accepted as a given, unalterable geometric constraint.



Figure 2.1: Test Rotor - Basic Geometry

Also, the approximate 90 degree inlet turn is believed to cause loss and possibly flow separation at the inlet, thus reducing the effective inlet area and resultant vane flow.

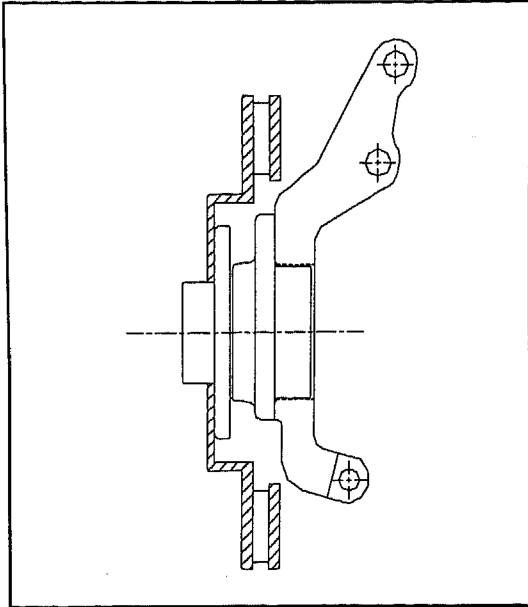


Fig 2.2: Test Rotor - Basic Mounting Geometry

Focus

Improving the flow of air through the brake rotor vanes centers on two primary factors:

(1) The centrifugal pumping aspects of vane orientation (i.e., straight, radial vs. curved, involute or angled vanes)
(2) Flow conditions at the inlet, including inlet geometry

First, the straight radial vanes of the current LH front brake rotors do not yield the best pump from the vane velocity/cooling standpoint [3,4]. However, involute or angled vanes are more difficult to manufacture and the loss of symmetry for such vanes results in left and right hand versions of the brake rotor. The associated production cost increase has effectively ruled out any non-symmetrical rotor designs. For these reasons, little consideration is given to vane orientation.

Focus is placed on the inlet flow conditions and inlet design. All modifications are chosen to result in symmetric brake rotor geometry. To avoid production cost increases, modifications should allow rotors to be horizontally sand cast, just as current production rotors.

3.0 BACKGROUND

Brake rotor vane flow correlations to estimate heat transfer coefficient have long been used in brake system thermal modeling. Unfortunately, these models can be only as accurate as estimates of heat transfer coefficients and boundary conditions allow. In general, these conditions require empirical adjustment to make the model fit actual brake temperature data. From a theoretical standpoint, estimating the air flow velocity (to predict heat transfer coefficient) over the "exposed" brake surfaces is almost impossible.

The literature to date suggests that one directly measure rotor air flow (as best possible) or turn to an empirical relation based on prior published investigations. Currently, there are two primary empirical relations for ventilated brake disc air flow. Both have been accepted as probative over the years, in one form or another. Both are (loosely) based on Reynolds Analogy.

The first relation for average vent velocity is from Sisson [2] and quoted later in Sheridan et al.[1].

$$V_{ave} = \xi\,\omega\,\frac{D_o}{2}\,, \quad \xi = \sqrt{-0.020 + 0.908D_i - 0.202D_i^2}$$

where ξ = correlation factor; ω = rotational rate; D's = the diameters at inlet & outlet. Inlet diameter must be in meters. Sisson's correlation is valid over the range of rotors tested, from 23 to 41 cm outer diameter, 10 to 18 cm inner diameter, and 0.6 to 2.5 cm vane widths. The number of vanes varied from 20 to 40. All models had straight radial vanes.

In addition, Sisson measured the vent heat transfer coefficient, by insulating all but the interior passages of the rotor. [Heat transfer coefficient, h_{air} is the ratio of heat transfer at the surface to the temperature difference; Nusselt number, Nu is a nondimensionalized form of h_{air};]. Sisson's heat transfer relation is

$$Nu = 0.045Re^{0.8}(D_o/2)^{0.2}\left[1 + 6.6(D_o/2)^{0.8}\right]$$

where Nusselt number, Nu is based on a characteristic length equal to rotor outer radius ($D_o/2$) and Reynolds number, Re is based on that same radius and average vent velocity from the first equation (above). However, some of Sisson's graphical data might suggest somewhat higher values for vent heat transfer than predicted by the above equations.

The other prominent vent air flow correlation comes from Limpert [3].

$$V_{ave} = \frac{V_{in}}{2}\left(1 + \frac{A_{in}}{A_{out}}\right), \quad V_{in} = 0.001655\sqrt{D_{out}^2 - D_{in}^2}$$

where V is mean air velocity; D is diameter in meters; A is vent area, all at inlet and outlet.

Limpert's average velocity values are roughly one-third lower than Sisson's. Limpert also splits two cases to estimate vane heat transfer coefficient, one for turbulent flow and one for laminar flow. For laminar flow ($Re < 10000$),

$$Nu = 1.65Re^{0.33}(d_h/l)^{0.33}$$

For turbulent vent flow ($Re > 10000$),

$$Nu = 0.0204\left[1 + (d_h/l)^{0.67}\right]Re^{0.8}$$

where d_h is hydraulic diameter of a vent passage; l is the length of a passage. Nusselt number is based on hydraulic diameter. Reynolds number is based on hydraulic diameter and Limpert's average vent velocity correlation. (Prandtl number appears in the original equation for laminar flow but has been given the standard 0.7 constant for air and combined in the proportionality constant).

Limpert's turbulent vane flow heat transfer correlation mirrors Sisson's in basic form but represents a more rigorously non-dimensionalized Reynolds' Analogy relation as opposed to a more empirical correlation. Either relation is viable, especially since vent inlet flow conditions are so complex and the influence of these complex geometric factors cannot be correlated in any simple way [10].

Both Sisson's and Limpert's relations for vent heat transfer coefficient have been widely used and agree within five percent at 600 rpm. (Limpert's coefficient is less sensitive to rotational rate than Sisson's). Though neither air flow estimate accounts for splash shields, suspension mounts, or other possible inlet restrictions, both provide an excellent reference for relating vent flow to vent heat transfer.

4.0 TURBOMACHINE ANALYSIS

Flow in the Rotor Vanes

The general theory and solution for the average air flow in the rotor vanes centers on these standard assumptions for turbomachine analysis:

(1) Air is an ideal, incompressible, inviscid fluid where gravitational potential energy is negligible.
(2) All shaft work is transferred to the fluid, i.e. no loss.
(3) Mean air flow is 2-D, axisymmetric flow for an ideal impeller with an "infinite number of infinitely thin blades", i.e. along a "mean flow surface" (streamline).

In this case, the mean flow surface is a path outlined by the rotating, straight radial rotor vane, i.e. the relative motion of the air is in the same direction as the rotor vane surfaces.

Based on the First Law of Thermodynamics, an ideal rotor's air pumping equation is represented as follows [see also ref. 14,15,17].

$$g\,(h_2^o - h_1^o) = \frac{c_2^2 - c_1^2}{2} + \frac{u_2^2 - u_1^2}{2} + \frac{w_1^2 - w_2^2}{2}$$

where c is the absolute air velocity; component velocities w and u are relative vane velocity and peripheral (tip) velocity, respectively; h^0 is enthalpy and g is the gravitational constant.

The first term on the RHS of the equation is the kinetic energy change due to the work input. The combination of terms containing peripheral and relative velocities is the static pressure change from the vane inlet to outlet. (Recall Bernoulli's Equation along a streamline). The relative velocity term represents the diffusion or retardation of the fluid relative to the impeller. The peripheral velocity term represents the change in static pressure due to the centrifugal force acting on the fluid.

It can be seen from the above equation that increasing the inlet pre-rotation (u_1) and reducing the relative velocity at inlet (w_1) will decrease the "effective" opposing pressure gradient. Note the trade off: minimum inlet relative velocity makes for a more efficient pump, but higher relative velocity over the length of the passage between the rotor rub cheeks makes for better cooling.

Since pumping work is negligible compared to engine output, any work penalties for higher relative velocities are not of practical concern. Higher vane flow velocities for cooling outweigh rotor pumping efficiency. In addition, the next section will show that accelerating the inlet stream may also help reduce flow separation at the inlet. Flow separation at the inlet can dramatically effect impeller performance by greatly reducing the effective inlet area.

Inlet Flow

Inlet flow geometry is highly case-specific. Complex inlet conditions (Figure 2.2) make modeling extremely difficult. Even the most extensive published analyses usually reduce inlet design to a few rules of thumb. Some general guidelines for turbomachine inlet design follow:

(1) *Smooth the transition* at the inlet to reduce loss or separation on the lip.

(2) *Accelerate the main stream* to combat separation (favorable pressure gradient tends to prevent excessive thickening of the boundary layer on the surface).

(3) *Impart fluid pre-rotation* and direct flow into the rotor vanes with minimum loss via the addition of guide vanes, inducer, or pre-impeller.

Loss associated with the flow around the inlet bend is small relative to loss from flow separation at the corner of this bend, which can result in partial or full blockage of the inlet.

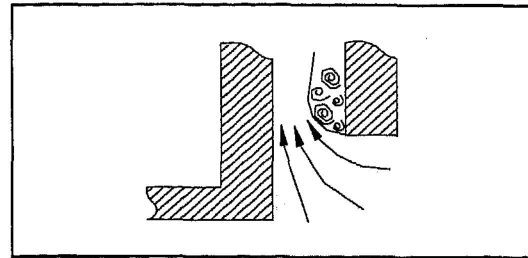


Figure 4.1: Flow Separation at Inlet

This accumulation of low-energy air can distort the effective air passage shape, thus reducing the total flow through the rotor (Figure 4.1). Thus, the inlet design should turn the air flow from the axial to the radial direction with minimum loss or separation [16].

5.0 DESIGNS

From the above turbomachine analysis, the following design modifications were selected to combat inlet separation by smoothing the intake, accelerating and directing the inlet air flow to the rotor vanes.

(1) Inlet lip chamfer/radius - Figure 5.2.

(2) Rounded inlet horn - Figure 5.3.

(3) Pre-impeller by means of extending the rotor ribs down into the inlet for greater active blade length - Figure 5.4.

All designs chosen to maintain rotor symmetry and manufacturability from a casting standpoint.

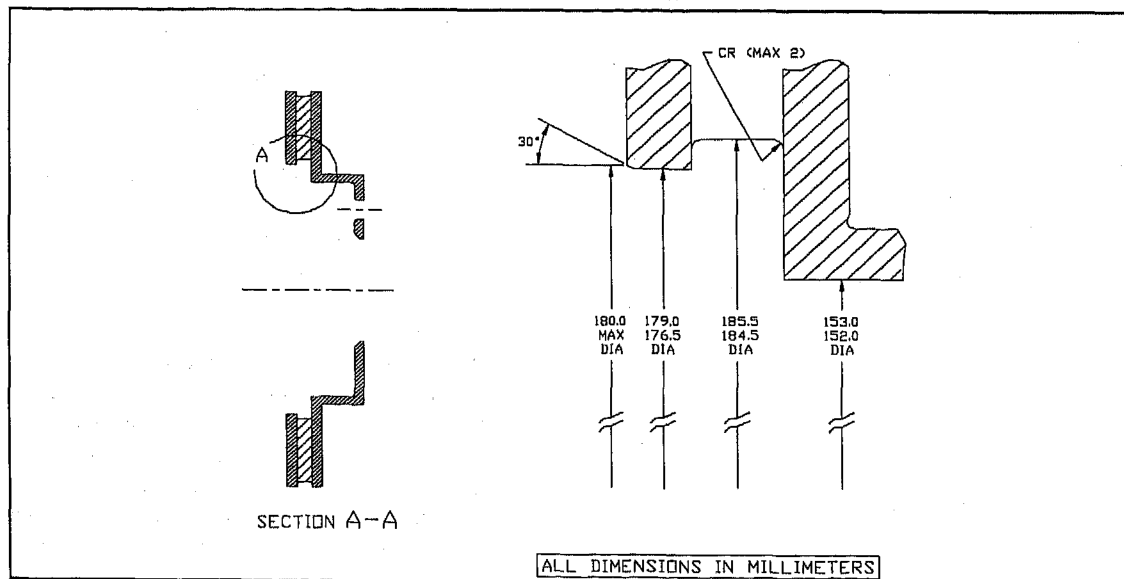


Figure 5.1: Original Rotor - Basic Inlet Geometry

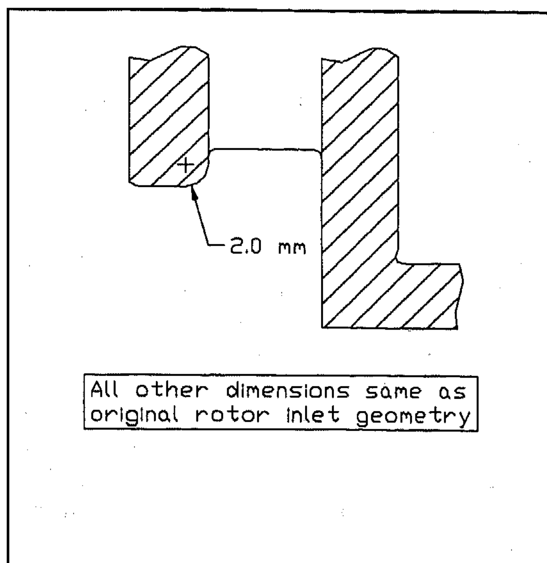


Figure 5.2: Smooth Inlet Lip -View in Circle A

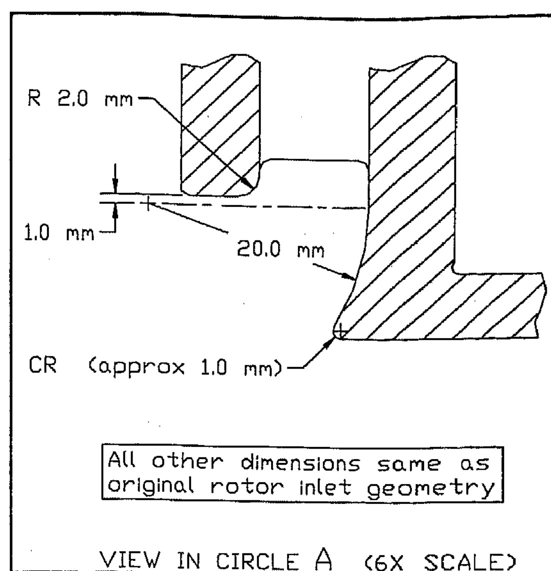


Figure 5.3: Rotor Inlet Horn -View in Circle A

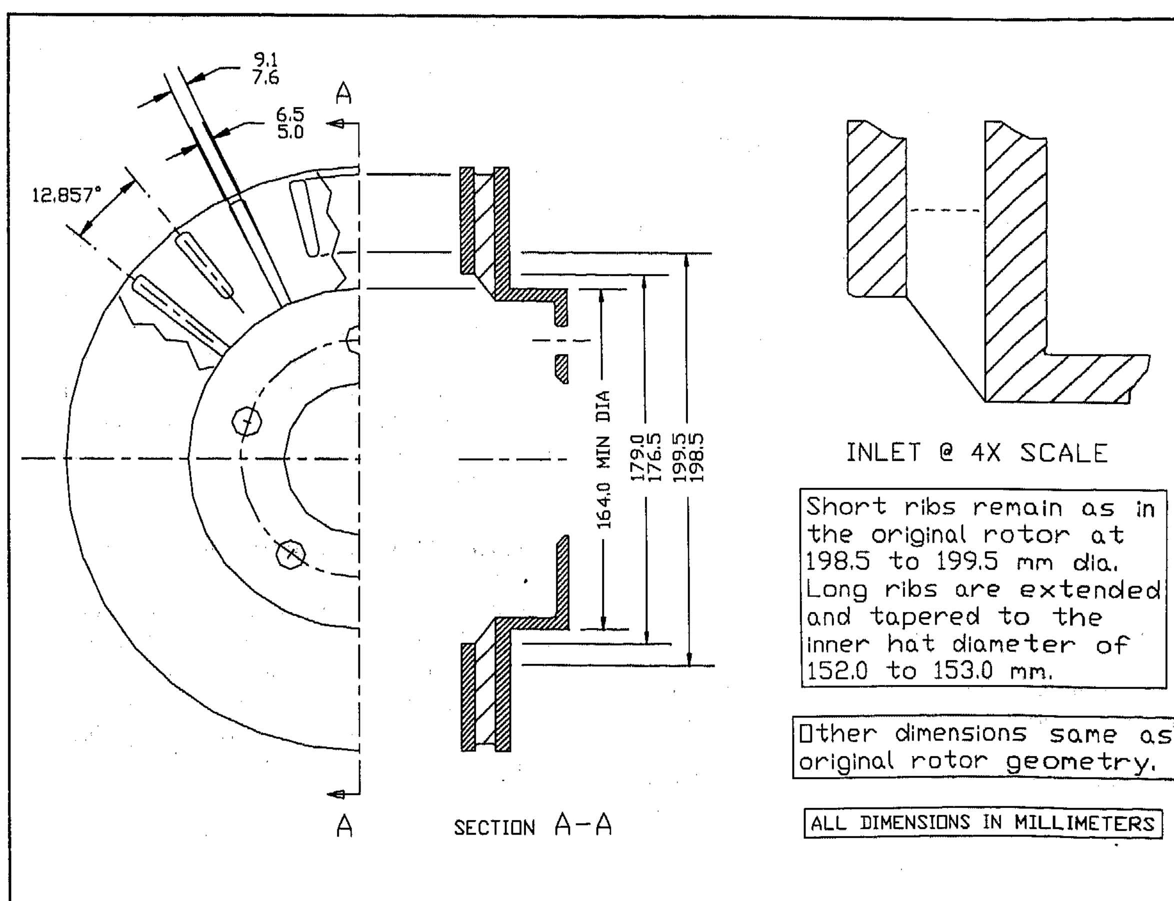


Figure 5.4: Rotor with Pre-Impeller Ribs

6.0 EXPERIMENTAL STUDIES

At rotational rates equivalent to 45 and 60 m.p.h., the production rotor outlet vane flow was measured for the current production rotor and all inlet modifications herein. Each rotor's outlet flow profile was measured at the various speeds with and without closed ventilation passages.

Direct measurement of interior vane flow was not accomplished due to the small vane cross-section and location of the associated measurement components in the rotating brake disc. Flow measurement at vane inlet is extremely difficult due to problems adequately defining the flow direction in the complex geometry (three-dimensional flow, high shear). Ventilated brake rotor vane air flow was measured at the outlet, where flow is two-dimensional (w/o wheel and tire blockage) - see Figure 6.1. Magnitude and direction of the outlet flow vector are measured using a three-holed wedge-shaped pitot-static pressure probe. Probe head design was adapted from a Bryer and Pankhurst [21] small wedge design, originally designed specifically for turbomachine research [21,22,23].

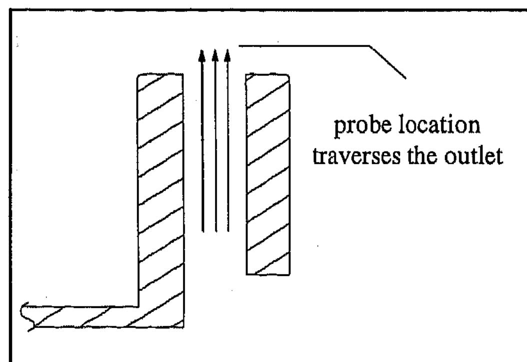


Figure 6.1: Two Dimensional Outlet Flow

Rigid steel framing served to mount the knuckle and modified bearing hub. The rotor was located on the wheel studs as on the actual automobile. And the bearing hub assembly was fit with a drive pulley and rotated by frequency controlled electric motor.

7.0 RESULTS

Flow separation at the inlet could not be conclusively shown (even when the rotor's outer diameter was machined down). The relative outlet vane velocity profiles do not show a backflow or zero flow region. Instead, the vane flow at the outlet appears to be very roughly parabolic. Figures 7.1 and 7.2 show typical measured vane flow velocity profiles. Table 7.1 summarizes the vent flow velocity measurement results.

Comparison with Vent Flow Correlations
Since the flow measurements were made at distance of one centimeter from the rotor vane outlet, the amount of fluid mixing that occurred (slowing the fluid) is not precisely known. Mixing of the outlet jet with stationary ambient air slows the fluid as it moves away from the rotor outlet. Thus, the reported average vent flow velocities tend to be slightly underestimated. (The mixing was minimized inasmuch as the outlet measurements were taken as close to the rotor as possible without the probe interfering with the outlet flow). It is clear that the relative velocity (in the rotor vane) is greater than that predicted by Limpert, and vent velocity may in fact be as large as that predicted by Sisson, since the measurements err on the small side. Clearly, the measured production rotor vane flow most closely matches the correlation of Sisson [2].

Predicted Heat Transfer for the Designs
Table 7.2 summarizes the heat transfer for the various designs as predicted by Sisson's and Limpert's correlations [2,3]. Using Limpert's research, the convective heat transfer coefficient associated with rotor vent flow improves by as much as 20%; upwards of 50%, using Sisson's model.

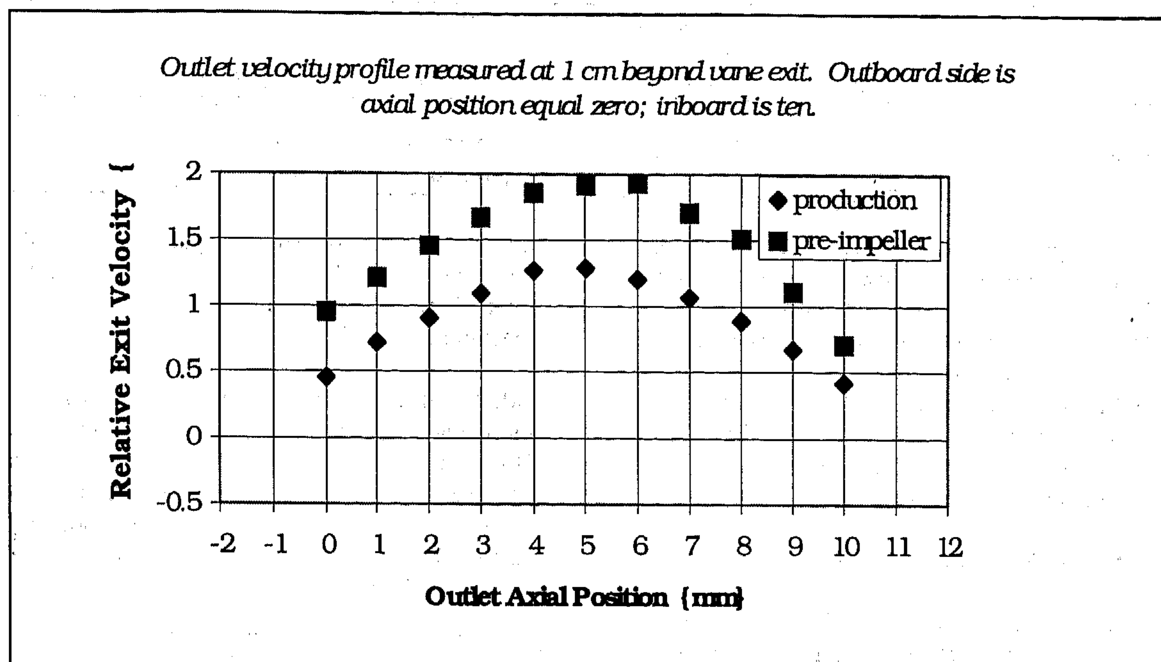


Figure 7.1: Pre-Impeller Rib Design's Outlet Velocity Profile - 600 rpm

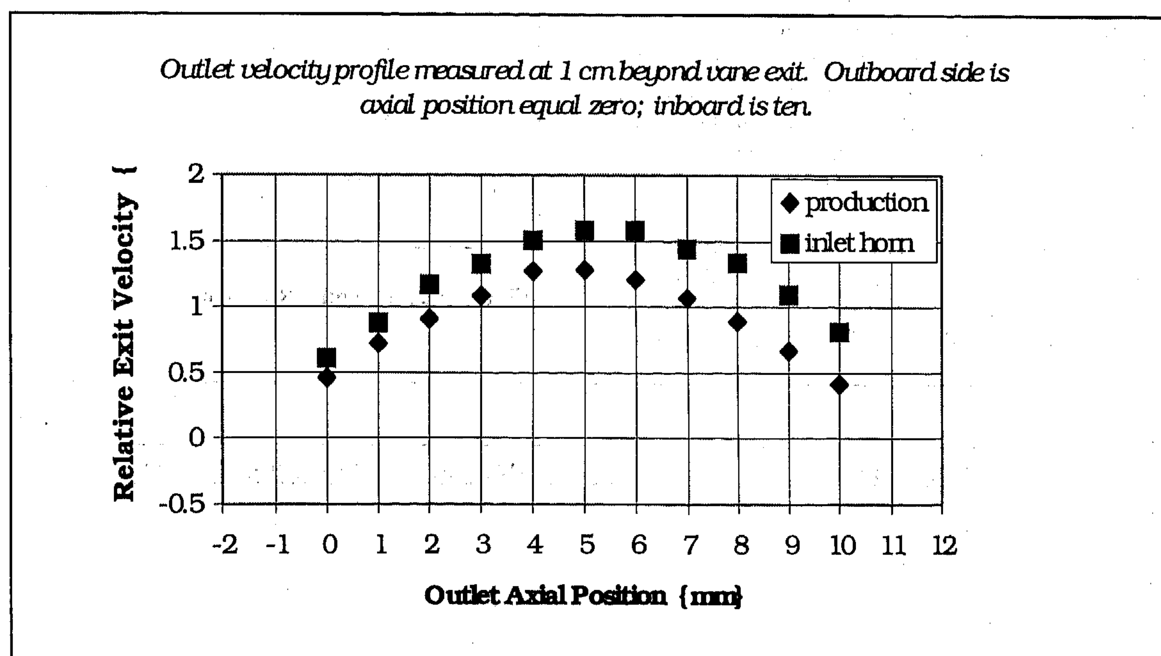


Figure 7.2: Inlet Horn Design's Outlet Velocity Profile - 600 rpm

Table 7.1: Average Vent Velocity as Measured at 1 cm Beyond the Vane Outlet

	Average Vent Flow Velocity {m/s}					
RPM	Limpert	Sisson	Prod.	Cham.	Horn	PreImp.
600	0.80	1.20	0.957	1.058	1.264	1.523
800	1.07	1.60	1.159	1.281	1.688	1.967

Table 7.2: Heat Transfer Coefficient Increase Based on Relative Velocity Increase

	Heat Transfer Coefficient {W/m^2K}			
	Prod.	Chamfer	Horn	PreImp.
RPM	600 / 800	600 / 800	600 / 800	600 / 800
Limpert	24.0 / 26.5	24.8 / 27.4	26.3 / 30.0	28.0 / 31.6+
Sisson	22.9 / 29.0	24.8 / 31.8	28.6 / 39.1	33.2 / 44.2

The improved heat transfer coefficient from increased cooling vent flow can be demonstrated by a simple fade test and lumped rotor model [1,7]. At the end of a ten stop (250 second) fade test the improved rotor's temperature above stabilization temperature is 3-6 % lower than the production rotor, depending on the chosen correlation. (Stabilization temperature is the rotor temperature at time infinity).

8.0 RECOMMENDATIONS

Measurements of current Chrysler LH series front brake rotors showed ventilation passage air flow corroborating Sisson's correlation [2]. Design modifications to improve the vane flow centered on the flow inlet conditions. The pre-impeller rib design yielded the most significant flow velocity improvement (greater than fifty percent) with negligible manufacturing cost increase, requiring only a slight change in casting shape and no machining operations.

If the pre-impeller concept is to be implemented, it is recommended that the modification of Figure 8.1 be examined in greater detail (i.e. optimized) to further reduce the inlet flow loss. A smooth inlet lip similar to that in Figure 5.1 with the combination of the pre-impeller and rounded inlet horn is desired within casting and machining limitations.

Additional possibilities include symmetric yet more egg-shaped rotor ribs similar to Figure 8.2. However, the curved flow path of the pressure rib-face is compromised by the restriction on the suction side (to maintain symmetry) and potential casting difficulty. Betz [15] and Lohmann [16] discuss the issues related to Figure 8.2.

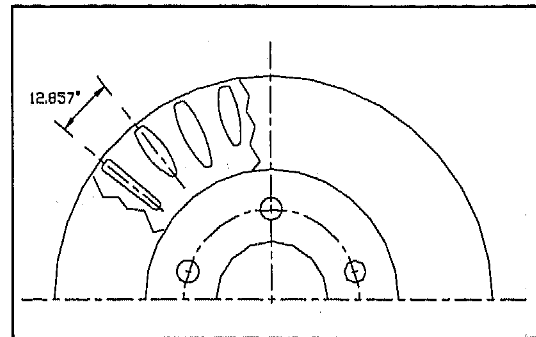


Figure 8.2: Potential Rib Shape Modification

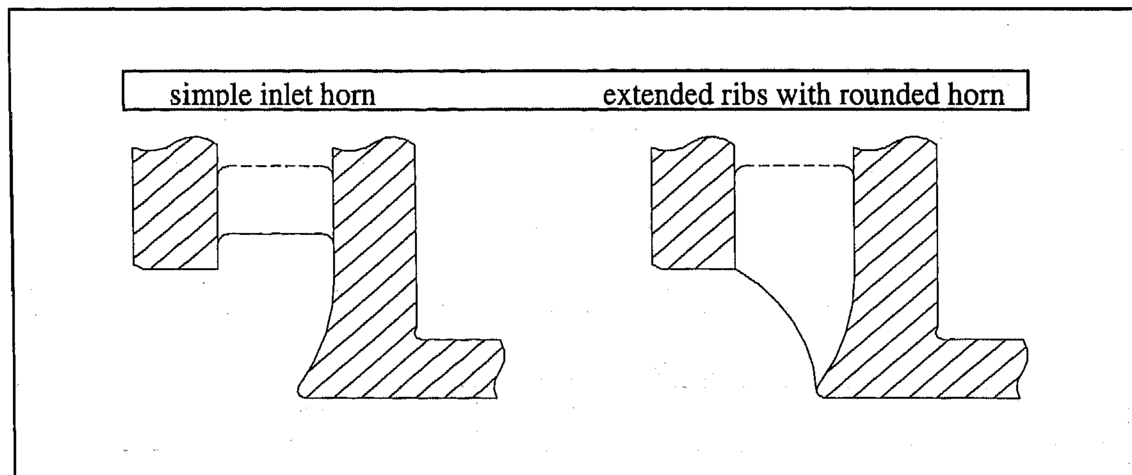


Figure 8.1: Pre-Impeller Rib Design Inlet Modification

9.0 REFERENCES

[1] Sheridan, D.; Kutchey, J.; Samie, F. "Approaches to the Thermal Modeling of Disc Brakes". SAE Paper 880256.

[2] Sisson, A. E. "Thermal Analysis of Vented Brake Rotors". SAE Paper 780352.

[3] Limpert, R. "Cooling Analysis of Disc Brake Rotors". SAE Paper 751014.

[4] Limpert, R. "The Thermal Performance of Automotive Disc Brakes". SAE Paper 750873.

[5] Hartter, L.; Schwartz, H.; Rhee, S. "Evaluating Copper Alloy Brake Discs by Thermal Modeling". SAE Paper 740560.

[6] Dennis, R. and Morgan, S. "A Theoretical Prediction of Disc Brake Temperatures and Comparison with Exptl. Data". SAE Paper 720090.

[7] Highley, F. H. "Techniques for Determining the Thermal Characteristics of Brake Drums and Discs". SAE Paper 710589.

[8] Dennis, R.; Newstead, C.; Eden, A. "The Heat Transfer from a Rotating Disc in an Air Crossflow". Proc. of the 4th International Heat Transfer Conference. Paris, France, F.C. 7.1. September 1970.

[9] Newcomb, T. and Miller, N. "Cooling Rates of Brake Drums and Discs". Proc. of Institution of Mech. Engineers 1965-6, Vol. 180, Part 2A, No. 6.

[10] Richardson P. and Sanders, O. "Studies of Flow and Heat Transfer Associated with a Rotating Disc". Journal of Mechanical Engineering Science, Vol. 5, No. 4, 1963.

[11] Kreith, F.; Maroti, L.; Deak, G. "Flow Phenomena of Partially Enclosed Rotating Discs". ASME Journal of Basic Engineering, September 1960.

[12] Kreith, F.; Taylor, J.; Chong, J. "Heat and Mass Transfer from a Rotating Disc". ASME Journal of Heat Transfer, May 1959.

[13] Kays W. and Bjorklund, I. "Heat Transfer from a Rotating Cylinder With and Without Crossflow". Transactions of the ASME, Jan 1958.

[14] Eck, B. Fans. Pergamon Press. 1973. {Translated from "Ventilatoren" 5th Ed. 1972}

[15] Betz, A. Intro. to the Theory of Flow Machines. Pergamon Press. 1966. {Translated from "Einfuhrung in die Theorie der Stromungsmaschinen"}

[16] Lohmann, R. "An Investigation of the Influence of Boundary Layers on the Performance of Centrifugal Compressor Impellers". ASME Journal of Basic Engineering, Paper No. 65-WA/FE-7.

[17] Hugelman, R. Emeritus Professor Department of General Engineering, University of Illinois, Personal Interviews, 1995.

[18] Chang, P. Control of Flow Separation. McGraw-Hill, Inc. 1976.

[19] Stepanoff, A. Turboblowers. Wiley. 1955.

[20] Goldstein, R. Fluid Mechanics Measurements. Hemisphere Publishing. 1983.

[21] Bryer, D. and Pankhurst, R. Pressure Probe Methods for Determining Wind Speed and Flow Direction. Her Majesty's Stationary Office. 1971.

[22] Lakshminarayana, B., et al. "Conventional Probes for the Relative Flow Measurement in Turbomachinery Rotor Blade Passage". Transactions of the ASME, Volume 103, April 1981.

[23] Chue, S. "Pressure Probes for Fluid Measurement". Progress in Aerospace Sciences, Volume 16, 1975.

971035

Standard Test Procedure for Passenger Car Brake Components Using Operating Strength

H. J. Tumbrink
Lucas Varity, Light Vehicle Braking Systems

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Vehicle breakdown statistics show a clear trend: quality will be a major target for the future. This, taken together with the growing market demand for greater reliability, shows that the automobile industry needs to take action.
A way to achieve this is being sought, in some cases too frantically. All the concepts which have been formulated, as yet, have a common aim:

- **Price or cost reduction**
- **A more flexible response to market trends**
- **To boost the consumer's quality consciousness.**

The right response to these challenges leads directly to a reduction in development time.
The first answer is a standardization of test procedures according to operating strength.
The first idea and outlook was done in the SAE-Paper 790464 in 1979: "The ASSET-Technique". Now in 1996 it has become a standard in the German car industry.[1]

INTRODUCTION

Similar to other branches of industry, the modern brake manufacturer has had to research into more efficient manufacturing methods and the optimization of material content in order to satisfy the even higher standards demanded of its product.
To achieve this objective, the disciplines of **Finite-Element-Analysis** (FEA), **Photo Stress Analysis** and **Operating Strength** have been employed to optimize the damaging stress in passenger car brakes.
In particular and specifically for brake designs, the following basics have been introduced.

- statistically calculated failure modes for components to ensure product design (product liability),
- light weight design to save energy and cost,
- suitable and cost effective selection of materials and production methods,
- improved component form in the highly stressed zones,
- minimized installation space, particularly resulting from the introduction of negative steering offset,
- correlation of stress tests in the function stress loading.

USAGE IN BRAKE DESIGN

But what were the considerations which led us to research into operating strength ?
In our first few years of disc brake production we knew precisely why a component failed after a given number of brake operations. Figure 1 shows a typical brake of the Colette type (floating caliper).
We had however no idea why other components of the same design withstood similar loads. The development of Woehler curves for components gave us the first chance to evaluate the factors responsible.

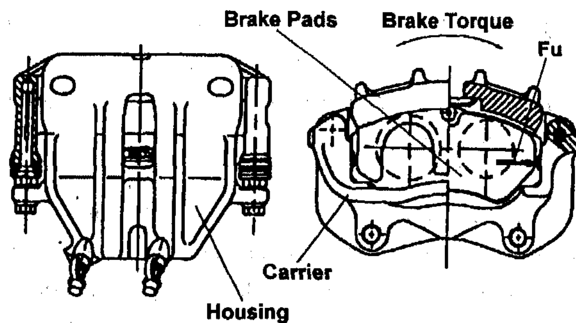


Figure 1: Fist Type Caliper Colette

Lightweight design was, is and will remain the "Leitmotiv" of the German Automotive Industry. We must however be in a position to create "tools" which will enable us to predict fatigue life fairly realistically and/or which we can modify using known correlation parameters.

Weight reduction with an ensured endurance life can only be achieved by using the method available through Operating Strength. [2]

Test specifications must be worked out with our customers so that there are not dozens of test procedures, differing for each customer, to be considered, for the self-same component. The tests should, in general, be as simple as possible and as rigorous as needed. [3]

In our case the required component strength must be established and the endurance life calculated. It is no longer possible to design on the drawing board optimal light weight components with an acceptable endurance life.

In short, it is now possible to load components not only to the limits of fatigue stress but in some cases up to and even over yield point always providing that the extreme loading does not occur too frequently in the normal working life of the component.

That brings us to the real problem in the fatigue life prediction theory:

"What is, in practice, too frequent ?" [4]

The solution is: Finite Element Analysis (FEA) and Operating Strength (FATIGUE) that means: "FATIGUE DESIGN" !

Since 1938 (and Gaßner's fundamental work) the logical opinion has been held that:

- There is no such thing as a service stress at constant amplitude and

- there are no un-notched components

This initiated the experimental determination of endurance life of components under realistic loading. [5]

We are now able to determine endurance life of components in advance (design stage) by using **Finite Element Analysis** and **Operating Strength,** and also with real components.

To answer the question the Woehler (S-N) - curve for the component must be considered in conjunction with the load histogram. The load histogram tells us the frequency at which a particular load occurs during the components normal life. This does not satisfy the customer, since he does not know, how frequently he may prudently allow overloads. In the case of fluctuating loads (brakes) the load cannot be defined numerically but rather by a characteristic function, i.e. the load collective.

Light-weight components cannot be laid out for to last forever but only for a defined life. Operating strength assessment and light-weight are therefore two sides of the same coin. It must be ensured that these collectives are realistic and apply to all components. The basic equation is an exponential function, derived from a discrete frequency function, illustrated in Figure 2. The scaled loading factor $\xi = 1.0$ represents 1.0 g deceleration.

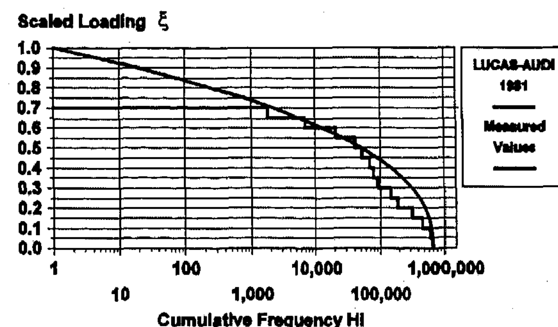


Figure 2: Cumulative Service Brake Load

A brake collective was collated between 1978 and 1981 in cooperation with AUDI (Dr. Anton Wimmer)[6] [2].

The collective was extrapolated to 300 000 km (186411 Miles) and therefore conserves a discrete frequency distribution which is also plotted in Figure 2.

The Fraunhofer Institute for Operating Strength (LBF) of Darmstadt has issued a technical report TM-81/77 which deals with the formal description of frequency distributions, by means of exponential functions. [7]

The collective can be described mathematically by using this well-known exponential procedure. The mathematically calculated collective is described in Figure 2.
The collective conforms to the equation:

$$H_i = 701269 * e^{(-13{,}4606 * \xi_i^{2{,}37338})} \tag{1}$$

$$\xi_i = \left(\frac{F_i}{F_1}\right) = \left(\frac{\sigma_i}{\sigma_1}\right) = \left(\frac{p_i}{p_1}\right) \tag{2}$$

The basic equation of this collective has the form:
H_0 = Zero passage for this function

$$H_i = H_0 * e^{[-\ln(H_0) * \xi_i^{n}]} \tag{3}$$

Safety related vehicle parts must last for at least **300000 km** at a failure probability of $\mathbf{10^{-5}}$.
The topic light-weight components and their loading has been a side issue till now.
Fundamentally:
Only when the service loading is accurately known or can be accurately estimated is it possible to dimension a component to suit the application.
Dubious loading assumptions lead to excessive safety margins and therefore impede the use of light parts [8].

TEST METHODS AND RELIABILITY

Teamwork between all specialist departments and all technologies is demanded for the development of a product. The part played by theoretical and experimental reliability assurance is very significant. Optimization takes place basically in the first design stages, so that the first prototype is 80-90% ready for production. The safe answer to the designer's frequent question to the stress engineer "Will it hold", can only be given by making sure that the component achieves the demanded endurance life under the specified **"Service loading collective"** conditions.

STANDARDIZED TEST METHODS - Taking a disc brake as an example, we show how generally applicable guidelines can be formulated, in cooperation with the German Automotive Industry and with the help of fatigue life analysis and modern statistical techniques. This has been worked out in a working circle of the Automotive Industry in concert with the two biggest brake suppliers. The aim is to carry out service load tests, based on a new collective, in the first stage of development. Until then a solution [9] has been worked out, which is based on a fixed collective and the Miner linear cumulative damage calculation.
First proposal of the "Brake Working Circle":
In the second meeting of the Brake Working Circle, an interim proposal of the German Auto-industry was agreed, as "lowest common denominator":
$N_{50\%}$ = 200 000 loading cycles and S_{log} = 0.2.
This implies testing against a mean life.
It is however clear from the literature [10] [11] that tests should not be based on the mean life. The reliability should always be determined for a minimum cycle count which represents a given endurance life or mileage.
Statistically it is reasonable to fix a minimum count (e.g. $P_{ü}$ = 99.9%, 99% or 90%). This led to a common proposal from the Brake suppliers [9] with an assumed collective and the test load cycle count N_{1k}.
The working circle initiated a comparison of all test stands at brake and car manufacturers with respect to damage and scatter.
In Figure 3 and Table 1 the test results are depicted.

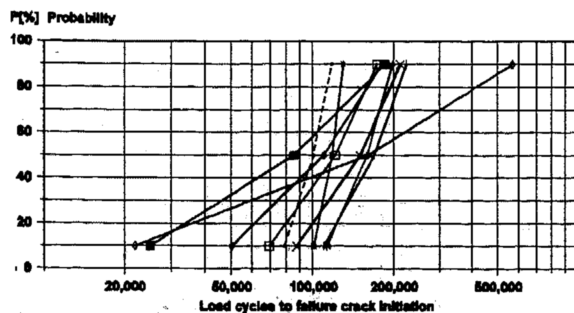


Figure 3: Test Rig Comparison Car Manufacturer and Brake Supplier

Lucas built a new test stand which closely simulates the vehicle setup.
The results showed very little test stand moment scatter (previously ± 5% and now ± 2%)
- Original stub axle and bearing
- Moment controlled directly at bearing.

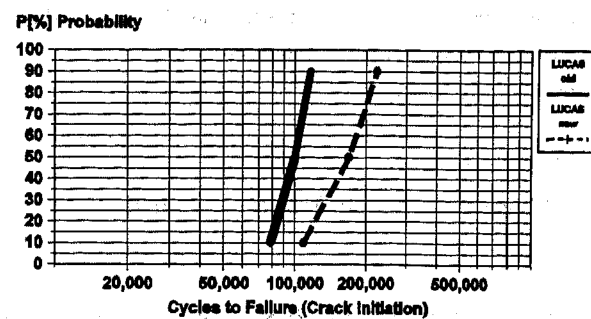


Figure 4: Test Rig Comparison Lucas old to Lucas new design

Figure 4 shows the difference between the results with the LUCAS rigid mounting (LUCAS old) and those from the new configuration using the original stub axle (LUCAS new). With the original stub axle "X" times as many cycles can be withstood. The mounting therefore plays an important part in running realistic tests. The results relating to 90% and 10% scatter, the log. Standard deviation and the b-value of the Weibull-Distribution are tabulated below.

Table 1: Test results of comparison tests

Tests:	T_N (90%/10%)	S_{log}	Weibull Slope b
Car Manufacturer (CM) 1 (all tests)	**7.45**	**0.340**	**1.54**
Car Manufacturer (CM) 1 (without min.)	**3.61**	**0.218**	**2.40**
Car Manufacturer (CM) 2	**1.75**	**0.094**	**5.57**
Car Manufacturer (CM) 3	**2.50**	**0.155**	**3.37**
Car Manufacturer (CM) 4	**25.98**	**0.552**	**0.95**
Brake Supplier 2	**1.29**	**0.043**	**12.30**
Lucas (old test stand)	**1.49**	**0.067**	**7.88**
Lucas (new test stand)	**2.04**	**0.121**	**4.33**
All (without 1 and 4 min. results)	**2.43**	**0.151**	**3.48**

It can also be seen from these results that the total evaluation exhibits a Gauss normal distribution. (b = 3.48; 3.2 - 3.6 = Gauss - normal distribution). [11]
The logarithmic standard deviation is 0.151, the max value 0.155 was chosen.
Recent Engineering Release Tests also confirm that the detected scatter is <2.5.

STATISTICAL BENCHMARKS - The results from table 1 lead to the following findings as regards statistical data. Statistical benchmarks must be determined from these tests and comparisons made over the past few years. The limiting quantities for minor single tests and the safety factor related to endurance life are to be chosen from the statistics. The risk from the fortuitousness of small sized random samples must taken care of, the confidence interval and P_a must be established.
It follows that:

$$J_{C,n}=S^{[\frac{U_C}{\sqrt{n}}]} \qquad (4)$$

S= 1.43 from S_{log} = 0.155, equivalent to a scatter T_s= 2.5, n=4, U_C = 1.282 for C = 90% (from TM68/78 (LBF) [12]) (1.282/2 = 0.641)
$J_{C,n}$ = $1.43^{0.641}$ = 1.26

and:

$$J_L=S^U \qquad (5)$$

S = 1.43 as equation (4) above
U = 3.10 for 99.9% survival probability. (P_a=99.9%)
J_L = $1.43^{3.10}$ = 3.03

This gives us a statistically based safety factor **S_F = 3.82**
(S_F= $J_{C,n}$ * J_L)

The factors are illustrated in Figure 5.

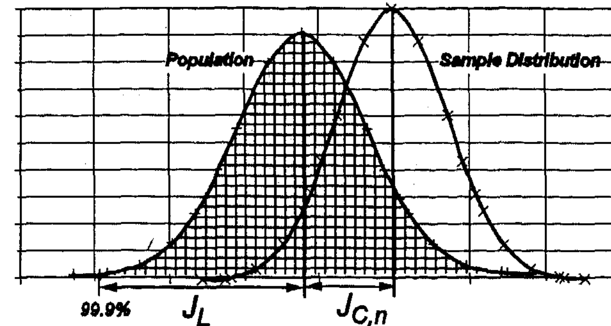


Figure 5: Statistical Factors $J_{C,n}$ and J_L

COLLECTIVE - The new brake load collective takes into account **20** non-negligible cases, from previous customer reports, where the tangential force exceeded the equivalent of more than 1.3 g (braking on extremely rough tracks, road edges etc.).
2130 stops at 1g are included in the scope of the collective. These could conceivably be harder stops, masked by ABS.
The remaining collective sections follow the so-called normal distribution, which applies for 1 to 1 200 000 cycles. The whole collective covers 4 braking operations per km = H_{ges} = 1.2 X 10^6 cycles, as recommended by LBF.
This collective can also be applied to light trucks as an initial approach.

The equation reads:

$$H_i=(H_0-2149)*e^{(-lnH_0+\xi_i^2)}+2149 \qquad (6)$$

Plus 20 cycles at 1.3 times 1.0 level.
H_0 = 1.2 * 10^6. In Figure 6 the new brake load collective is illustrated and is also valid for ABS applications.

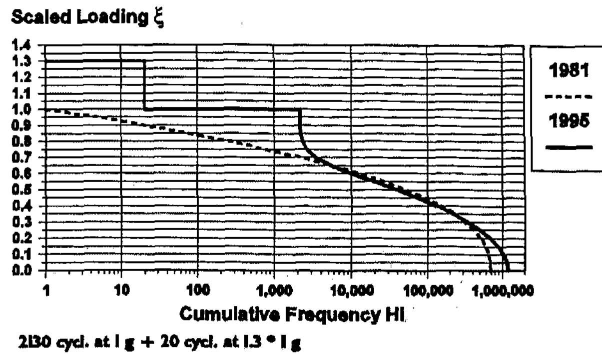


Figure 6: New Standard Brake Load Collective

WOEHLER (S-N) CURVES - A Woehler curve must be plotted for every brake or component to be released. This should be based on the brake assembly or its weakest part.

When testing components it must be ensured that the weakest part always fails first. 12 items (brake or component) minimum are to be tested, at two loading levels (6 items min. per load case) and the $P_{\ddot{u}}$= 50% curve drawn.

The failure criteria are either a 1 to 2 mm technical crack or a dynamic load drop of 5%. The $P_{\ddot{u}}$=99.9% survival probability curve is calculated using the factors $J_{C,n}$ and J_L.

Although the Woehler curve was calculated for 12 items, $J_{C,n}$ remains at the value 1.26, which affords us additional safety. With n=12: $J_{C,12} = 1.14$.

The equation for the 99.9% survival probability Woehler curve is:

$$N_{99.9\%}=\frac{N_{50\%}}{J_{C,n}*J_L} \tag{7}$$

$N_{99.9\%} \geq N_{1k(D=1)}$

REQUIREMENT

Using the new collective as described above and taking the slope (k-factor) of the Woehler curve and the individual frequencies n_i into account, we calculate the cycle count of the rectangular collective of equivalent consistent maximum load test.

THE N_{1k} - FACTOR - The minimum requirement for a brake is covered by the simple equation:

$$N_{1_{k(D=1)}}=[\Sigma(\frac{F_i}{F_1})^k*n_{i_k}] \tag{8}$$

$N_{1k(D=1)} \leq N_{99.9\%}$

However we know that we cannot always use D=1 (rel. Miner = 1) in our calculations. we must introduce a factor which leads to the specified cycle count.

THE $N_{1(spec)}$ - FACTOR - From the literature [13] [14] it is apparent that the rel. Miner-Count is dependent on the collective, the component and the material.

It follows that different brake designs:

- **Hammerhead (Brake Supplier 2)**
- **Colette (Lucas)**
- **Frame-type caliper**
- **Fixed (opp.piston) caliper**

are allocated a different D-value.

Service load testing at Brake Supplier 2 and at Darmstadt TH (university) showed a clear effect due to design differences.

Tests also showed a strong dependence on crack length and excess loading.

For 1.3 x max.load and 1g a rel. Miner count results from different crack lengths.

The max. collective load for the scaled loading factor ξ_i=1 during simulated service testing must not exceed 1g; plus the 20 cycles at 1.3 * 1g.

	1.3 X 1.0g	1.0g
0.3 mm:	D=0.2	D=0.6
0.6 mm:	D=0.7	D>1
2.0 mm:	D=1.2	D>1
Break:	D=2.6	D>1

Taking 1g as the highest loading figure resulted in a Miner-count of > 1.

The supplier gave for each brake type the D-factors to achieve $N_{1(spec)}$.

The equation therefore reads:

$$N_{1_{k(D=X)}} = N_{1_{k(D=1)}} * J_{C,n} * J_L * \frac{1}{D} \qquad (9)$$

FATIGUE LIFE AND TEST SPECIFICATION

All relevant parameters have now been developed or worked out. The following test specification is based on them. The collective has a more realistic loading profile. We are, nevertheless, still waiting for the test results of the working circle "Customer-Collective.

During the development stage the Engineering will carry out simulated service load tests based on the finalised collective in order to reduce the imponderables of the Miner-Factors and further inaccuracies to a minimum.

Single stage test on a complete caliper assembly on a pulsating disc, with pads of the lowest possible friction, connected to the disc via a three-way ball joint. Alternatively a test rig with original axle parts set up to drawing and including a rotating disc (friction grip) under controlled torque. Brake torque increase to lag pressure rise and torque decrease to synchronise with pressure fall.

Moment M_E and pressure p equivalent to a 1g stop, torque tolerance +5%; pressure tolerance +20%. M_E to be set by calculation equivalent to the ideal or installed brake force distribution for gross vehicle weight. The max. value has to be taken.

The required cycle count is $N_{1k(spec)}$.

If k-factors lie between given figures the next lower figure is to be chosen, in order to keep N_{1k} on the safe side, (e.g. 4.84 becomes 4.5).

The required cycle counts relating to the slope of the Woehler curve can be determined from the following table:

The required minimum cycle count is

$N_{1k(spec)} = N_{1k(D=1)} * J_{C,n} * J_L$ and shown in the Table 2 below.

The diagram in Figure 7 illustrate the influence of the slope factor of the S-N-Curve (Woehler-Curve).

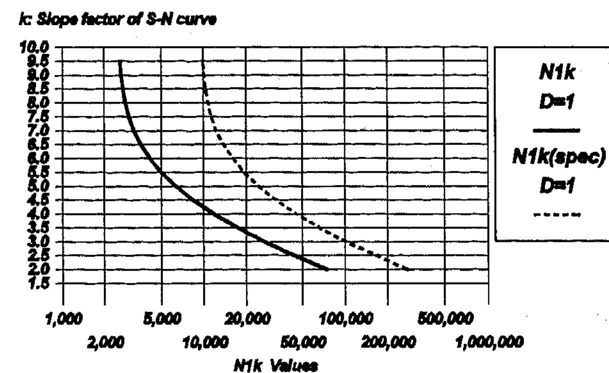


Figure 7: N_{1k} and $N_{1k(spec)} = f(k,n_i)$

Table 2: Test cycles versus slope factor k of the Woehler (S-N) curve

k-factor	Cycles $N_{1k(D=1)}$	Test cycles [*1] forward $N_{1k(spec)}=N_{1k}*J_{C,n}*J_L$ $J_{C,n}*J_L = 3.82$	Test cycles reverse 10% of forward application
2.0	**74546**	**275000**	**27500**
2.5	**43633**	**167000**	**16700**
3.0	**26832**	**102500**	**10250**
3.5	**17329**	**66500**	**6650**
4.0	**11766**	**45000**	**4500**
4.5	**8411**	**32500**	**3250**
5.0	**6334**	**24500**	**2450**
5.5	**5017**	**19500**	**1950**
6.0	**4167**	**16000**	**1600**
6.5	**3608**	**14000**	**1400**
7.0	**3237**	**12500**	**1250**
7.5	**2988**	**11500**	**1150**
8.0	**2822**	**11000**	**1100**
8.5	**2713**	**10500**	**1050**
9.0	**2644**	**10500**	**1050**
9.5	**2604**	**10000**	**1000**

[*1] *The test cycles are rounded to the next 500.*

The cycles for reverse applications have to be carried out at the beginning of the test
(i.e.: For k = 7: 1250 cycles reverse + 12500 cycles forward)
For brakes with an integrated handbrake (combined brake) the following additional test has to be carried out with the same brake parts:

15 000 cycles with 400 N hand brake lever load or 500 N foot lever load without tangential force to the brake carrier.

5 000 cycles with a hand brake lever load or a foot lever load with tangential force to the brake carrier corresponding 30% hill hold.

The procedure is now clearly defined, it follows the rules below:

Table 3: Test Procedure

Evidence:	Test Procedure:
1. First tests	Identify the weakest component by means of FEA analysis or test.
2. Woehler (S-N) curve	Determine the k-factor for the Woehler curve by testing $n \geq 12$ components or assemblies ($n \geq 6$ each load level). Find $N_{50\%}$ and $N_{99.9\%}$.
3. relative Miner-count	To be fixed for each brake type.
4. N_{1k} D=1	$$N_{1_{k(D-1)}}=[\Sigma(\frac{F_i}{F_1})^k * n_{i_k}] \quad (8)$$ = $f(k,n_i)$ elementary Miner
5. Minimum test cycle count $N_{1k(spec)}$ D=X	$$N_{1_{k(D-X)}}=N_{1_{k(D-1)}}*J_{C,n}*J_L*\frac{1}{D} \quad (9)$$ Test $n \geq 4$ up to $N_{1k(spec)}$ cycles without failure, then break off. Torque, pressure and cycles count given by suppliers.
6. Woehler curve for calculation. (The rel. Miner number will be use during life prediction.)	$$N_{99.9\%}=\frac{N_{50\%}}{J_{C,n}*J_L} \quad (7)$$ 99.9% Survival probability; determined using $J_{C,n} * J_L$; Assumption: log Normal-distribution.
7. Endurance life determination	Woehler curve from component or assy. test versus collective with Miner elementary. L*D > 300 000 km.

The final test procedure can also illustrated in a diagram and Figure 8 shows the theoretically S-N-Curves and the experimental S-N-Curves.

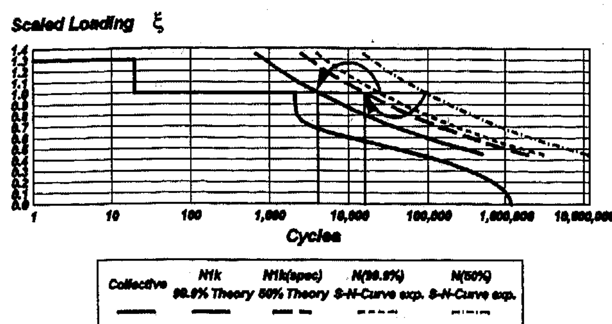


Figure 8: Theoretical and Experimental S-N-Curves

Now it is clearly seen that in the worse case the experimental curves will meet the theoretical curves calculated from the k-factors and the statistical factors J_L and $J_{C,n}$.

That means:

The 99.9% survival probability curve must fulfill 300 000 km and this is the N_{1k} curve (worse case) for a rel. Miner No. D=1 or as described in Point 7 in Table 3: L * D = 300 000 km min.

Additional to the Fatigue Life other tests has to be carried out.

SEAL TEST- Load to p = 100 bar (Pressure increase rate approx. 500 bar/s, piston stroke: 0.3 mm)

for 135 000 cycles at: RT; f=800-1200 [1/h]
45 000 cycles at: -40 ± 2°C; f=600- 800 [1/h]
70 000 cycles at: +150 ± 2°C; f=800-1200 [1/h]

After 250 000 cycles total:
the following conditions must be met:
- no hydraulic system leakage
Test pressure p = 150 bar; $\Delta p/\Delta t \leq 10$ bar within 2 minutes.
- no wear on piston or cylinder ID which impairs performance.
- Piston push-out pressure: 0.2 - 3.0 bar
- Piston push-in force: 50 - 700 N

Before measurements: Piston to move ± 5 mm 2 time; then the piston has to be moved back in position of new pad condition and 1 static pressure application with 100 bar.
Spot check quantity n ≥ 4.

Only one release is required for each housing diameter, providing that the expansion is in the same tolerance range and that the geometric results (e.g. roll-back) and the materials are identical.

SERVICE BRAKE - M_{ST} = 1.75 M_{req} + 5%; pressure and moment rise together.; M_E and p from 3.4; after M_{ST} has been applied there must be no hydraulic leakage and the following conditions must be fulfilled:

Housing sliding force:	10 - 150 N
Piston push-out pressure:	0.2 - 3.0 bar
Piston push-in force:	50 - 700 N

Before measurements: Piston to move ± 5 mm 2 time; then the piston has to be moved back in position of new pad condition and 1 static pressure application with 100 bar.
Spot check quantity ≥ 4.

SERVICE BRAKE WITH INTEGRATED PARKING BRAKE
- This test has to be carried out at a suitable test stand.
Installation corresponding to the required procedure.
One "reverse" brake application with a brake clamping force corresponding to a hand brake lever load of 800 N or a foot lever load of 1000 N. The lever ratios, the lever geometry of the appropriate cars and the efficiencies have to be taken into account.

No functional failure of parking brake.

Spot check quantity ≥ 4.

BURST PRESSURE - p > 350 bar
Spot check quantity ≥ 4.

SUMMARY

By developing the new, state of the art, test specification, we have been able to cut through much of the present tangle and clear the way for the future.
The same procedure is valid for other brake components (brake booster etc.).
Since 1981 we have used **Fatigue Life Prediction Procedures** in testing all our brake component prototypes and in their further development.
All brakes were analysed before release using the Palmgren-Miner Rule against the collective described.
We have built more than 200 million brakes of the Colette type for world-wide sale. To date not a single component has failed.

ACKNOWLEDGEMENT

I say thank you to all my colleagues at Lucas CAE department for the discussions during the preparation of the procedure and to my colleagues from ITT Automotive: K.Stoerzel and H.Rueckert and to the members of the working circle : A.Pradt, Chairman (Mercedes Benz), K.H.Boss (BMW), M.Luedeke (AUDI), R.Martin (PORSCHE) and Dr.L.Jung (VW) for the help during the discussions.
Thank you also to Prof. Dr.-Ing. V. Grubisic from the Fraunhofer Institute of Operating Strength in Darmstadt (Germany) and Prof. Dr.-Ing. Erwin Haibach for the help in the statistical determination of the load collective and safety factors.

ABOUT THE AUTHOR

The author Dipl.-Ing.(FH) Hermann Josef Tumbrink is Chief Engineer CAE/Fatigue, Simulation at Lucas Varity, Light Vehicle Braking Systems, Koblenz, Germany. He has 20 years experiences with fatigue on brake components.
His mailing adress is: TUMBRIH@LIKOBL2.LI.CO.UK
Telephone and Fax: +49 261 895 515 Phone
+49 261 895 601 Fax
+49 2641 34423 home phone
+49 2641 34408 home phone
+49 2641 34407 home Fax

NOMENCLATURE LIST

ξ_i	Scaled load factor of step i
σ_1	Max. stress
σ_i	Stress at step i
C	Confidence level
CAD	Computer Aided Design
CAE	Computer Aided Engineering
CAM	Computer Aided Manufacturing
D	Relative Miner
F_1	Max. load
FEA	Finite Element Analysis
F_i	Load at step i
H_0	Zero passage of the collective
H_i	Cumulative frequency at load step i
$J_{C,n}$	Risk factor for confidence level C at n test samples
J_L	Safety factor to be applied to fatigue life
k	Slope factor of Woehler (S-N) curve
n	1. Exponent of the collective 2. Number of test samples
$N_{1k(D=1)}$	Number of cycles for Miner No.: D=1
$N_{1k(spec)}$	Test cycles for a rel. Miner No.: D=1
$N_{1k(D=X)}$	Test cycles for a rel. Miner No.: D=X
$N_{50\%}$	Number of cycles for 50% survival prob. for Woehler curve
$N_{99.9\%}$	Number of cycles for 99.9% survival prob. for Woehler curve
p_1	Max. pressure
p_i	Pressure at step i
$P_ü$	Survival probability
S	Standard deviation
S_F	Statistic safety factor
S_{log}	log standard deviation
T_N	Scatter of 90% to 10% probability
U	Standardised normal distributed variate
U_C	Scaled variate for determination of C

REFERENCES:

[1] Tumbrink, H.J.; Stroh, D: "The ASSET-Technique - Assessment of Simulated Service Endurance Test", SAE Congress and Exhibition 1979, USA 1979

[2] Wimmer, Dr. A.: "Erarbeitung fahrzeugunabhängiger Parameter zur betriebsnahen Dimensionierung lebenswichtiger Bauteile". Vortrag zur 10. Sitzung des AK-Betriebsfestigkeit im DVM. 7. u. 8.2. 1984

[3] "Moderne Betriebsfestigkeitsprüfung", 16. Vortragsveranstaltung des DVM-Arbeitskreises Betriebsfestigkeit im Oktober 1990, S.278

[4] Maaß, H. : Kloeckner-Humbold-Deuz AG, Koeln: "Welche Hilfe kann die Betriebsfestigkeitslehre dem Konstrukteur im Entwurfsstadium geben?"; 4. Sitzung des Arbeitskreises Betriebsfestigkeit, 12.10.1978, Darmstadt, Berichtsband Seite 37-45

[5] Zenner, Prof. Dr. H.: "Überlegungen zur Betriebsfestigkeitsprüfung" TU Clausthal, DVM Arbeitskreis Betriebsfestigkeit 1990 Berichtsband

[6] Tumbrink, H.J.: "Betriebsfestigkeit im PKW Bremsenbau" Eigenvertrieb: H.J. Tumbrink, Ahrallee 21, D-53474 Ahrweiler. Original vergriffen, nur Kopien.

[7] Lütteke H./Zaschel J.M.: "Zur formelmäßigen Beschreibung von Häufigkeitsverteilungen mit Hilfe von Exponentialfunktionen". TM-Nr.: 81/77, LBF Darmstadt

[8] Zenner, Prof. Dr. H.: "Leichtbau und Betriebslasten" TU Clausthal, DVM Tag 1991, 6.-.8.5. in Berlin, Berichtsband S.184 ff.

[9] Tumbrink, H.J. /Stoerzel, K. : "Testvorschrift Radbremse." Gemeinsamer Vorschlag für AK-Bremsen im Strategiekreis Betriebsfestigkeit der deutschen Automobilhersteller. Lucas Automotive GmbH/ITT Automotive, 1996

[10] Brunner, Franz J. : "Angewandte Zuverlässigkeitstechnik bei der Fahrzeugentwicklung -Teil 1 und 2"; ATZ 89 (1987) 7/8

[11] "Qualitätskontrolle in der Automobilindustrie", Zuverlässigkeitssicherung bei Automobilherstellern und Lieferanten. Verfahren und Beispiele, 2., überarbeitete u. erweiterte Auflage. Verband der Automobilindustrie e.V. (VDA), Westendstr. 61, Frankfurt/Main, Copyright 1984

[12] Haibach, Prof. Dr.-Ing. E.; Ostermann, H.; Köbler, H.-G. : "Abdecken des Risikos aus den Zufälligkeiten weniger Schwingfestigkeits - Versuche", TM68/78 LBF Darmstadt 18. Mai 1973

[13] Haibach, Prof. Dr.-Ing. E.: "Betriebsfestigkeit", Verfahren und Daten zur Bauteilberechnung, VDI-Verlag GmbH, Düsseldorf 1989, ISBN 3-18-400828-2

[14] Neugebauer, Dr.-Ing. J.; Grubisic, Dr.-Ing. V. : "Zum Betriebsfestigkeitsnachweis von Motorradfahrwerkskomponenten", Fortdruck aus VDI-Berichte Nr.: 657 (1987), Seite 407 bis 427; LBF, Darmstadt.

971041

Development and Improvement of Finite Element Side Impact Dummy (EUROSID) Model Based on Experimental Verifications

Copyright 1997 Society of Automotive Engineers, Inc.

Chinmoy Pal, Hideaki Ichikawa, and Koichi Sagawa
Nissan Research Center

ABSTRACT

In side-crash phenomena, finite element modeling is essential in investigating the occupant's post-impact dynamic behavior after contact with the door panels. A number of modifications have been made to the model described here based on combined simulation and experimental verifications of the dynamic and pseudo-static characteristics of different materials such as foam, damper and individual sub-assemblies. This report illustrates how the modified material and structural modeling of different components improve the accuracy of the overall dynamic behavior of the FEM model in simulating different HYGE experiments to speed up and optimize the vehicle design process. The rib-module drop test results with two different polypropylene pads clearly indicate the effect of the pad unloading characteristics on rib displacement. The verification of HYGE test results, showing the variation in response with two different speeds and relative timing of impact at the chest, abdomen and hip, clearly indicates that the present model can be effectively used in selecting the optimum deformation mode of door and B-pillar structures to control the input load on the occupant. The shoulder clavicle is found to be extremely important in controlling the arm movement which influences the load path to individual ribs. Proper modeling of the jacket is also found to be important in predicting the influence on lower rib deformation.

INTRODUCTION

More stringent safety regulations worldwide have made the development of a vehicle a difficult process. The automotive industry has been working hard over the last decade in searching for new methods and processes to develop new models to meet the market timing and safety standard requirements. Consequently, various computer-aided engineering tools have been integrated in the vehicle design process and have become key elements in reducing the development time and number of prototypes required for new vehicles. Two major side impact test procedures are used for verifying the ability of a vehicle to protect the occupants in a side impact. The proposed European Experimental Vehicles Committee (EEVC) procedure is based on the injury level of a driver measured with the EUROSID-I dummy, and in the National Highway Traffic Safety Administration (NHTSA) procedure, the injury level of the driver and that of the rear driver's side passenger are measured using two SID dummies. As full-scale side impact tests are expensive and time-consuming, they are supplemented with sled tests and extensive computer simulations. Numerical simulation techniques are commonly used to assess the crash performance of automobiles and guide their design during the development stage. Mathematical models of vehicle structures, restraint systems, and dummies are developed and verified under different conditions to ensure effective usage. Several analytical dummy models have been developed to simulate occupant responses in front and side impact phenomena. These models can be broadly classified into two groups: (i) rigid bodies of ellipsoids connected by springs i.e. mass spring models and (ii) elastic/plastic bodies of proper material strength and geometrical shapes connected together with or without springs and joints. Although rigid mass-spring dynamic models are widely used in frontal crash simulations because of their ease of modeling, they can only capture 2-D gross movements. Such ellipsoidal models fail to represent the actual 3-D contact mechanism which is extremely important in side impact phenomena. However, the recent versions of some of the numerical codes, such as MADYMO, based on mass-spring models, have included some extra features to simulate and visualize more realistic crash phenomena. On the other hand, the later approach based on finite element techniques is capable of predicting very complex and complicated responses such as deformation, contact force and so forth which can hardly be measured experimentally.

This paper describes the validation of the EUROSID FEM model. Some of the basic modifications that are incorporated to improve the degree of accuracy of the present model are described in later sections. The component level modifications are based on identifications of (i) static responses of hip-pelvis foam compression tests, (ii) dynamic responses of lumbar spine and neck bending tests and (iii) dynamic responses of rib module drop tests. The assembly level modification includes (i) identification of the effect of upper-lower connection on pubic symphysis force with the help of rigid impactor pelvis tests and (ii) identification of the effect of the pre -tension clavicle string on thorax rib displacements. With the help of a tentative HYGE impact configuration very close to an actual crash test, the overall dynamic behavior and different injury response criteria, such as, the rib displacement criteria (RDC), the abdominal peak force(APF) and the pubic symphysis peak force(PSPF), are initially verified at a speed of 35 km/hr. Further validation is made for different sets of configurations by changing the input force timing at the abdomen and hip in order to simulate a board spectrum of probable input patterns in real side crash tests. The explicit finite element code RADIOSS is used to obtain the numerical solutions. Finally, the data are also correlated with the results obtained at a lower impact speed of 25 km/hr.

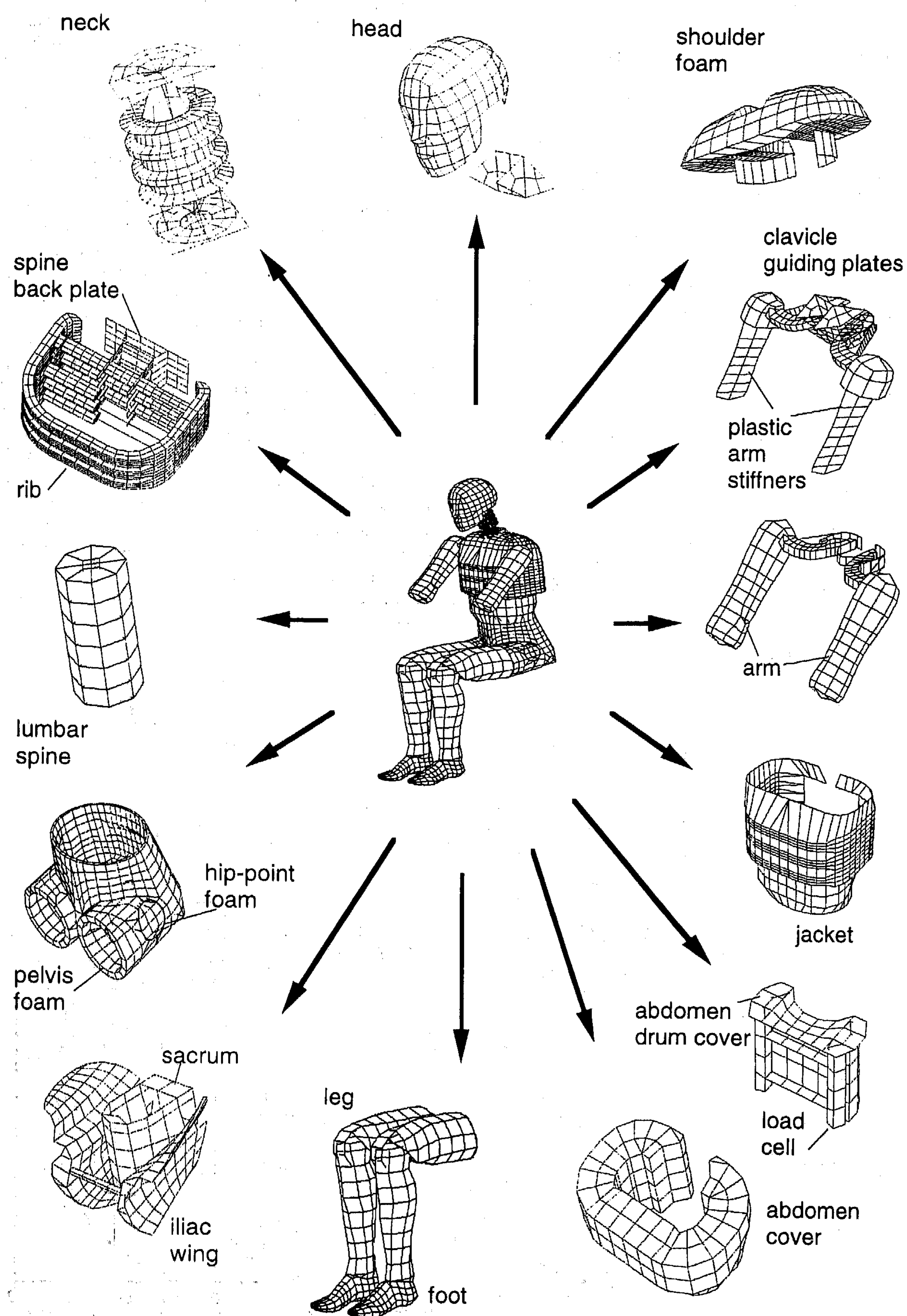


Figure 1. Different Parts of EUROSID Dummy FEM Model

FINITE ELEMENT MODEL

The overall geometry of the different parts of the present model are shown in Figure 1. It is originally based on the MECALOG model and consists of solid brick, thin shell and 6 degrees of freedom spring and beam elements. Different parts which undergo small deformation, such as the spine box, head and the legs below the hip are treated as rigid bodies to reduce the computational time by eliminating the extra degrees of freedom for unimportant parts. However, care has to be taken to represent the 3D shape of each of them for accurate estimation of contact forces. To improve the degree of accuracy of the present model, the following items have been checked and modified to obtain good correlation between the experimental and numerical results.

IDENTIFICATION OF INDIVIDUAL PARTS

(a) Nonlinear Characteristics of Hip and Pelvis Foam

The hip and abdomen part are modeled with sophisticated porous foam material. However, to set the parameters of the material requires accurate tests of samples of the same material. Further, it is to be noted that these parts are made airtight with a thin outer polyvinyl rubber completely covering the inside core foam (Figure 2a). Since the in-plane membrane stiffness of the outer rubber skin and the shape of the individual items greatly affect the overall equivalent compression characteristics of the item, a simple material test of the inner core material will not truly represent the actual compression behavior. Hence, as an alternative, direct compression tests of the relevant part, say the hip point flesh of the EUROSID dummy, were conducted to find the equivalent stress-strain relationship from the measured force-displacement characteristic under different compression conditions, the details of which are described below. The unknown material characteristics of the nonlinear foam material were determined by an iterative optimization procedure based on the concept of pseudo-static simulation performed by explicit nonlinear finite element codes. Figures 2b-2d show the process of evaluating the equivalent volumetric stress-strain characteristics. For example, the abdomen was initially tested in a frontal compression and the material characteristics were identified by the above-mentioned procedure. Using these material characteristics, it was then tested in a lateral compression to check the validity of the previous stress-strain characteristics. A comparison of the test and the final simulated results (Figure 2e) indicates fairly accurate identification of the nonlinear characteristics of the hip foam material. Similar results were obtained in a pseudo-static test of the pelvis foam.

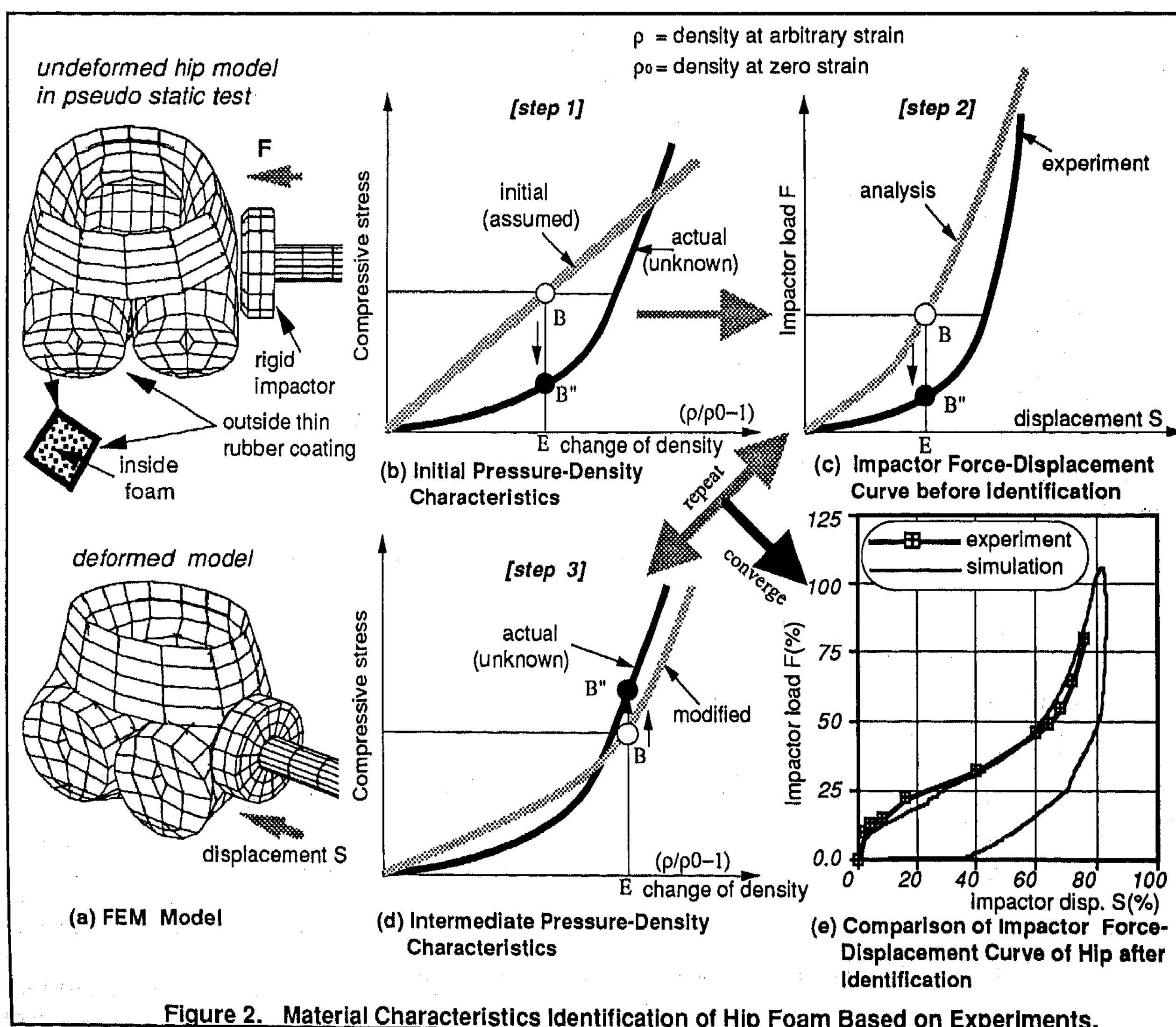


Figure 2. Material Characteristics Identification of Hip Foam Based on Experiments.

(b) Nonlinear Dynamic Characteristics of Lumber Spine and Neck

Since the lumber spine and the neck connect the hip and the head of the dummy with the spine box at the bottom and at the top of the spine, both of them behave as cantilever beams. Numerical simulation of a free fall pendulum (Figure 3a) as stipulated in the EUROSID manual [2] was performed for input pendulum acceleration (Figure 3b) having an impact velocity of 6.05 ± 0.1 m/sec in the direction of motion. A nonlinear solid material modeling (material 35 of RADIOSS) was chosen based on the Maxwell-Kelvin-Voight viscosity as shown in Figure 3c. Figure 3d indicates that the present nonlinear model has satisfactory accuracy. Typical deformation modes are shown in Figure 3e. Similar results were obtained in a pendulum test of the neck assembly.

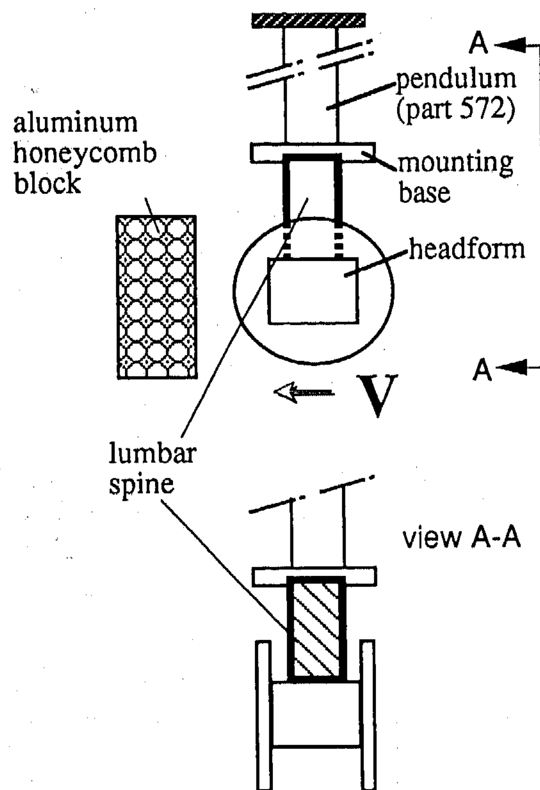


Figure 3a. Pendulum Test

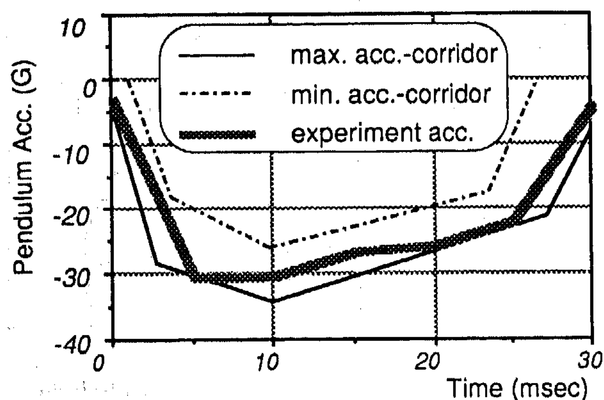
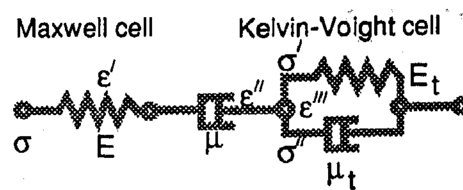


Figure 3b. Pendulum Input Acceleration of Lumbar Spine (V = 6.05 m/sec)



Basic relationship

$$\varepsilon = \varepsilon' + \varepsilon'' + \varepsilon''' \qquad \varepsilon' = \sigma / E$$

$$\sigma = \sigma' + \sigma'' \qquad \dot{\varepsilon}'' = \sigma / \mu$$

$$\dot{\varepsilon} = \dot{\sigma}/E + \sigma/\mu + \dot{\varepsilon}''' \qquad \varepsilon''' = \sigma' / E_t$$

$$\sigma = \mu_t \dot{\varepsilon}''' + E_t \varepsilon''' \qquad \dot{\varepsilon}''' = \sigma'' / \mu_t$$

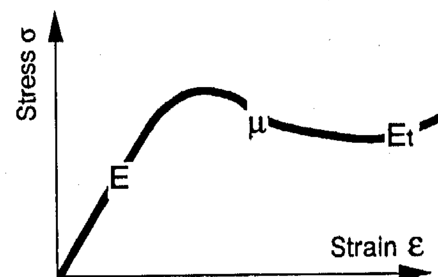


Figure 3c. Nonlinear Damping Material Model (Maxwell-Kelvin-Voight)

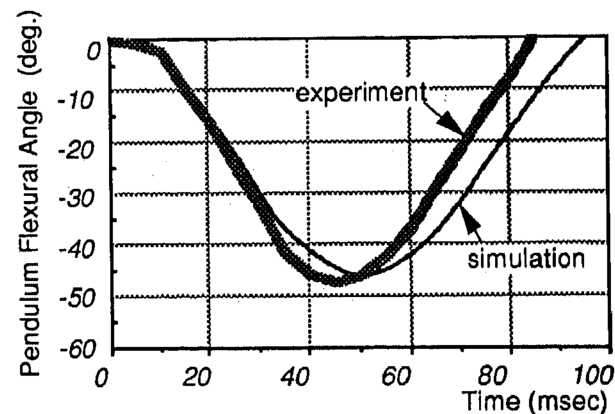


Figure 3d. Pendulum Angular Deflection of Lumbar Spine (V = 6.05 m/sec)

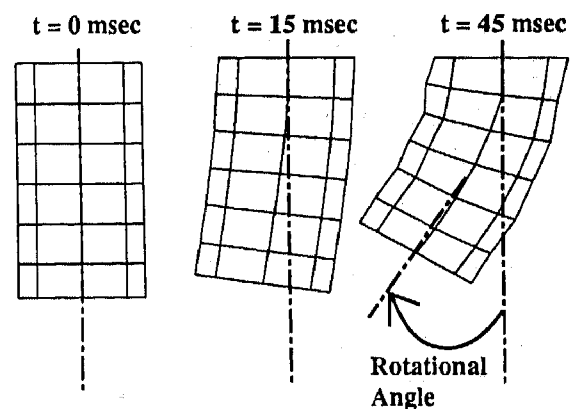


Figure 3e. Bending Mode of Lumbar Spine

(c) Effect of Nonlinear Dynamic Characteristics of Polypropylene on Rib Module in Drop Test

The rib module assembly was tested in a standard drop test (Figure 4a, 4b), as stipulated in the EUROSID manual, using (i) a rigid impactor and (ii) soft polypropylene pad 45PP. A comparison of the experimental results are shown in Figure 4e. The padded drop-test results are comparatively shorter in peak height, flatter at the top and longer in duration. The shortness is due to absorption of the input energy in the loading curve of the pad. The flatness of the peak is due to the high rebound effect of the pad during the unloading phase, and elongation of the time period is due to the small amount of energy loss to hysteresis. Accurate modeling of this part is extremely important in simulating the exact deformation characteristics such as the damper, return spring and tuning spring. Padded impacts were conducted to investigate the dynamic behavior of the thorax when impacted by the door trim. The pad was modeled with material 33 of RADIOSS [1] to incorporate the effect of hysteresis. Material law 33 is very similar to material law 35 as described in the previous section. A comparison of the simulation and padded impact test results, as shown in Figure 4f, indicates fairly accurate identification of dynamic response. The significance of this test from the hysteresis point of view will be more clear in later sections where full scale sled test results are compared.

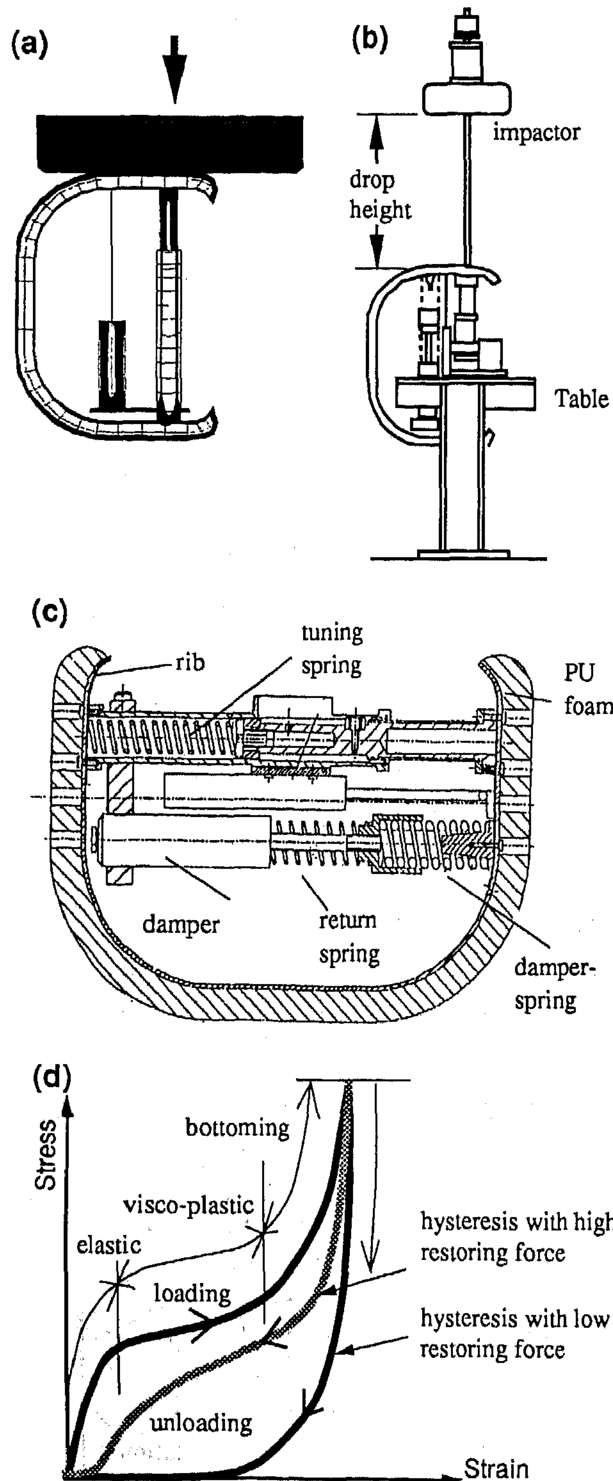


Figure 4. (a) FEM Model (b) Experimental Set-up (c) Rib Module (d) Stress-Strain Characteristics of Pad

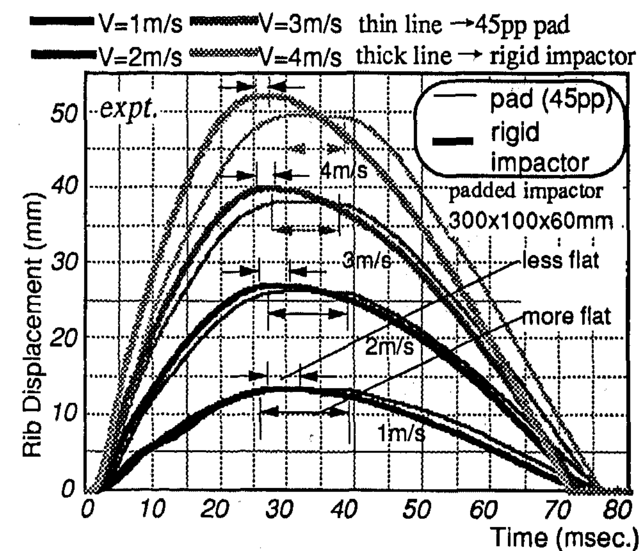


Figure 4e. Rib Drop Test Results for Different Impact Speeds V=1.0,2.0,3.0,4.0 m/sec.

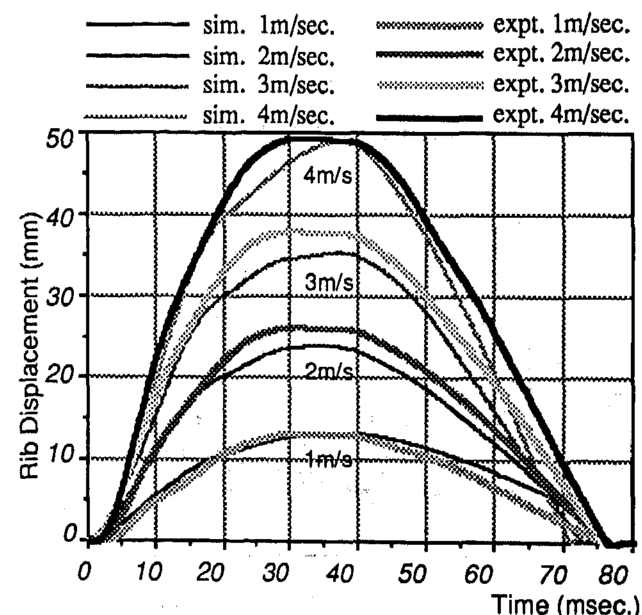


Figure 4f. Comparison of Drop Test Results and Simulation for V=1.0,2.0,3.0,4.0 m/sec.

IDENTIFICATION AT ASSEMBLY LEVEL

(a) Modeling of Lower Part of Leg and Its Effect on Pubic Symphysis Peak Force (PSPF)

The outer surfaces of the upper leg and the lower leg are modeled with shell elements. The inside rigid rod is modeled with beam elements and the joint between the upper and lower leg is modeled with general springs as shown in Figure 5a. Since the product of the mass of the legs and the distance of its center of gravity (G2) from that of the whole dummy (G1) is of a very high order, it greatly affects the total movement of the dummy as whole, especially for lower end impacts, say at the hip point. Hence, it is a key factor in designing the door and surrounding structures. The effect of legs on PSPF was studied by numerical simulation for two types of models: (i) one with beam elements to incorporate the inside steel rod and the links and (ii) the other without beams at an impact velocity of 32km/hr in the direction of motion. Figure 5(b-e) clearly illustrates the validity of the present model which satisfactorily predicts the impactor deceleration, PSPF and the overall lateral movement of the dummy as calculated from the response of the line joining the pelvis, lower spine and upper spine.

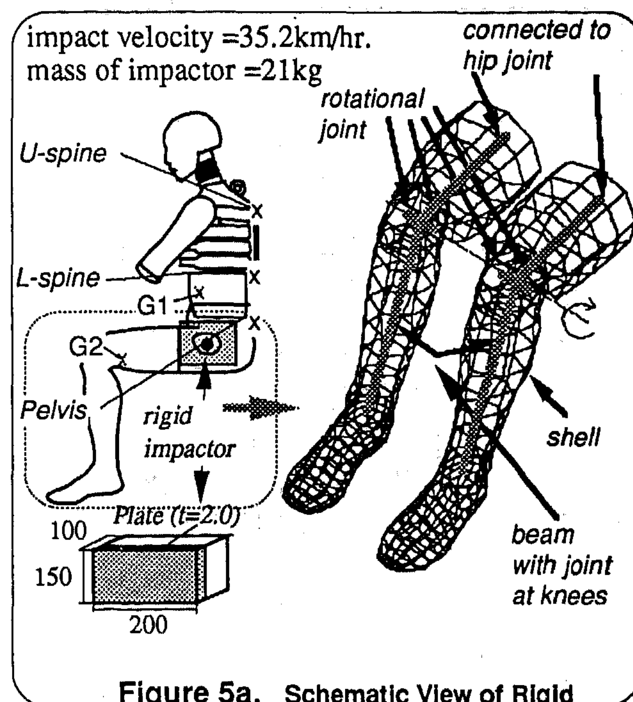


Figure 5a. Schematic View of Rigid Hip Impact Test.

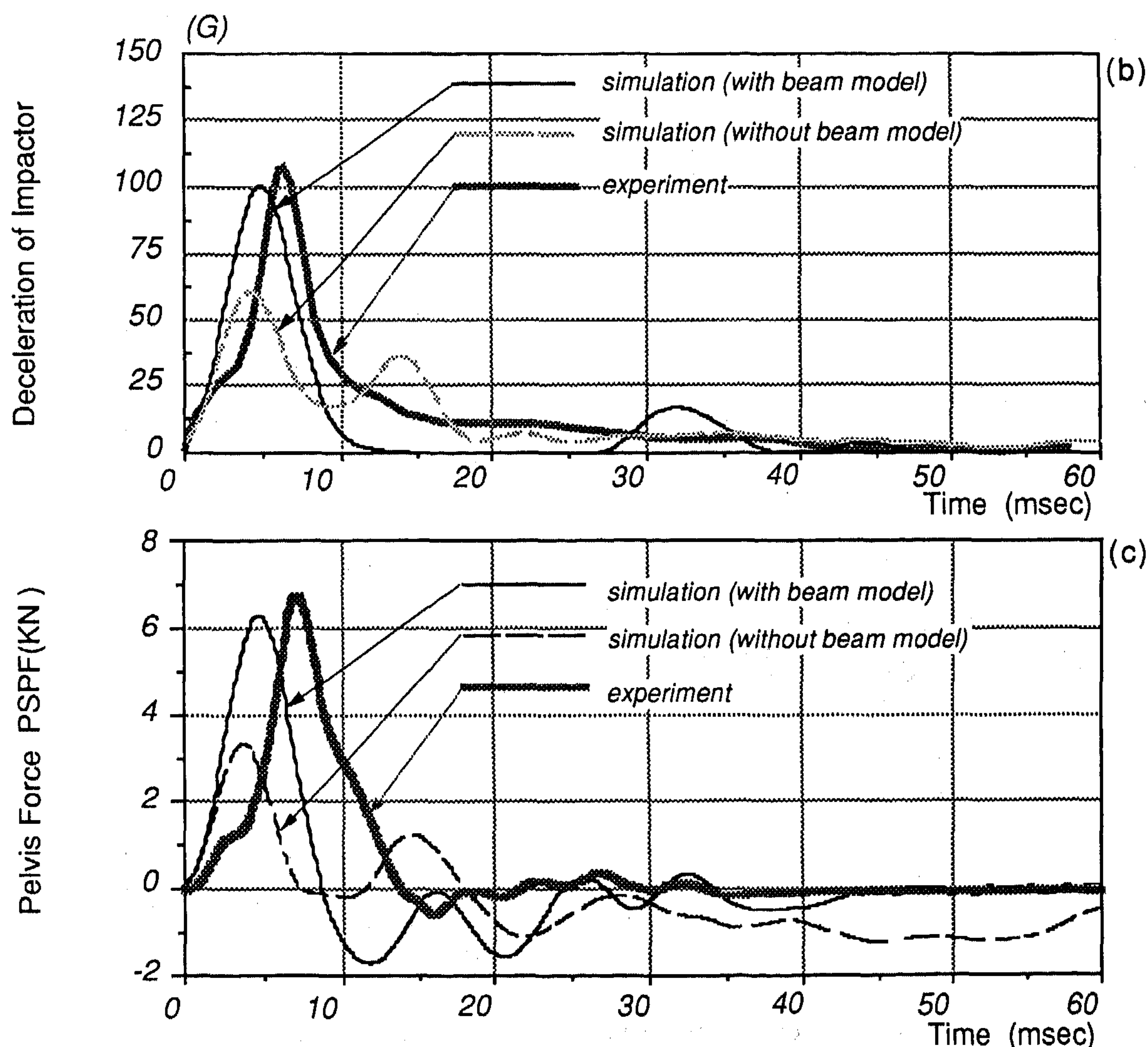


Figure 5. Comparison of Time Histories of (b) Impactor Acceleration and (c) Pubic Symphysis Peak Force (PSPF)

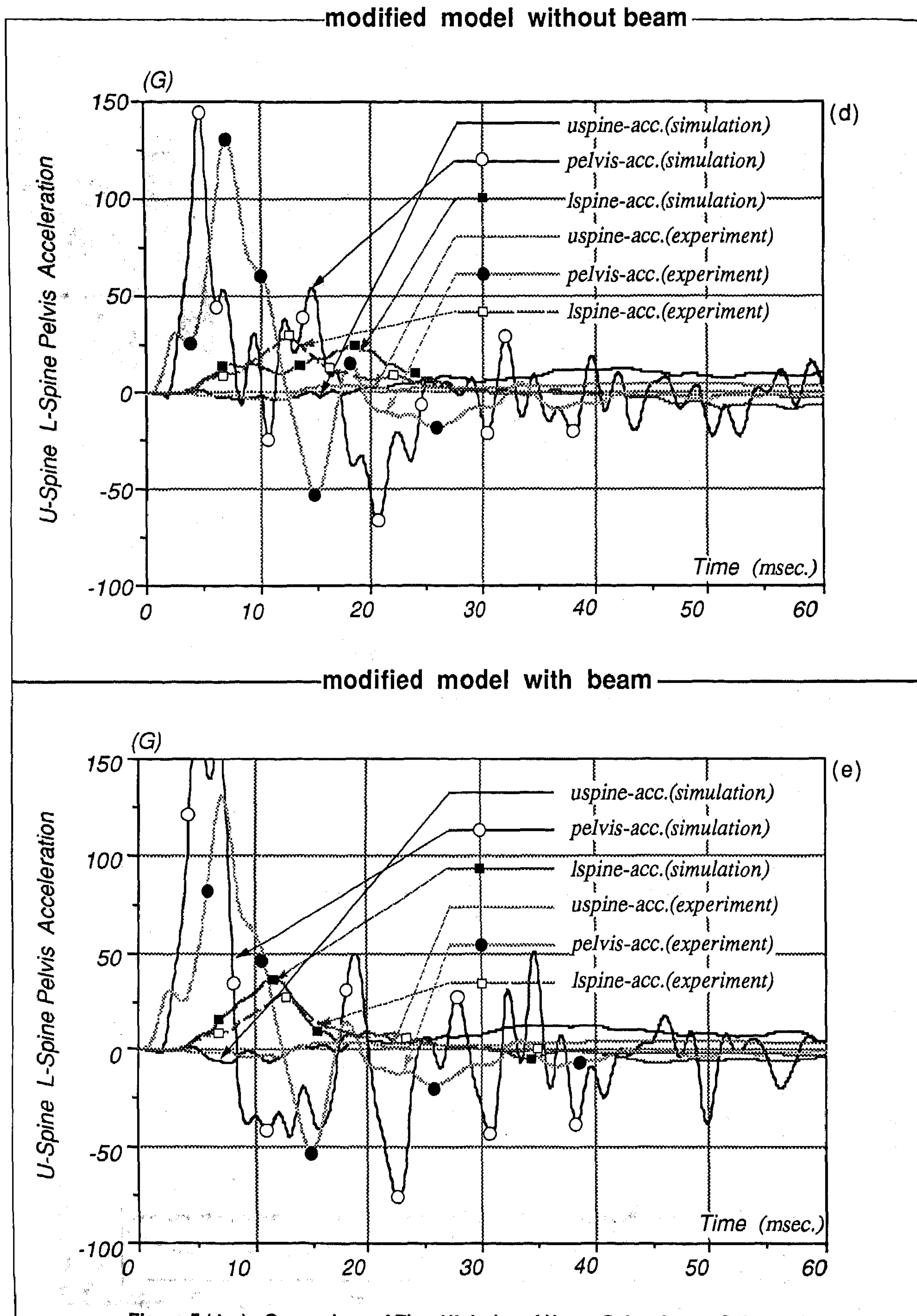
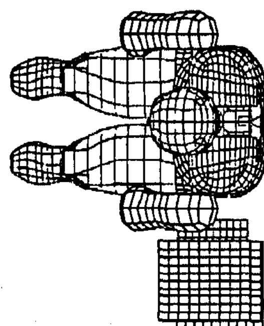
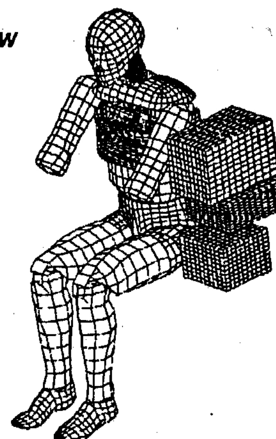


Figure 5 (d, e) Comparison of Time Histories of Upper Spine, Lower Spine and Pelvis Acceleration For Rigid Hip Impact.

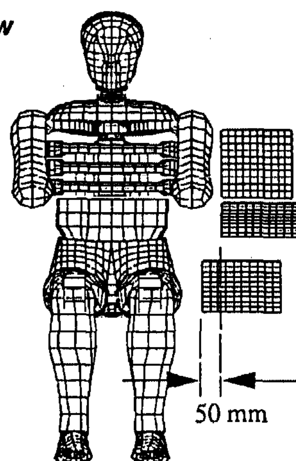
Top View



Isometric View



Front View



Side View

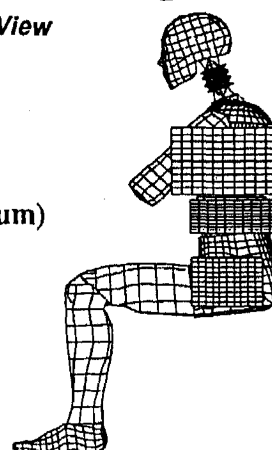


Figure 6a. Padded Sled Test Configuration with an Offset of 50 mm at Hip Pad (Case 4)

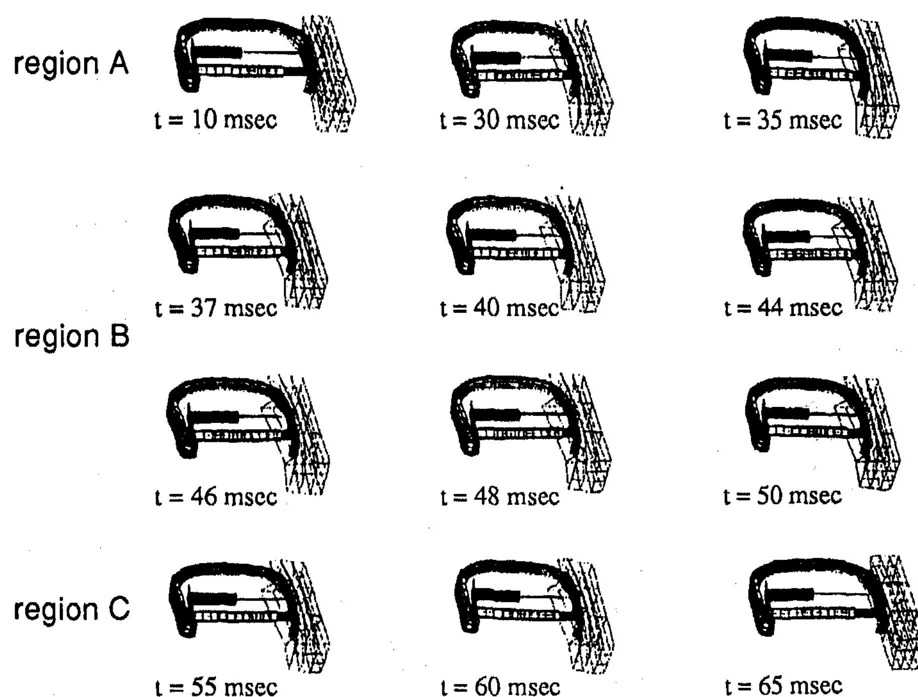


Figure 6b. Deformation of Rib Module with Padded Impactor

(b) Effect of Hysteresis of Impactor Pad and Thorax Module on Rib Displacement

Figure 6a shows the test configuration for multi-input padded sled impact tests for case 2 (refer to Figure 12 for other cases) with an offset in the hip pad. Typical deformation modes of the rib module for padded impactor drop tests are shown in Figure 6b. Figure 6c shows the effect of hysteresis on the chest deformation response. Without a hysteresis mechanism, as illustrated in Figure 6c, the predicted response of the upper rib displacement shows greater deviation from the experimental value due to a lack of pad restoration force in region B (refer to Figure 6b, d) where the relative velocity of the ribs and the spine box is very low. This mechanism can be properly explained by the force-displacement response characteristics of 45PP padded impactor drop tests for a single rib module as shown in Figure 6e. Similar results were obtained for all the other eight different cases of HYGE tests with various offsets at the abdomen and hip pad with respect to the chest pad.

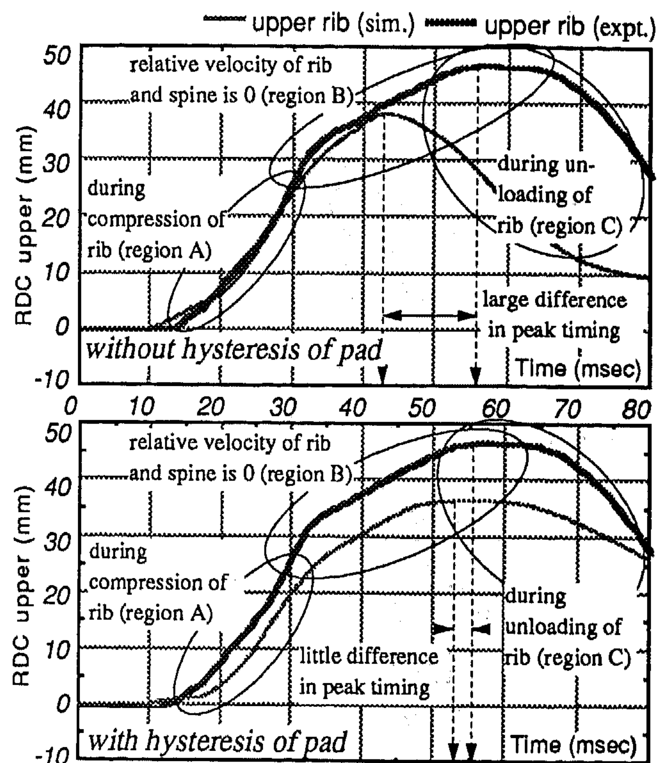


Figure 6c. Variation of RDC (case 2); V=25km/hr.

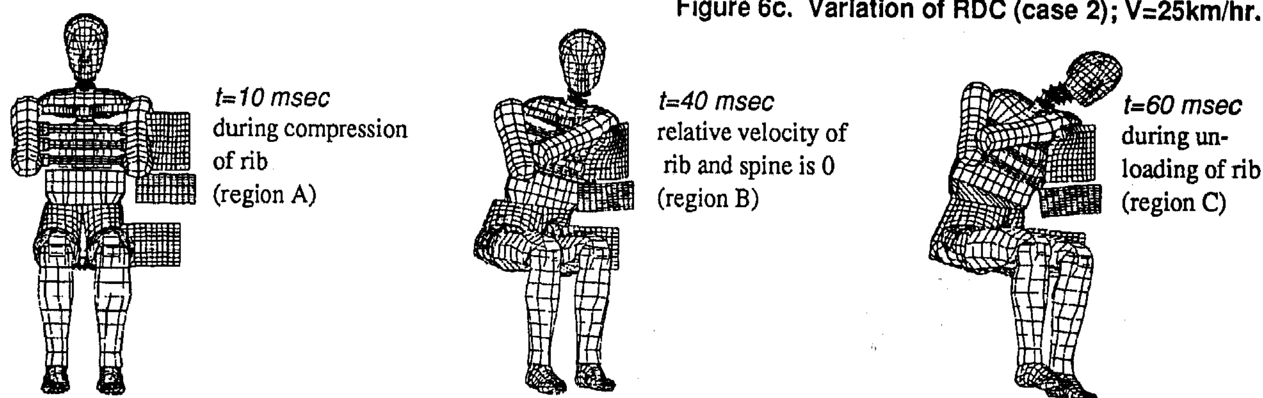


Figure 6d. Deformation Mode of Impactor Pad Corresponding to Regions A, B and C

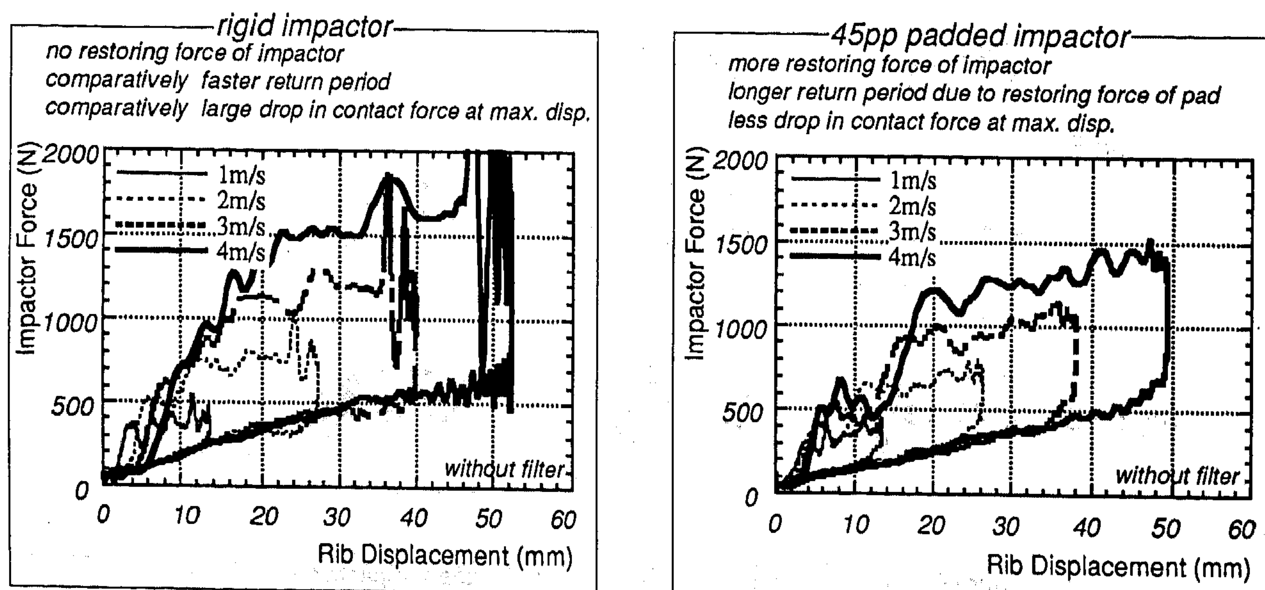


Figure 6e. Comparison of Rib Module Force Displacement Characteristics For Different Impact Speeds

(c) Effect of Hysteresis of Impactor Pad and Thorax Damper on Rib Displacement

The relative rotation of the arm is controlled by the movement of the clavicle within the top and bottom shoulder plates. The end of the clavicle is tied at the back of the shoulder block by a pre-tension elastic chord. This thin elastic string is modeled with 10 shell elements in order to simulate its contact with the sides of the shoulder block. The force displacement characteristics of the elastic chord is measured and modeled with 20 springs attached at the edges of the shell elements. Figure 7a shows the effect of the pre-tension of the elastic rod on rib displacements. Typical relative rotational deformation modes of the arm and shoulder and the consequent compression of the rib module for padded impactor tests are shown in Figure 7b. Without a pre-tension mechanism, the predicted response of lateral chest displacements shows greater deviation from the experimental values due to an improper distribution of contact force at the upper, middle and lower ribs. The distribution of contact forces at three different heights has a large influence on the relative compressive characteristics of these ribs[4]. Similar results were obtained for all the other eight different cases with various offsets at the abdomen and hip pad as shown in Figure 12. The experimental value of the lower is are not available.

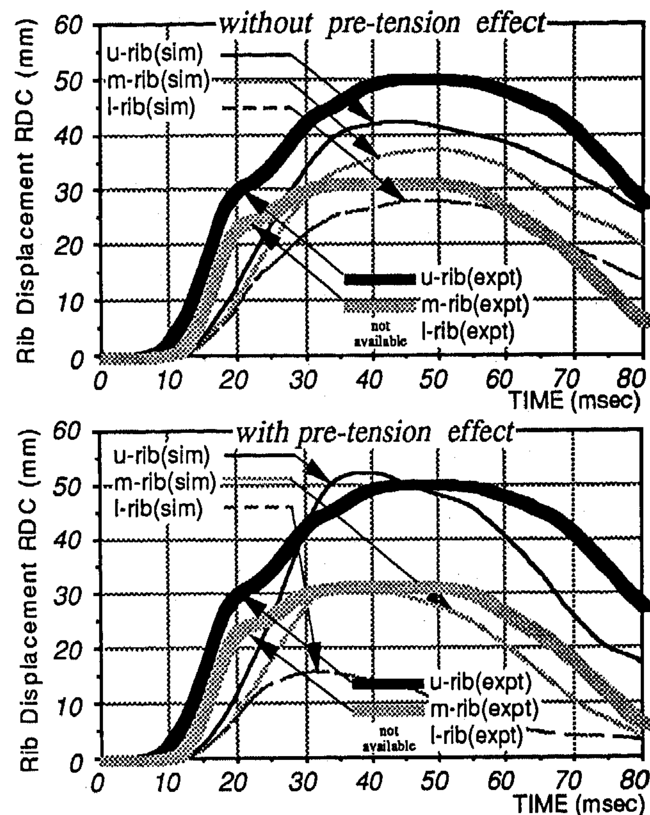


Figure 7a. Variation of RDC (case 2); V=35km/hr.

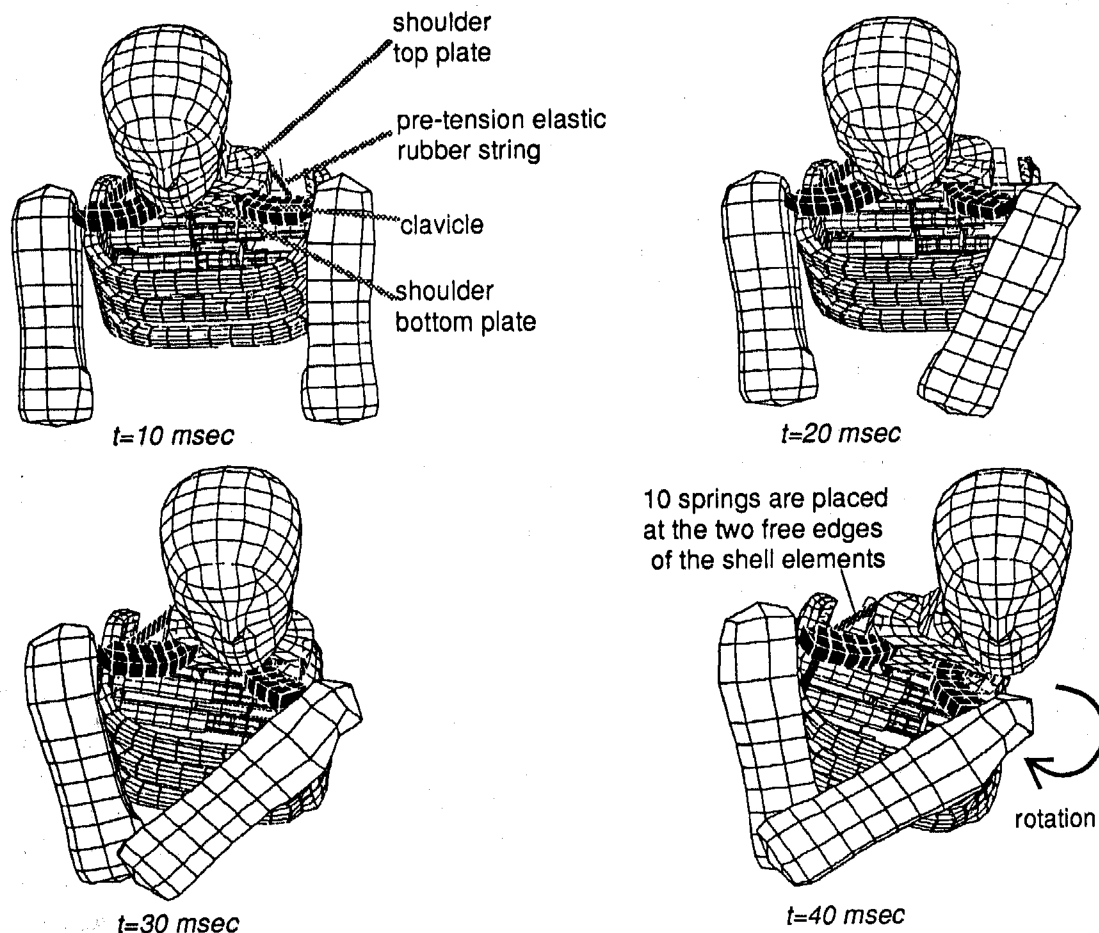


Figure 7b. Rotational Deformation Mode of Shoulder and Arm (case 2); V=35km/hr.

DETAILED VERIFICATION OF FEM MODEL IN DIFFERENT HYGE TEST SIMULATIONS

(a) Mechanism of Side Impact Simulation

An essential requirement for side impact applications is a validated database of the dummy and the barrier models for assessing whether the design meets safety regulations. Simulation results can provide a fundamental understanding of occupant behavior and an evaluation of design variations affecting the occupant responses in side impact crashes, rather than relying on expensive vehicle tests. The prediction of occupant responses in side impact events depends on how well the geometry, the material properties and the mechanism of the dummy can be represented and on how well the responses of the vehicle and the barrier can be predicted as individual items.

Figure 9b represents the typical velocity-time histories of the MDB, the struck vehicle, the impacted door and the dummy's spine in a FMVSS 214 test as illustrated in Figure 9a. The force input from the bumper of the MDB was transmitted via the door hinges to the A and B pillars. The MDB rapidly decelerate due to the reaction force of these pillars. A schematic diagram of the transmission of force from the arm-rest, B pillar and door to the dummy is illustrated in Figure 9c. The complex impact phenomenon can be analyzed with a simple HYGE test as shown in Figure 9d. The abdomen pad has a 50 mm offset with respect to the chest and hip pad to simulate an input pattern similar to that of real crash tests. The door impacts the dummy approximately at a speed of 35 km/hr.

(b) Verification of Dynamic Behavior and Different Injury Response Criteria at a Speed of 35km/hr.

The overall accuracy of the dynamic behavior of the present model is checked by comparing the acceleration of the upper spine (Figure 10a), lower spine (Figure 10b) and pelvis (Figure 10c). On the following page, the injury response criteria of the upper rib displacement, middle rib displacement, abdominal force(APF) and pubic symphysis force(PSPF) are plotted in Figure 10d, 10e, 10f and 10g, respectively. All of them show fairly good correlation which validates the present model. However, the following modifications are necessary to further improve the accuracy of the model.

(a) The shape of the upper rib displacement near the peak. Unlike the simulation results, the experimental result clearly shows two different slope before it reaches the peak.

(b) The shape of the pubic force(PSPF) near the second peak. Unlike the simulation results, the experimental result clearly shows a high peak around 40 msec.

(c) The shape of the experimental abdominal force(APF) is flatter at the top than that of the simulation result.

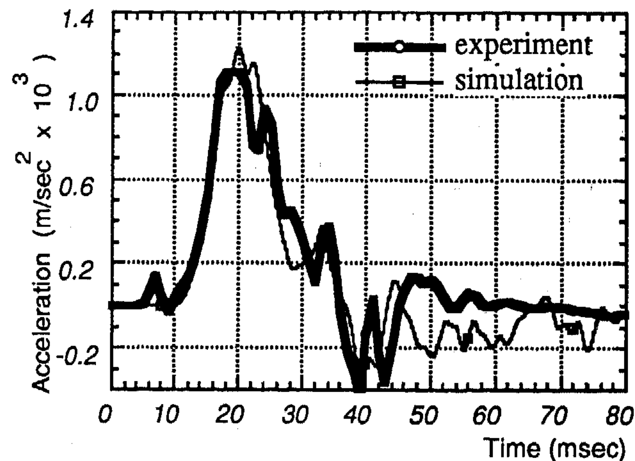


Figure 10a. Comparison of Pelvis Acceleration

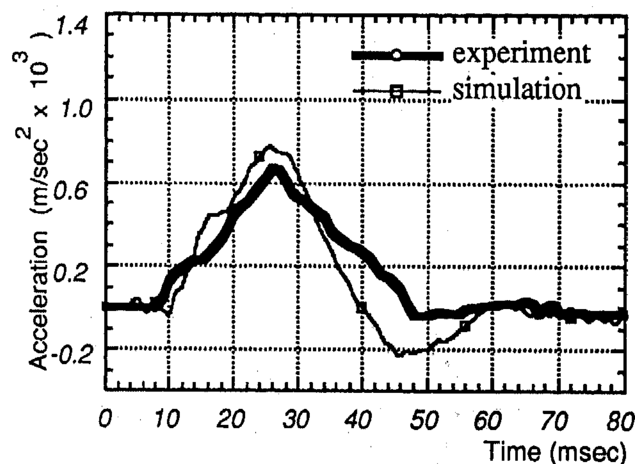


Figure 10b. Comparison of Lower Spine Acceleration

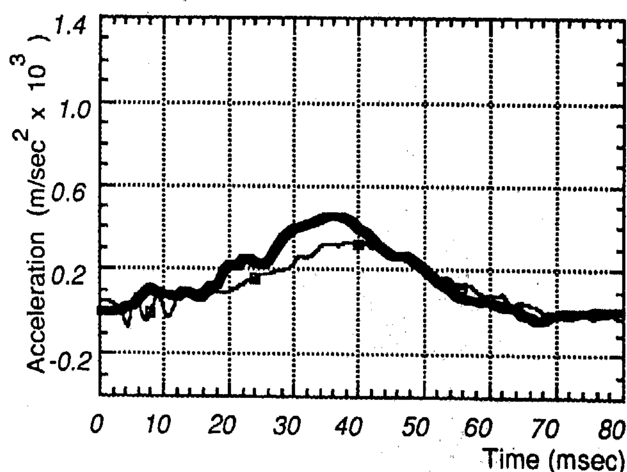


Figure 10c. Comparison of Upper Spine Acceleration

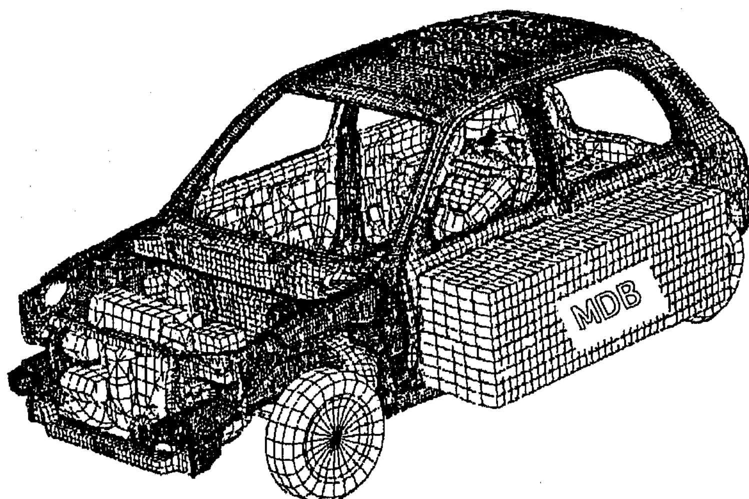


Figure 9a. Full Car Side Impact FEM Model with EUROSID Dummy and Deformable Barrier

Conditions of EEVC

Impact Angle: 90 deg.
Impact Velocity : 50 km/hr
Impact Point: passing through the R-point of the struck car with respect to the center of barrier
Mass of Barrier: 950 kg
Shape: Flat + Bumper
Number of Blocks: 6
Material Type: aluminum honeycomb (20, 50 and 60 psi)
Ground Clearance: 300 mm
Height: 550 mm
Width: 1500 mm
Thickness: 500 mm

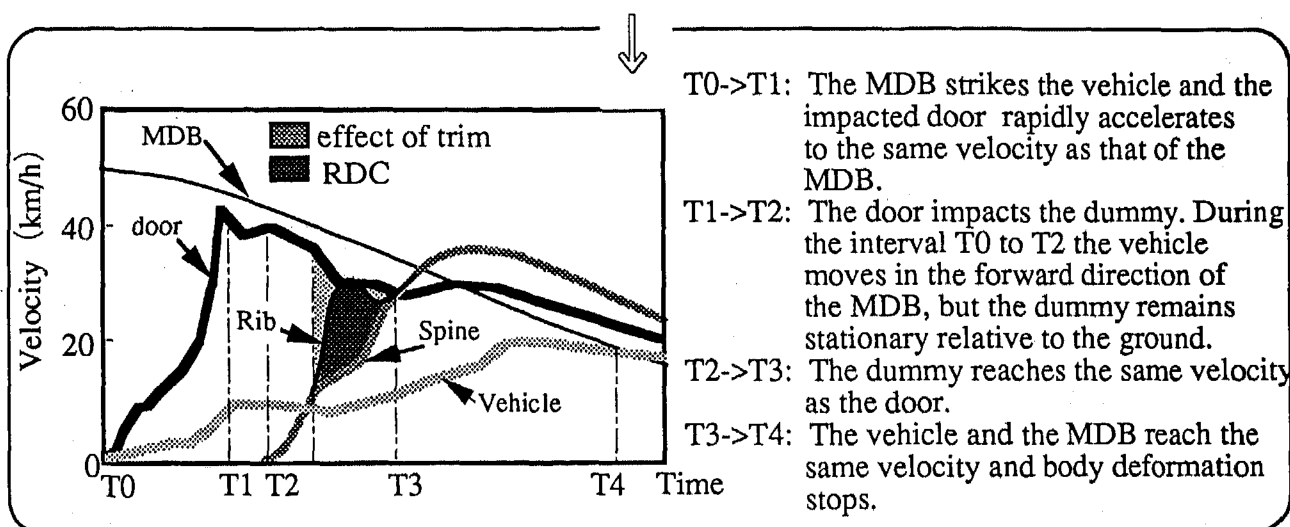


Figure 9b. Velocity-Profile of Dummy and Deformable Barrier

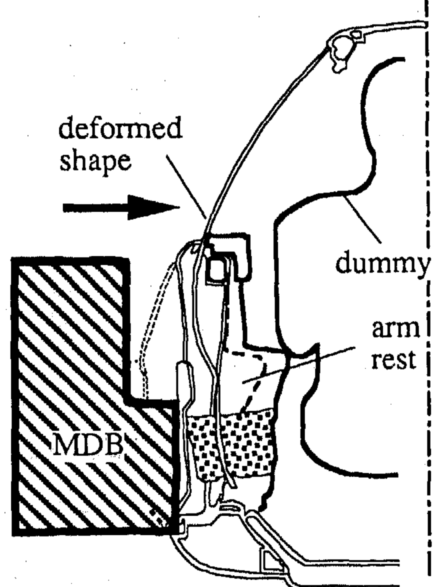


Figure 9c. Schematic Diagram of Deformed Shape of Door at the Time of Contact with the Dummy

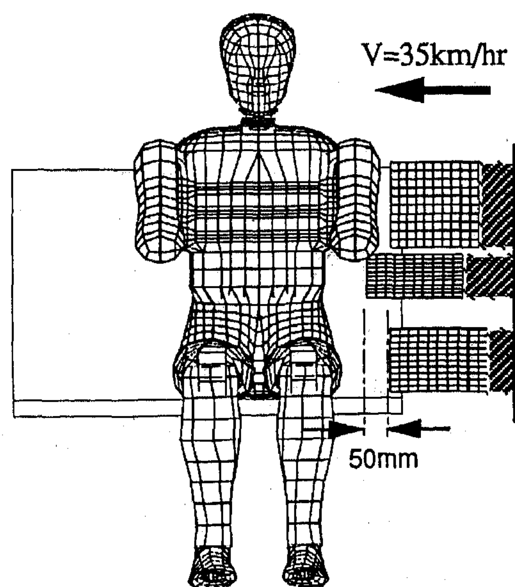


Figure 9d. Approximation of Side Crash Phenomena with Simplified HYGE Test.

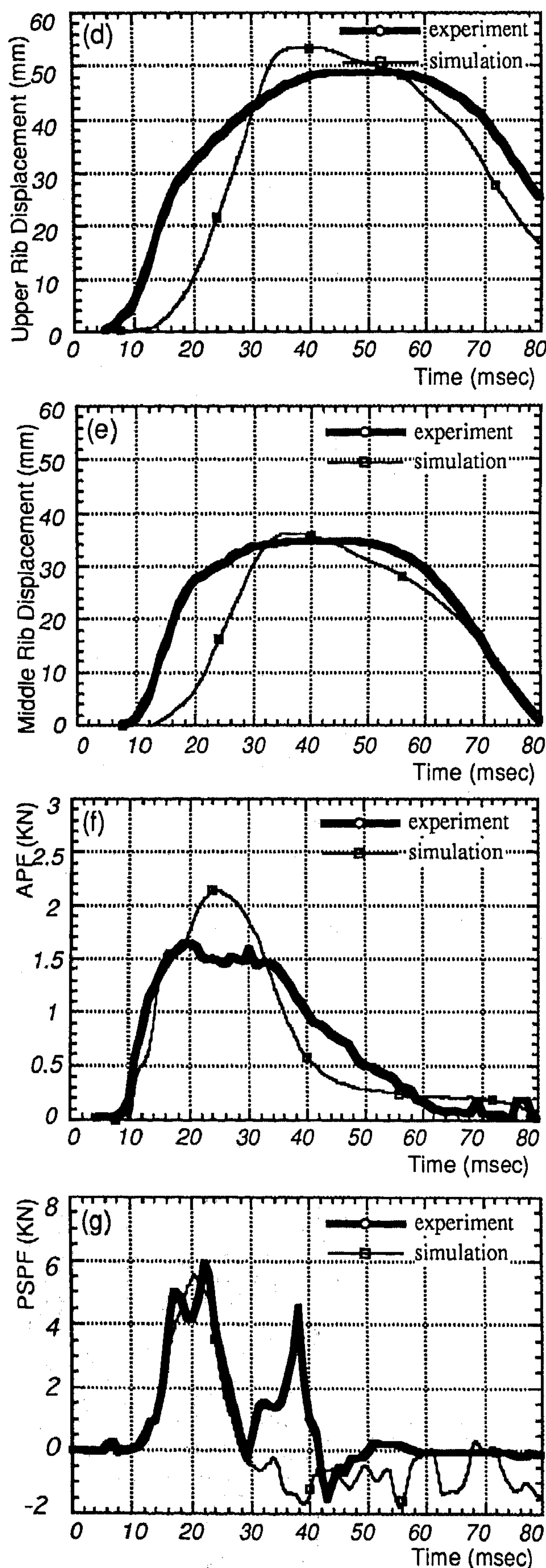


Figure 10. Comparison of (d) Upper Rib Displacement, (e) Middle Rib Displacement, (f) Abdominal Force (APF) and (g) Pubic Symphysis Peak Force(PSPF) (case 4, V=35 km/hr)

(c) Variation of Injury Criteria at a Speed of 35km/hr for Different Input Load Patterns

A series of nine different sled tests was performed by varying the input load pattern (Figure 12) at the abdomen and hip to investigate the variation of the injury criteria such as RDC, PSPF and APF, simulating a broad spectrum of possible input load patterns on the occupant in actual side crashes. The results are plotted in Figure 11.

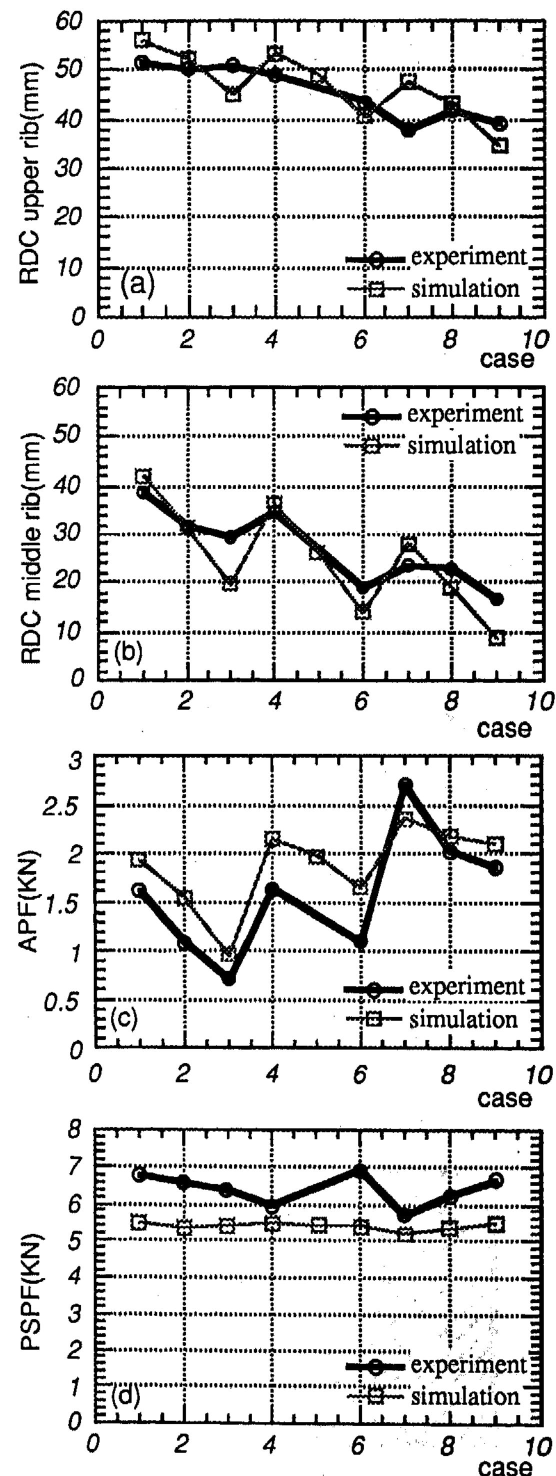


Figure 11. (a) Variation of Upper Rib Displacement, (b) Middle Rib Displacement, (c) Abdominal Force (APF) and (d) Pubic Symphysis Peak Force(PSPF) (V=35 km/hr)

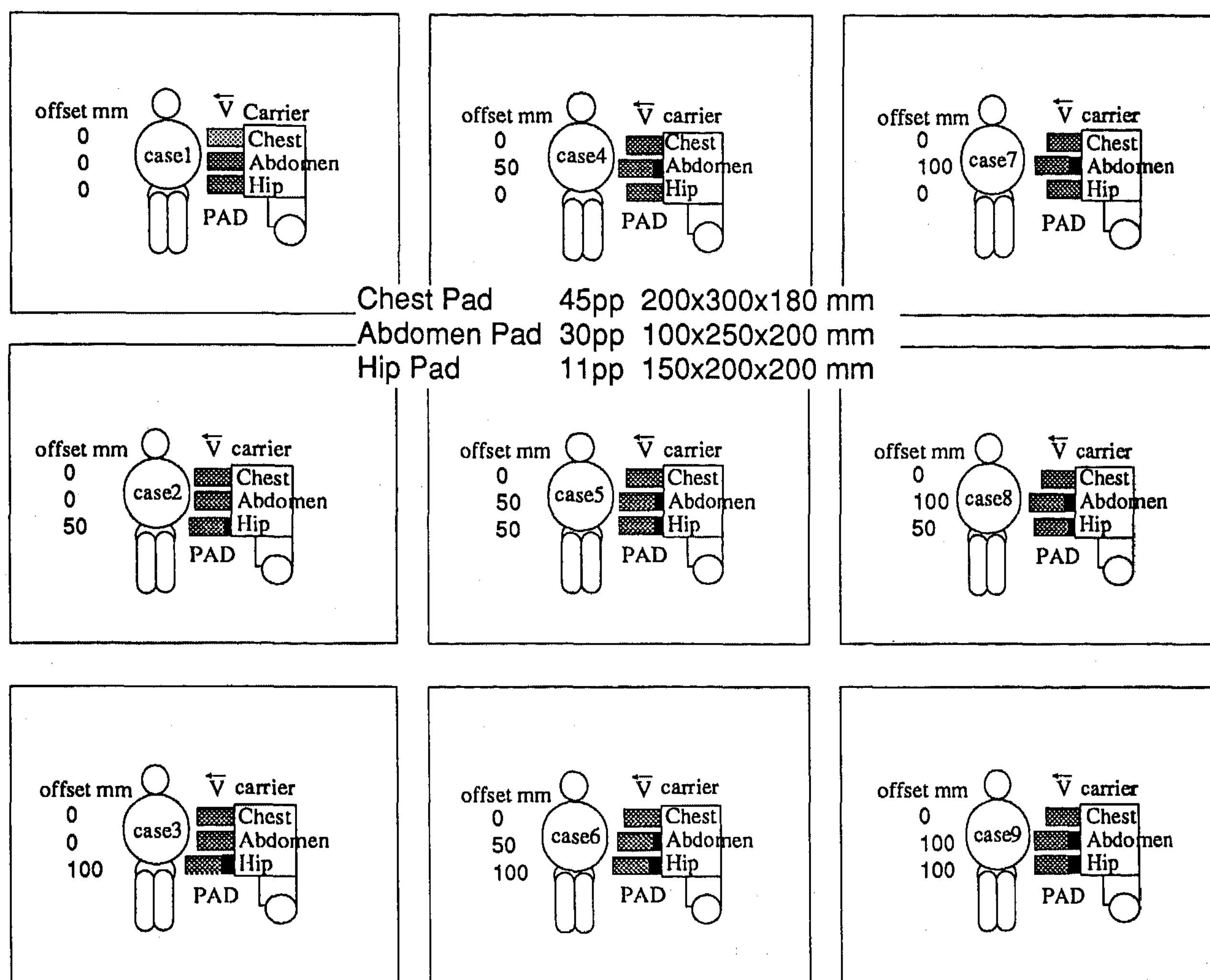


Figure12. Different Input Load Patterns used in a Series of HYGE Tests

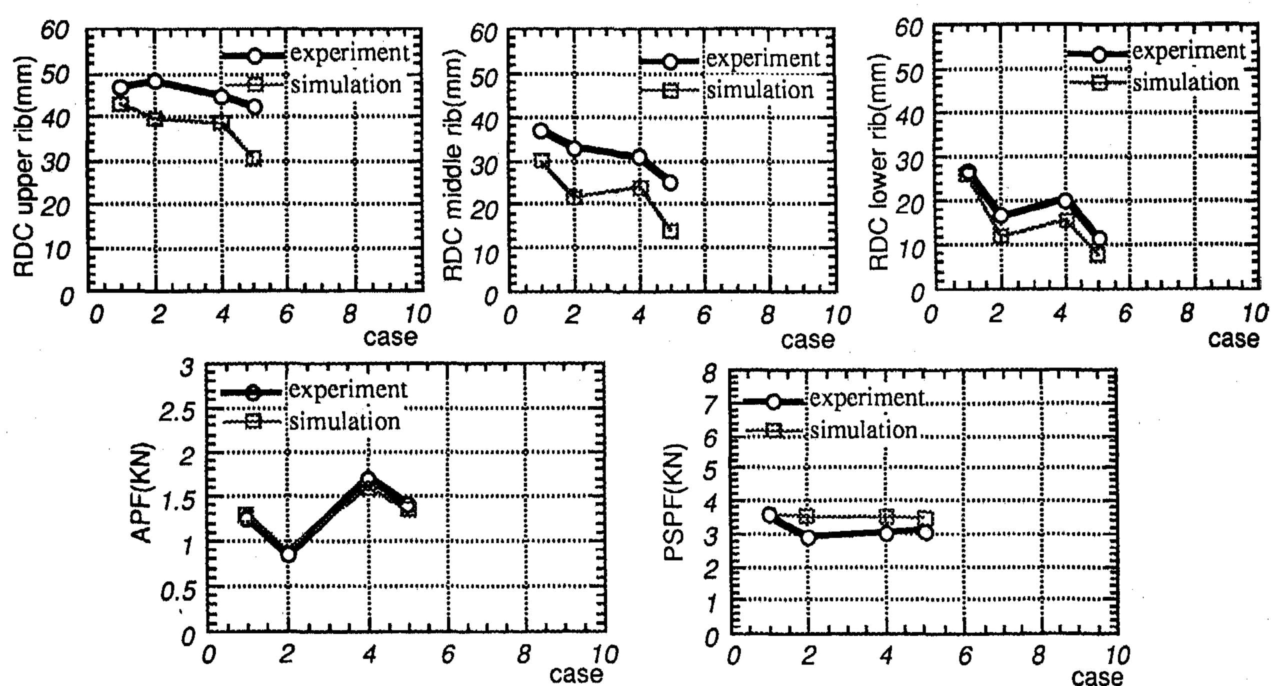


Figure13. (a) Variation of Upper Rib Displacement, (b) Middle Rib Displacement, (c) Abdominal Force(APF) and (d) Pubic Symphysis Peak Force(PSPF) (V=25 km/hr)

(d) Verification at a Lower Speed of 25km/hr

To further check the accuracy of the present model at a much lower speed (25 km/hr) than actual crash tests, four different cases (case 1, case 2 case 4 and case 5) were chosen. Case 2 has a 50 mm offset at the hip with respect to case 1 which has no offset. On the other hand, case 4 has an offset of 50 mm at the abdomen relative to case1 and case 5 is chosen as a combination of cases 2 and 4. Figure 12 shows a comparison of the tests and experimental results for the upper, middle and lower rib displacements, the abdominal force and the pubic symphysis pelvis force. The results are very similar to those at an impact speed of 35 km/hr.

CONCLUSION

Based on the results of the foregoing simulations and experimental verifications, the following conclusions can be drawn.

- Comparing the peak values of thorax damper displacement for different velocities and types of impact, a satisfactory level of correlation can be achieved with the present FEM model for studying the dynamic behavior of the EUROSID dummy.

- The above results indicate that proper modifications of material modeling lead to better representation of bottoming of the arm pad, rib deformation mechanism and lumber spine bending characteristics and so forth, all of which are very important in analyzing the distribution of the load path from the ribs to the spine.

- Single-input rigid hip impacts were studied to investigate the effect of the legs. The results clearly show the effect of leg inertia and the modeling of joints on pubic force PSPF.

- The sensitivity of the pubic force(PSPF) is comparatively low in comparison to the abdominal force(APF) with respect to the timing of the input load as discussed in the present analysis. However, rib displacements can be considerably reduced with a comparatively faster input at the hip. It can be well explained by analyzing the mechanism of rib displacement which is nothing but the area under the curve of the velocity profiles corresponding ribs and the spine as illustrated in Figure 9b. A faster input at the hip will accelerate the spine box with little rib displacement. This initial velocity will prevent a strong impact at the thorax.

- In spite of the reduction in rib displacement criteria(RDC)due to a faster input-load at the abdomen, the high sensitivity of the abdominal force with respect to the input-load timing at the abdomen negates the benefit of having a large offset at the abdomen.

- Good correlation with the experimental results at a speed of 35km/hr indicates that the present FEM model will allow the CAE engineers to evaluate doors, reinforcements, foam padding and B-pillar sections in side crash tests with sufficient accuracy.

- Further improvement is necessary in the pelvis model to predict the variation of PSPF values with higher accuracy. It may also be worthwhile to study the sensitivity of the location of the abdominal and pelvis input loads on different injury criteria.

DISCUSSION

This paper has focused on the basic dynamic behavior of the EUROSID dummy and the corresponding modifications incorporated in the FEM model. Only a small number of multi-input padded sled-impact dynamic responses were tested and verified. These were studied to investigate the effect of the timing of the door impact on the injury level of the occupant. Further study is needed to improve the cross-correlation of the hip-thorax input-output relation and vice-versa. A series of multiple sled tests with simultaneous impacts at the chest, abdomen and hip is also under way with the help of the present FEM model to determine the best feasible input load pattern for achieving the optimum dynamic response.

REFERENCES

[1] RADIOSS User's Manual, Version 2.3, MECALOG, France.

[2] Eurosid Side User's Manual, TNO, July 1991.

[3] Midoun, D., Abramoski, E., Rao, M., and Kalidindi, R., *Development of a Finite Element Based Model of the Side Impact Dummy*, SAE International Congress and Exposition, SAE Paper No. 930444, March 1993.

[4] Detlev Maurer, Stephan Kohlhoff and Paul A. DuBois, *Advances in Side Impact Simulation Procedures*, 14th International Technical Conference on Enhanced Safety of Vehicles, Paper No. 94 S6 W 24, May 1994.

ACKNOWLEDGEMENT

The authors would like to thank Ishikawa, Honda and Satoh of the Crash Safety Laboratory of Nissan Technical Center at Atsugi for their cooperation in carrying out different experiments. They would also like to thank Hideo Tai of Mecalog Japan for his valuable suggestions.

971042

Development of an Advanced Finite Element Model Database of the Hybrid III Crash Test Dummy Family

Steve Moss and York Huang
First Technology Safety System, Inc.

Tim Keer and Bhavik Shah
Ove Arup & Partners International, Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The paper describes the development process of a detailed finite element model database of the Hybrid III crash test dummy family using the manufacturer's product data and test facilities. Specific reference is made to the Hybrid III 50th dummy. The models are validated at material, component, assembly and sled test application levels, and the modeling methodology is described. Each component of the dummy model is discussed together with the specific component impact tests used to validate the model. The specially designed sled tests used to generate highly controlled validation data are detailed.

INTRODUCTION

Crash test simulation has rapidly evolved from the research and academic fields into an everyday design and development tool at automobile manufacturers worldwide. Rigid body models of the crash test dummies are generally accepted and widely used, but the rigid body technique is limited when modeling the vehicle and restraint systems. The rigid body codes are coupled with finite element codes so as to use the rigid body dummy model with a detailed finite element model of the vehicle and restraint system.

Various finite element models of the crash test dummies have been developed by industry and academia to different levels of validation and application and they have not yet been widely accepted. The model database described here is developed from the manufacturer's product data and uses many specially designed impact test conditions to validate the models at the material, component, sub-assembly and full assembly levels.

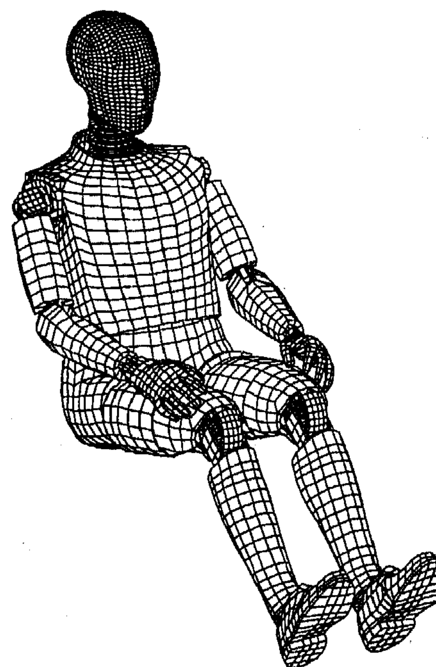


Fig. 1 Assembled Model

MODELING METHODOLOGY

The primary goal was to develop a model that will perform accurately and robustly in a variety of real application test conditions without the need to modify the model for each specific application. To achieve this objective, the models are developed and validated separately at the material, component, sub-assembly and full assembly levels.

Geometry for the models are taken from the manufacturer's engineering drawings for the machined parts. Three dimensional scans were made of the aluminum and brass castings, such as the skull, pelvis

bone and femur hip joints. The vinyl and foam parts were scanned from the manufacturer's production molds or master patterns. It was not assumed that each part was symmetrical, and left and right sides were scanned where there is asymmetry in the physical dummy. Model masses were checked against the production parts.

The component parts are all modeled exactly and without ambiguity. The non-deformable components, typically the dummy skeleton, are rigid bodies but are exact physical representations of the actual parts. The mass distribution is therefore also exact and avoids the need to rely on measured moments of inertia, which can be a source of considerable error.

The robustness of the model is also checked at the component and sub-assembly level. Several analytical test conditions were devised to check the interaction of the component models with special attention paid to the contact surfaces and hour glass energies.

TESTING METHODOLOGY

Individual components and sub-assemblies are validated to many specially designed impact tests as well as the standard calibration tests. Over 600 tests have already been performed for the Hybrid III adult family dummies to date. The full assembly dummy is then validated to pendulum calibration tests and to controlled sled tests.

The selection of the dummy components for the validation testing is carefully controlled. The dummies are certified to perform according to standard calibration corridors which are based on human impact response corridors. These corridors are essentially tolerance bands either side of a nominal performance curve. Manufacturing variations from dummy to dummy can cause individual dummies to perform either higher or lower then the nominal curve. All components selected for the validation tests are brand new and are taken from a dummy or dummy sub-assembly that meets the standard calibration specification as close to the nominal curve as possible.

Wherever practical, material properties are derived from material tests of samples made in special molds or sectioned from production parts. The manufacturer's production equipment was used to make the material samples. All component tests were conducted in the manufacturer's laboratories under controlled temperature and humidity conditions. Material tests were conducted at outside test laboratories.

HEAD

A highly accurate solid model of the head assembly was built using CAD data scanned from the production molds and finished parts. This model was used to calculate the moments of inertia which are within 2% of the nominal specifications for the skull assembly.

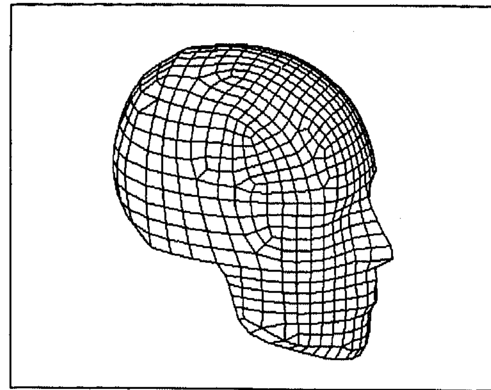


Fig. 1 Head Model

The regulation specification has a very wide tolerance band because of the inherent inaccuracy of physically measuring moments of inertia.

The actual head model has shell elements as part of a rigid body to represent the external surface of the skull. Additional nodal masses are distributed about the rigid body to accurately represent the correct inertia.

The deformable skin is modeled as visco-elastic solid elements. The material properties are derived from the basic vinyl material. Special molds were machined to produce flat vinyl samples at the same thickness as the head skin, and were tested statically and dynamically, and models of these tests built and correlated to the test.

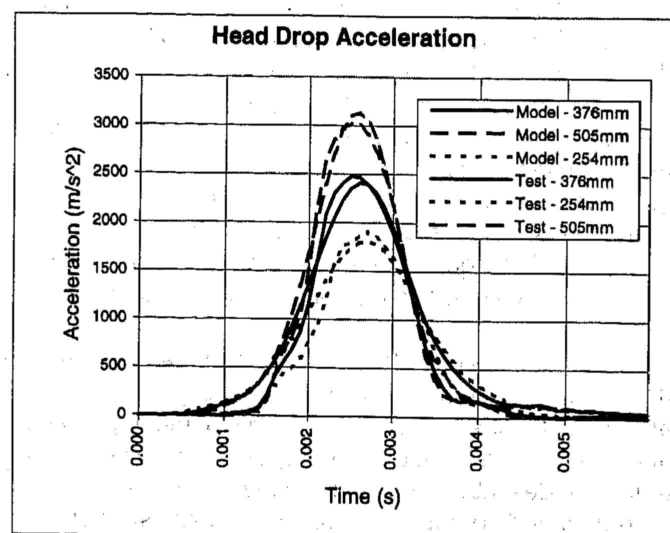


Fig. 2 Head Drop Correlation

Drop Height (mm)	HIC Test	HIC Analysis
254mm	**420**	**412**
376mm	**756**	**784**
505mm	**1282**	**1276**

Table 1 Head Drop Test HIC

The complete head assembly model is validated to three different head drop speeds, to drop impacts to irregular surfaces, to other pieces of vinyl and to a deformable sheet metal section. Peak accelerations and HIC values are within 5% of the test values, as shown in Table 1. The correlation to the three drop impacts is shown in Fig. 2.

NECK

The neck assembly is divided into two mechanisms. The occipital condyle is represented by a lateral pin pivot with anterior and posterior rubber stops controlling the nodding motion of the head. A component test was designed to isolate this joint and measure the moment/rotation stiffness at various rates. This joint was modeled in LS-Dyna as a revolute joint with a non-linear rotational spring and damper per the test measurements.

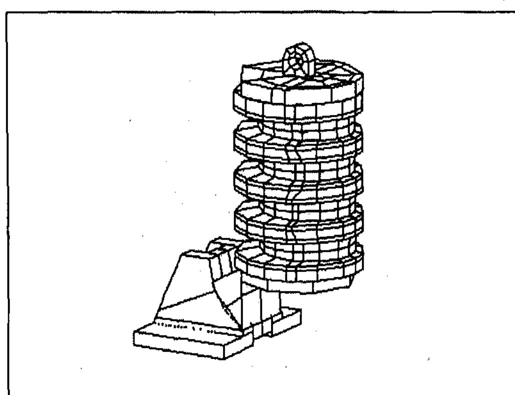


Fig. 3 Neck Model

The neck is composed of four butyl rubber segments sandwiched between five aluminum discs with a steel cable running through the center. The entire assembly is molded in one operation and the bonding between rubber and aluminum has to meet a specification. The rubber segments discs have a partial-depth horizontal slit on the anterior side. This results in a less stiff response in neck extension (head bent back) than in flexion (head bent forward) to match the human neck response. Each slit has a different depth to give a prescribed deformed shape.

A new certified neck assembly was rough sectioned and the samples of the rubber were milled in a fixture cooled by liquid nitrogen. The process made the rubber hard enough to produce precise, high finish samples for material testing. These samples were put through tension, compression and oscillatory shear tests to define the base material properties. The samples had to be cut from actual parts because the high temperature and pressure processing of rubber components determines the final material properties.

The geometry for the neck is taken from the manufacturer's drawings. The aluminum disks are modeled as rigid solid elements and the rubber segments are modeled as visco-elastic solid elements. Eight-point integration was necessary to avoid hour-glassing at the extreme deformations. The steel cable is represented by elastic truss elements with contact defined to the aluminum disks. A single surface contact surface was used for each neck slit. In order to increase the stiffness of these slit contact surfaces without causing time-step problems, the density of the neck rubber adjacent to the slits is increased and the densities of the remaining rubber and the aluminum disks reduced accordingly. The mass redistribution is not believed to affect the response.

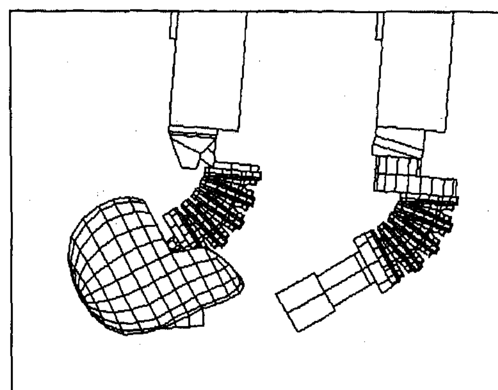


Fig. 4 Neck Pendulum Tests

To test the neck without the nodding blocks, a rigidly mounted steel dead weight was mounted to the neck to replace the head. This assembly was then tested at different speeds on the neck pendulum fixture. The test was exactly modeled (see Fig 4) and the test deceleration pulse used as input to the model. A six axis load cell measured moment and shear load, and rotary potentiometers recorded the angular deflection.

Final validation was completed with the complete head, nodding joint and neck assembly on the neck pendulum test fixture. The standard test calibration impact velocities are 7.06ms^{-1} for flexion and 6.07ms^{-1} for the extension test. Both tests were repeated at a lower velocity of 4.34ms^{-1}. The deformation of the neck during the standard extension test is shown in the left side of Fig. 4. Note the opening of the neck slits. The neck rotation vs. time correlation results are shown in Fig 5. "D-plane rotation" is defined as the rotation of the base of the head relative to the pendulum.

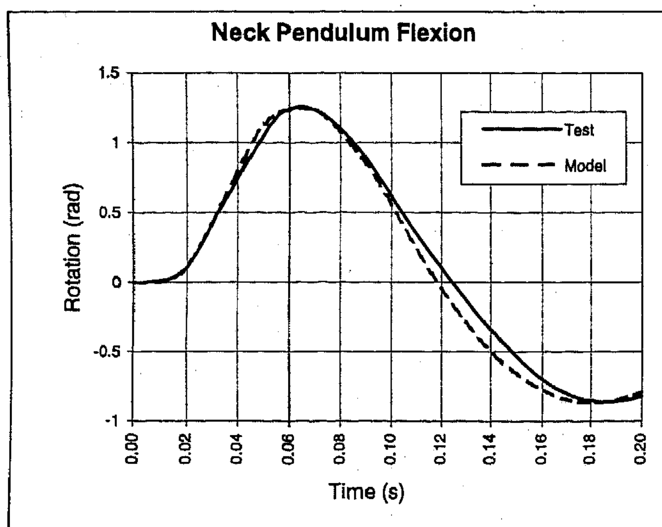


Fig 5 Neck Flexion Correlation

THORAX ASSEMBLY

The thorax is the most complicated of the Hybrid III components. A ballasted steel spine box acts as the backbone . Six spring steel ribs are attached by screws to its rear with additional leaf springs to prevent permanent deformation of the ribs. A high density damping material is bonded to the inside of each rib. The thickness of this material is modified to control the dynamic performance of the thorax assembly. The sternum assembly which joins the open end of the ribs, is a sandwich of Delrin (a brand name nylon), a hard polyurethane and aluminum.

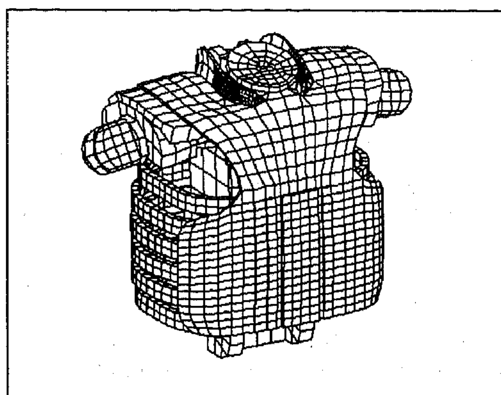


Fig. 6 Thorax Assembly

Each shoulder is represented by three aluminum castings, the clavicle link, clavicle and shoulder yoke, which are connected by joints that allow a few degrees of angular motion. These shoulder components have contoured upper surfaces for interaction with a vehicle's seat belts. Surrounding the thorax assembly is a jacket made from a vinyl-skinned polyurethane foam. An Ensolite pad is bonded to the inside of the front of the jacket. The hard polyurethane part of the sternum, called the bib, extends upwards and splits into three. The center of the bib is clamped between the spine box neck bracket and the neck's lowest aluminum disk. Ensolite pads are bonded to the side pieces of the bib and sit over the clavicles.

Chest deflection is the relative displacement between the sternum plate and the spine box as measured by an aluminum rod attached to a rotary potentiometer pinned to the base of the spine box. The other end of the aluminum rod slides in a groove in the aluminum sternum plate. As the rib set is compressed, the rod rotates and its outer end slides up the sternum. Chest deflection is calculated by applying a transfer function to the output of the rotary potentiometer.

THORAX MODELING

The spine box is modeled in detail as rigid solid elements including the internal lead ballast. The three accelerometers used to calculate the resultant chest acceleration are individually represented as the standard LS-Dyna accelerometer feature. Beams are added through the active axis of each accelerometer to aid visualization.

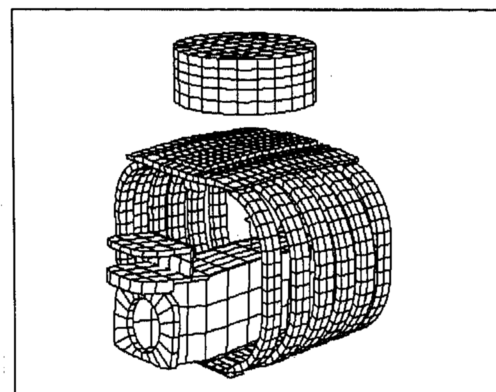


Fig. 7 Thorax Assembly Drop Test

The ribs are modeled as elastic shell elements, two rows deep, with a single layer of visco-elastic solid elements using single point integration for the damping material. The Delrin and aluminum plates of the sternum are represented by rigid solid elements and the bib as elastic thick shell elements. The jacket and ensolite pads are modeled as low density foam solid elements with 8-point integration. Fig. shows the thorax assembly model without the jacket.

The clavicle, clavicle link and shoulder yoke are modeled as separate solid rigid bodies. The clavicle link is pinned to the spine box, with resistance provided by a rubber shoulder bumper. The clavicle is pinned to the clavicle link, with resistance provided by a Delrin pad sitting between the two. The shoulder yoke is attached to the clavicle by a joint with Delrin bushings and cushioned

rubber bump stops. All these joints were modeled in LS-DYNA as generalized stiffness joints. Moment/rotation characteristics and stop angles were taken from test data.

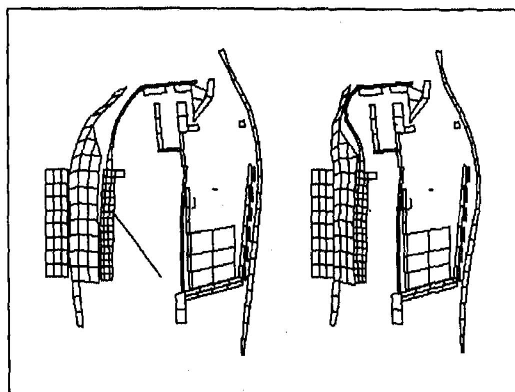


Fig. 8 Section Through Thorax Model

The chest displacement transducer is modeled rigorously. The aluminum rod is represented as a rigid beam with its lower end attached by a revolute joint to the adapter assembly which is similarly attached to the base of the spine box. Its upper end was connected to the sternum rigid body via a cylindrical/spherical joint combination that allows it to slide up and down the sternum. A zero-stiffness rotational spring at the base of the rod measures the rotation.

THORAX VALIDATION

Base material properties were measured from standard tension, compression and oscillatory shear tests of all the plastic materials. The rib steel is assumed to remain elastic.

A drop tower test was devised to perform individual rib and rib assembly impact tests. The ribs are clamped vertically on a bed plate with a load cell beneath it. A cable-guided drop head with accelerometer and velocity trap drops onto the ribs from varying heights generating impact velocities up to 10m/s. A vinyl pad was mounted on the drop head to avoid metal to metal impact. A linear potentiometer measures rib deflection.

Each rib of the six rib assembly was tested with and without the damping material. The ribs came from a new certified set of ribs. Models of each test condition were built to ensure the response of the damping material was correctly characterized at different velocities.

Drop tests were also performed on a rib-cage sub-assembly comprising the spine box, ribs and sternum. The model of this test is shown in Fig 7. An Ensolite pad was mounted on the drop head face to prevent metal to metal contact to the sternum screw heads. Correlation for this test is shown in Fig. 9. Note the accurate prediction of the residual chest compression.

To test the robustness of the material stack up between the chest jacket and rib assembly, an analysis of a drop test to the full thorax with jacket was performed to ensure that the many components in the thorax would interact properly. A section through this model is shown in Fig. 8. The left side shows the initial, undeformed geometry of the thorax and drop head. The right side shows the deformed geometry at the time of maximum chest compression.

The interaction with LS-DYNA one-dimensional seat-belt elements was also evaluated. The full thorax model with an initial velocity of 6ms^{-1} was decelerated by a shoulder belt with fixed end points. The maximum belt load during the analysis was 2000N.

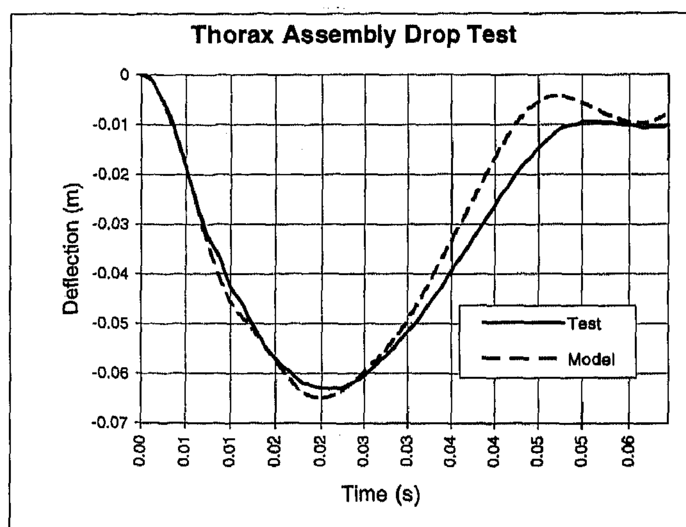


Fig. 9 Thorax Assembly Drop Test

ABDOMEN

The Hybrid III abdomen is a single C-shaped component molded as a one piece unit. The part is a polyurethane foam with a vinyl skin and locates in a cavity above the pelvis assembly.

The geometry for the abdomen model was scanned from the production mold and it is modeled with low density foam solid elements with single point integration (see Fig. 10). Drop tests at 4.50 ms^{-1} and 6.76 ms^{-1} were used to validate the model. Correlation for the 6.76 ms^{-1} test is shown in Fig. 11.

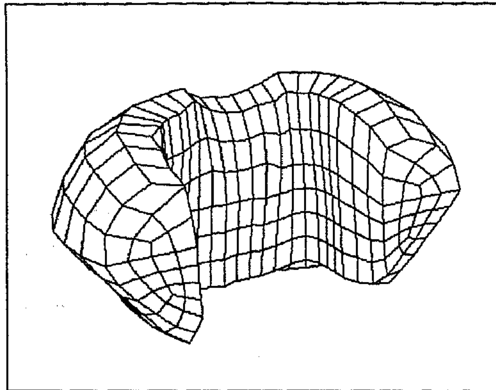


Fig. 10 Abdomen Model

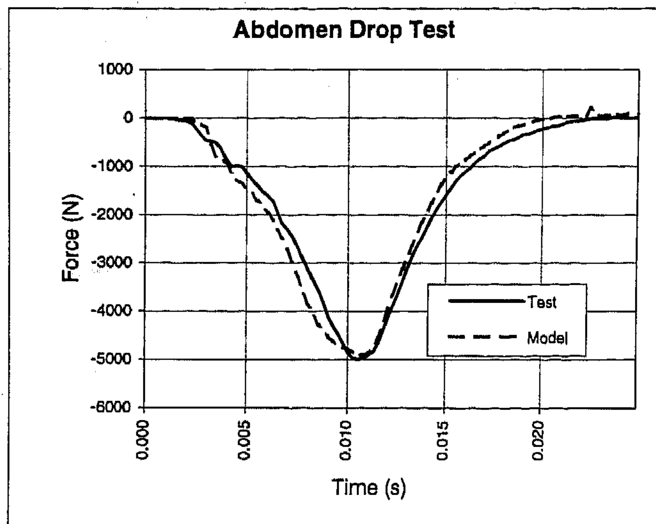


Fig. 11 Abdomen Drop Test Correlation

LUMBAR SPINE

The lumbar spine is a curved polyacrylate elastomer member, molded between steel end plates. Two internal steel wire-wound cables provide lateral stiffness and limit extension of the spine. The geometry for the lumbar spine model was taken from the manufacturer's drawings. The rubber is modeled as visco-elastic solid elements with single point integration. The fundamental material properties and short- and long-term shear moduli are derived from oscillatory shear tests performed on a sample of the polyacrylate elastomer material cut from a new lumbar spine. The steel cables are modeled as a series of elastic truss elements spaced from the rubber by rings of rigid shell element. Spherical contact entities were evaluated for these cables but the CPU costs were prohibitive.

The lumbar spine is a key load path between the upper and lower body in a vehicle impact, especially when the dummy is only restrained by an air bag and knee bolsters. If there is a time delay between the restraint of the upper and lower torsos, then a significant velocity difference can occur between these two bodies, until the lumbar spine corrects the difference. Shear loads in the lumbar spine can exceed 5000N which is a major component of the decelerative forces on the spine box and thus the measured chest accelerations.

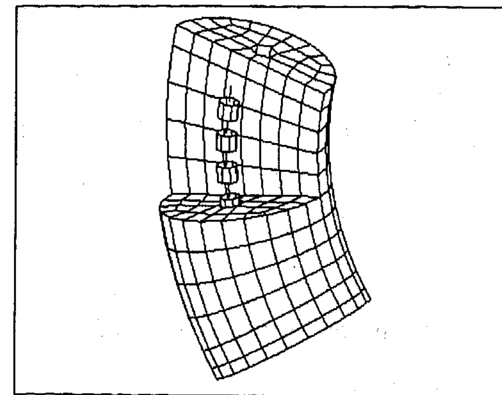


Fig. 12 Lumbar Spine

There is no specific calibration test for the lumbar spine as a component, but there is a static bending test on the fully assembled dummy. Because of the lumbar spine's important role in a crash test, this static test was ignored for the model and a dynamic component test was devised.

Special fixtures were built to attach the lumbar spine to the neck pendulum test fixture with a six axis load cell. A rigid steel mass was attached to the other end of the spine. Extension and flexion tests were performed at 4.34, 5.97 and 6.99m/s using an aluminum honeycomb block to provide the deceleration pulse. Fig. 13 shows the deformed and undeformed shapes of the lumbar spine undergoing the extension test, with the pendulum swinging from right to left. The correlation of the spine rotation with time is shown in Fig. 14 for the 5.97m/s flexion test.

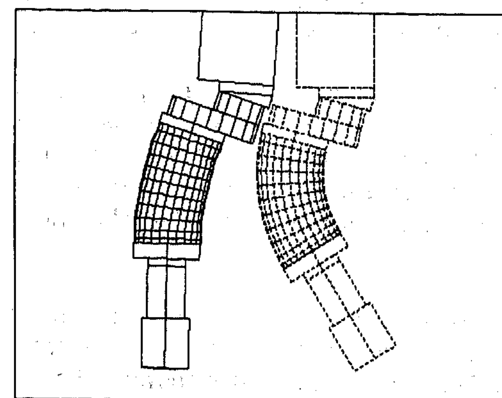


Fig. 13 Lumbar Spine Flexion Test

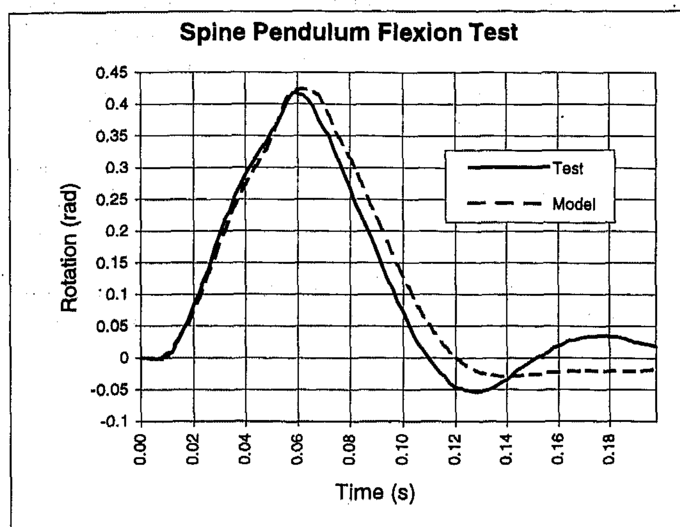


Fig. 14 Lumbar Spine Pendulum Flexion Test

PELVIS

The hip assembly of the Hybrid III is a vinyl-skinned polyurethane foam molded around a machined and ballasted aluminum casting representing the pelvic bone The pelvis is molded in a permanently seated position. The upper femur joints are brass castings with a ball joint that clamps into ball sockets machined into the pelvis bone casting. The pelvis casting is a regulated design and was originally cast from an actual human pelvis bone. Accordingly, the pelvis bone is not symmetrical.

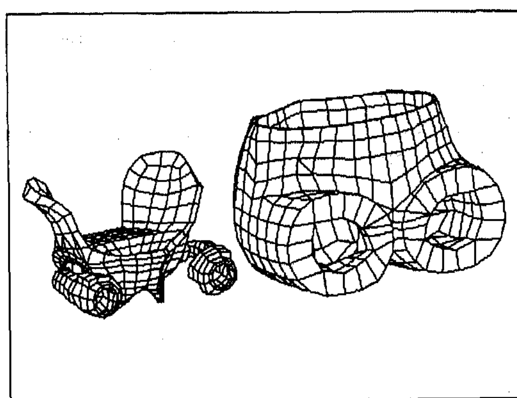


Fig. 15 Pelvis Model

In certain vehicle tests, the femur to hip joint can rotate up to the limits of its range of motion and the ball joint bottoms out, causing excessive loads in the lumbar spine and chest acceleration. To address this, NHTSA has recently proposed an additional calibration test that sets moment/rotation corridors for the hip to femur joint. This requires a design modification to the femur attachment and the addition of a reinforced polyurethane pad to achieve the necessary moment/rotation requirements. This new design has been incorporated into the model.

The pelvis model geometry was scanned from an actual pelvis casting with the asymmetry retained. The pelvic bone surface is modeled as rigid shells and the correct mass moments of inertia applied. The foam and vinyl skin is modeled with low density foam solid elements with eight point integration. The femur attachment member is made up of rigid solid elements, with the urethane pad modeled as elastic solid elements. Fig. 15 shows the pelvic bone and femur attachment members to the left and the pelvic skin to the right. The deformable pelvis elements are matched with the surface of the pelvis bone and the mating surface nodes are incorporated into the rigid body of the pelvis bone.

To assess the robustness of this model, a belt deceleration analysis was performed, similar to the analysis performed for the thorax. The maximum belt load during the analysis was 4000N and several modifications were made to the model to improve the robustness.

KNEE

The knee assembly is a machined steel part containing the knee pivot joint, with a steel clevis weldment attaching to the lower leg. On each side of the clevis pivot and the knee is a knee slider joint with a rubber block bonded to both sides of the slide. A vinyl-skinned urethane foam molding fits over the knee assembly with an additional rubber pad between the knee cap portion and the vinyl skin.

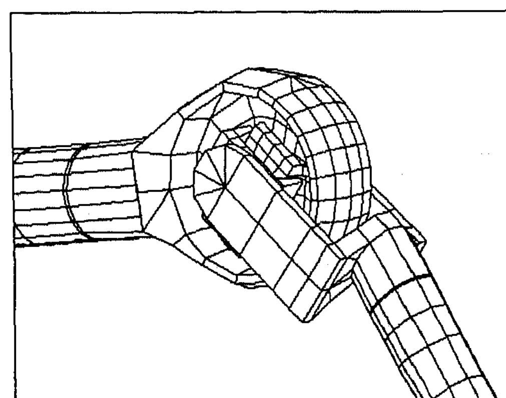


Fig. 16 Knee Model

The knee flesh geometry was obtained by scanning production molds. The remaining geometry was taken from engineering drawings. The LS-Dyna model of the knee is shown in Fig. 16. To the left is the femur (upper leg) skeleton and to the right is the tibia (lower leg) skeleton. The knee cap, friction block and slider are modeled using rigid solid elements and the knee pad is a single layer of visco-elastic solid elements

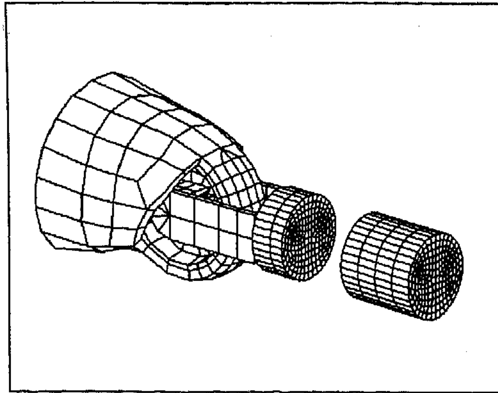


Fig. 17 Knee Slider Test

with eight point integration. The knee slider is modeled using a non-linear spring and a linear damper in parallel.

The two knee calibration tests are used to validate the knee model. The upper leg is rigidly mounted to a load cell and struck with a pendulum to the knee cap. The pendulum deceleration must fall within set performance corridors. Fig. 18 shows the analysis correlation for this test. The knee slider calibration test measures the response of the sliding joint inside the knee to a pendulum impact to a modified clevis bracket. See Fig. 17.

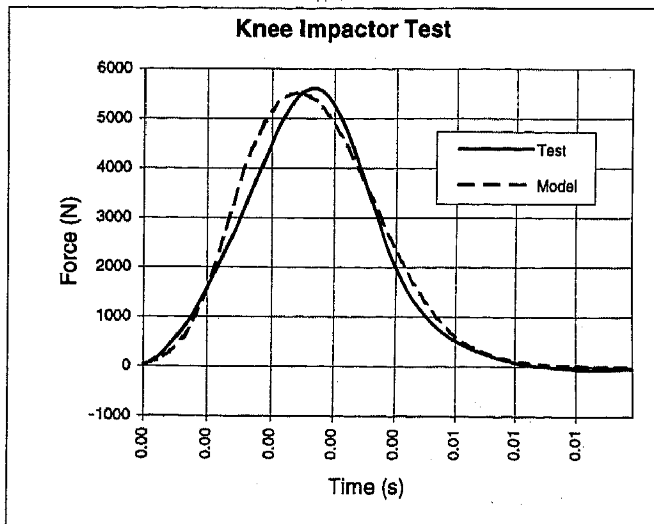


Fig. 18 Knee Impact Correlation

LIMBS

Geometry for the limb skeletons was taken from engineering drawings. The geometry of the external "flesh" was scanned from the master patterns and/or the production molds of the components. The skeletal structures are modeled as solid elements as part of rigid bodies. The deformable vinyl-skinned foam on the arms and legs are modeled as low density foam solid elements. Fig. 19 shows the legs, one with flesh removed to show the underlying skeletal structure.

The elbow, wrist, knee and ankle all include one or more joints to allow twist, rotation or flexion. All these were modeled as generalized stiffness joints. Many of these joints are adjustable in the physical dummy, and are set at "1g" friction - meaning that they should barely support a horizontally extended limb. The correct friction torque or moment was calculated for each joint from moment equilibrium equations based on the mass and center of gravity of the extended limbs. Each joint also has rubber stops to limit the rotation. Quasi-static range of motion tests produced the data for the generalized stiffness joints.

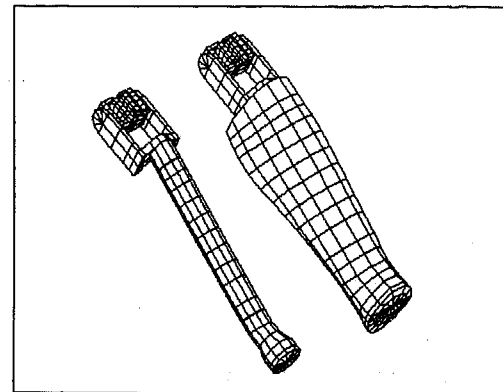


Fig. 19 Lower Leg Model

HANDS

The hands are molded in one piece with solid vinyl and have no skeletal structure except at the joint to the forearm and they play a significant role in a vehicle test. The hands are placed on the steering wheel for test set-up but are typically forced off the wheel as the arms are pushed outwards by the airbag. Also the hands often impact other portions of the vehicle and can cause peak load transfer though the arms to the spine box.

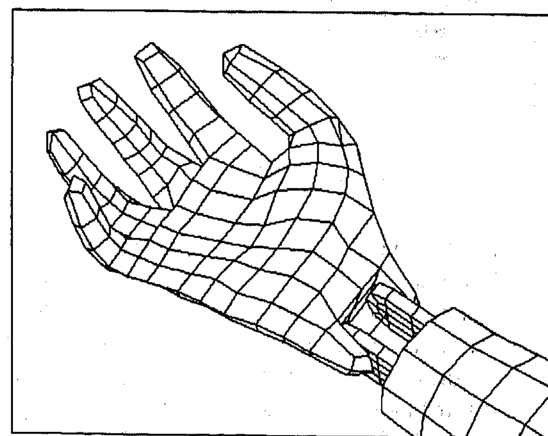


Fig. 20 Hand Model

For these reasons specific detail was put into the hand models which are modeled as visco-elastic solid elements with eight-point integration. There was initial concern that an accurate geometric model of the fingers could lead to such high deformation of the finger elements that the model could not be guaranteed to be robust. The hand model was analytically drop tested onto a rigid surface at 10m/s with added masses to represent the arms. The model. proved to be very robust at several different impact orientations.

ANKLES AND FEET

The model uses a generalized stiffness joint to represent the ankle and allows up to 30 degrees of rotation from the perpendicular position before it hits a hard stop. There is also a 45 degree ankle option with the Hybrid III dummy, and a new NHTSA proposal puts a rubber soft stop into the ankle joint. These will be parts of future model upgrades.

The Hybrid III dummy is always tested with regulation shoes, and these are included by default in the model. The specified regulation size 11EE shoe was scanned and modeled as an integral part of the foot model. The feet are currently modeled as rigid bodies, but a new proposed NHTSA regulation puts a vertical force/defection requirement on the foot which will be part of a model upgrade.

MODEL ASSEMBLY

After the development and validation of the component models, the full dummy model was assembled and subjected to further validation against pendulum and sled tests. Model robustness was a particular focus here - the ability of the model to interact with seat belts, airbags and other vehicle interior components is important.

Most of the components could be assembled easily as they fit together unambiguously. In some instances, however, the physical components undergo deformation during manual assembly, and this deformation needed to be represented in the model. The rear of the pelvis flesh, for example, needs to be stretched rearward by 30mm in order for the lumbar spine to be inserted. Dynamic relaxation analyses were performed for this and other instances (e.g. abdomen insertion). The assembled geometry represents the output from these dynamic relaxation analyses (but note that the model is not pre-stressed).

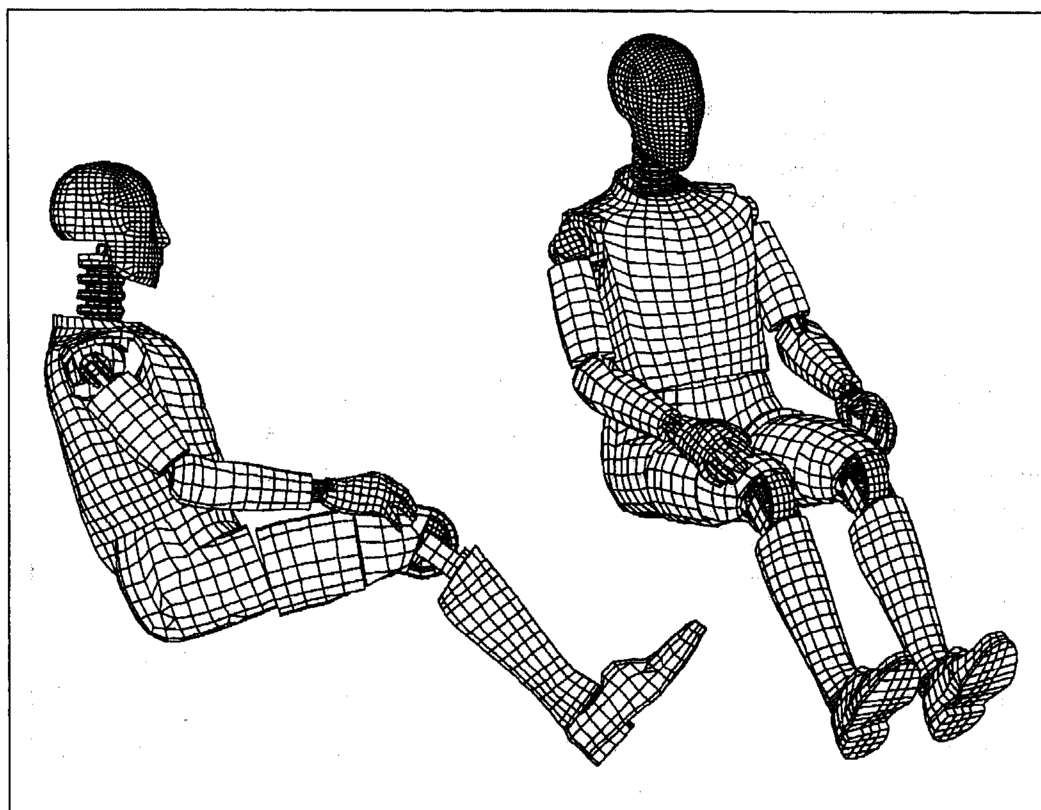


Fig. 21 Full Model Assembly

INSTRUMENTATION

The Hybrid III dummies are available with several instrumentation options in addition to those mandated in FMVSS208. All the optional instrumentation is modeled - See Table 2.

Load cells are modeled in LS-Dyna as stiff springs between nodes on two rigid bodies. Each load cell comprises three translational elastic springs of stiffness $10^5 Nmm^{-1}$ and three rotational elastic springs of stiffness $10^7 Nmm.rad^{-1}$ with orthogonal orientation vectors to allow the extraction of forces in the load cell local axis systems.

Component	Instrumentation	LS-DYNA Feature
Head	Tri-axial accelerometer set	Accelerometer
Neck	Upper - six-axis load cell	6 stiff springs
Neck	Lower - six-axis load cell	6 stiff springs
Thorax	3 accelerometers	3 accelerometers
Thorax	Thoracic spine 5-axis load cell	6 stiff springs
Thorax	Rotary potentiometer	Zero-stiffness rotational spring
Lumbar Spine	3-axis load cell	6 stiff springs
Pelvis	3 accelerometers	3 accelerometers
Upper Leg	Upper femur 6-axis load cell	6 stiff springs
Upper Leg	Femur 6-axis load cell	6 stiff springs
Lower Leg	Instrumented lower leg	12 stiff springs

Table 2 Model Instrumentation

ELEMENT SUMMARY

Table 3 lists the numbers of parts, nodes and elements in each component. Of the parts, 61 are rigid beams whose purpose is to allow the visualization of accelerometers and joint coordinate systems. Null elements are used for some contact definitions and are not included in the totals.

Component	Number of parts	Number of nodes	Number of deformable elements	Number of rigid elements
Head	5	1686	824	830
Neck	13	1749	710	361
Thorax	57	7475	2736	1194
Lumbar Spine	21	1059	464	224
Abdomen	1	825	576	0
Pelvis	15	2013	446	956
Knees & Limbs	85	6708	2206	2048
Total	**197**	**21515**	**7962**	**5613**

Table 3 Element Summary

MASSES

The masses of the individual components are summarized in Table 4. Note that the total is not the sum of each column of figures as each dummy has two sets of limbs. The model masses are within the allowable tolerances.

Each LS-Dyna generalized stiffness joint definition requires two coordinate systems, one for each connected part. To aid visualization during pre- and post-processing, beams were added to the model to show the coordinate system locations and orientations.

Component	**Mass (kg) Actual**	**Mass (kg) Model**
Head	4.5	4.5
Neck	1.5	1.4
Upper Torso	17.2	16.8
Lower Torso	23.0	22.4
Upper Arm	2.0	2.0
Lower Arm	1.7	1.7
Hand	0.6	0.6
Upper Leg	6.0	6.0
Lower Leg	4.3	4.3
Foot	1.4 (no shoe)	1.7 (inc. shoe)
Total	**78.2**	**77.7**

Table 4 Masses

TIME STEP

The component requiring the smallest time step is the thorax at 1.54μs, controlled by a shell element in the steel rib.

FULL DUMMY VALIDATION

The first validation test for the assembled dummy was the standard thorax pendulum test. A rigid pendulum of mass 23.4kg and velocity 6.7ms^{-1} impacts the dummy in the thorax, as shown in Fig. 22. The correlation results for the head, chest and pelvis accelerations and chest deflections are shown in Fig. 23 to Fig. 26. The three acceleration graphs are plotted to the same scale. In the physical pendulum test, the vertical and lateral components of acceleration were not measured, and so the accelerations compared here are for the longitudinal component only. The accuracy of the head and pelvis accelerations is encouraging. It demonstrates the correct load transfer between the dummy components.

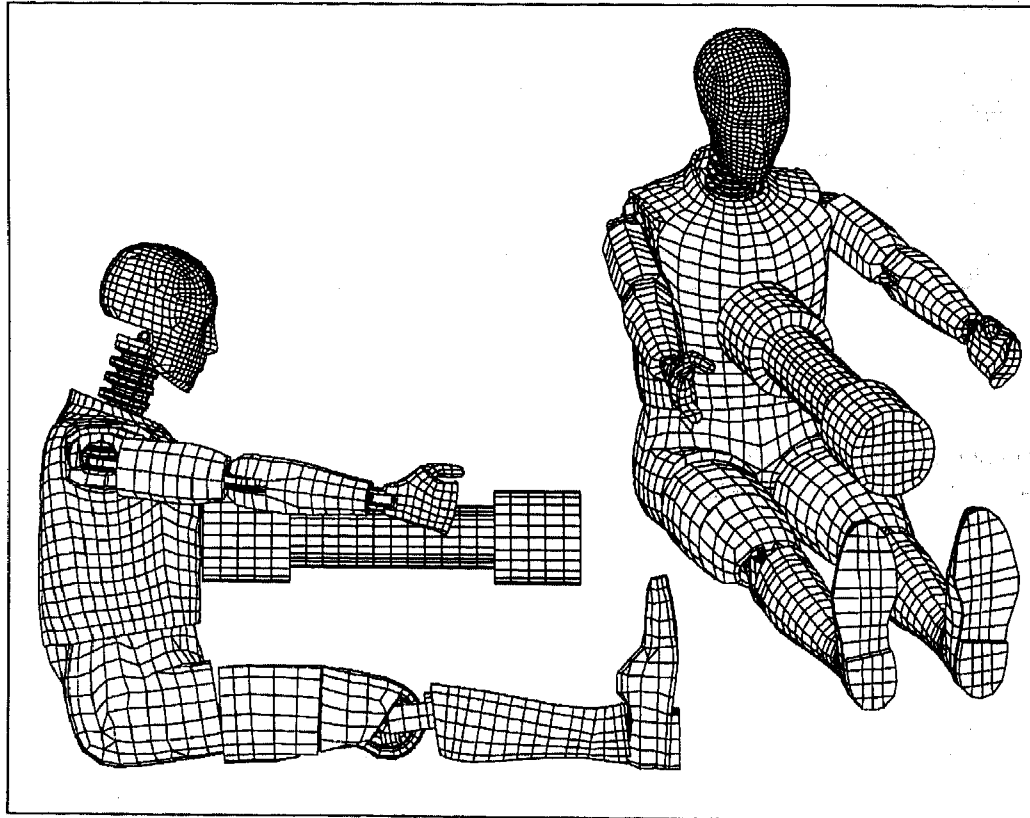


Fig. 22 Thorax Pendulum Test

THORAX PENDULUM TEST VALIDATION

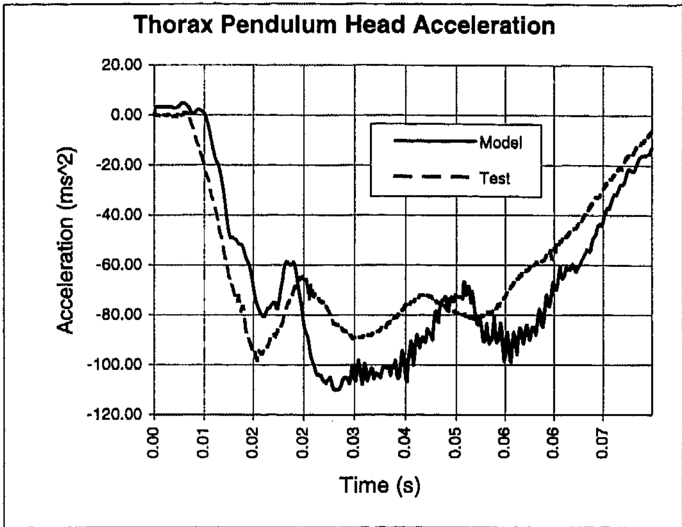


Fig. 23 Head Resultant Acceleration

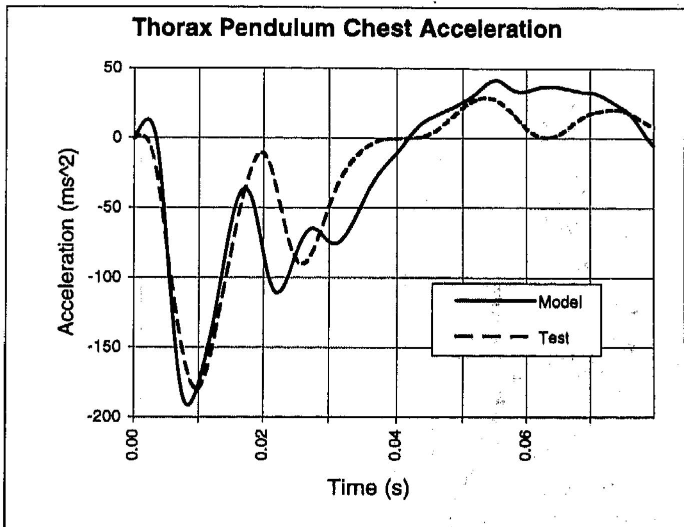


Fig. 24 Chest Resultant Acceleration

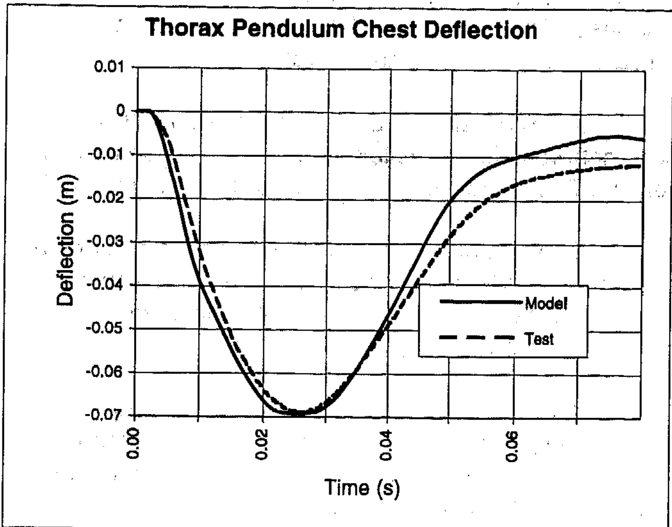


Fig. 25 Chest Deflection

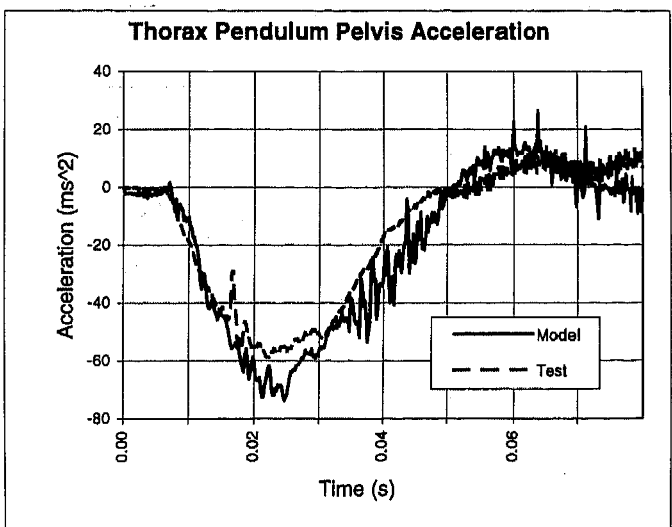


Fig. 26 Pelvis Resultant Acceleration

BELTED SLED TEST VALIDATION

To prove the validation of the model, the sled testing was specifically designed to be highly controlled with as few unknowns as possible, so that the pure response of the dummy model could be measured and validated and not the restraint system.

A rigid seat structure was fabricated with seat belt anchorage locations taken from an average mid-size vehicle. There was no knee bolster or instrument panel.

The restraint system comprised a lap belt and a separate shoulder belt made from low strain webbing and bolted directly to the rigid seat. No slip-rings or retractors were used.

A sled acceleration pulse was developed representing a typical 30mph front impact and standard Hybrid III instrumentation was used. Seat belt load cells were used to measure seat belt forces.

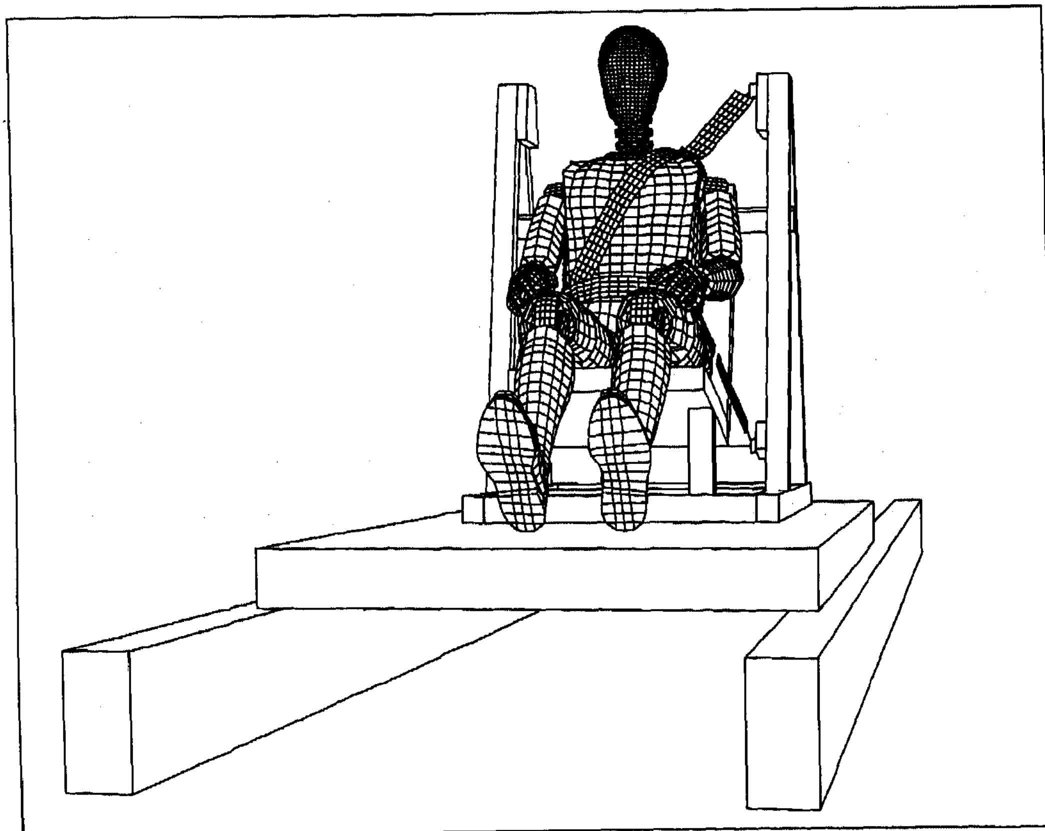


Fig. 27 Belted Sled Test Validation

SLED TEST SIMULATION

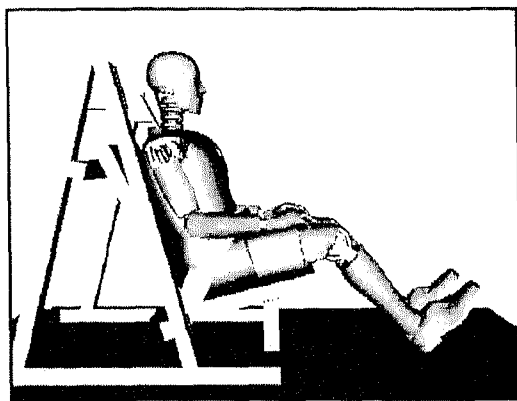


Fig. 28 Model Time 0ms

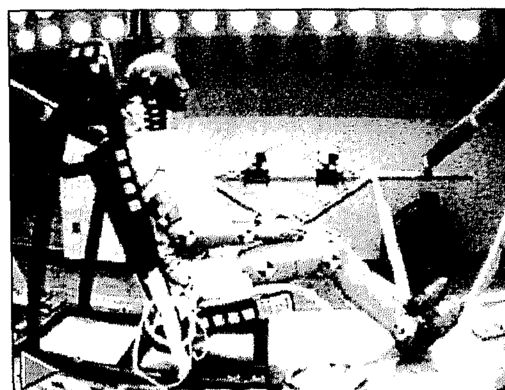


Fig. 30 Test Time 0ms

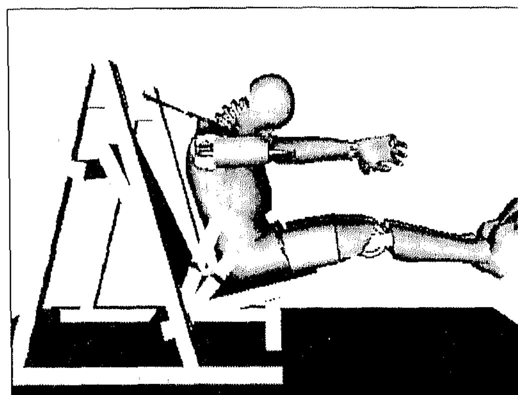


Fig. 29 Model Time 100ms

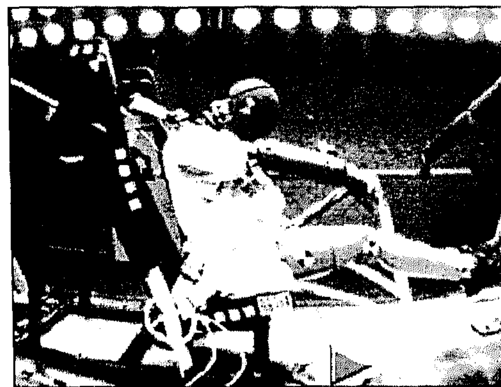


Fig. 31 Test Time 100ms

BELTED SLED TEST RESULTS

Despite the controls put on the sled test specifications, the initial belt slack proved difficult to measure. The seat belt load cell mass caused significant belt sag under static conditions. Parametric studies were used to achieve an accurate representation of the belt shape in the model and the shape of the belt was varied until the timing of the initial increase in belt loads matched the test data.

High speed cameras recorded the test. Fig. 28 to Fig. 31 compare the kinematic response of the model and the physical test. Correlation is generally good, except that at 100ms the hand motion is different. Parametric studies have shown that the limb motion is extremely sensitive to the exact torque settings of the friction joints, and there may be significant variation in this joint stiffness on the physical dummies depending on test set-up.

Two nominally identical sled tests were performed with the same physical dummy and there was very good repeatability between the tests. The correlation for key dummy injury criteria is shown in:

Fig. 32 Head Resultant Acceleration.

Fig. 33 Chest Resultant Acceleration.

Fig. 34 Pelvis Resultant Acceleration

Fig. 35 Chest Compression.

Femur loads are not discussed here since there was no knee bolster used in the test, and the loads would be inertial only.

When run on a HP J210XC workstation, this sled test model required 17.1 hours of CPU to reach 150ms, using a time step of 1.54μs.

This sled test is just the first of a series of tests designed to provide a comprehensive database for the validation of the Hybrid III dummy models. An unbelted sled test with a driver's side airbag is currently being designed and will be the next validation condition for the dummy model.

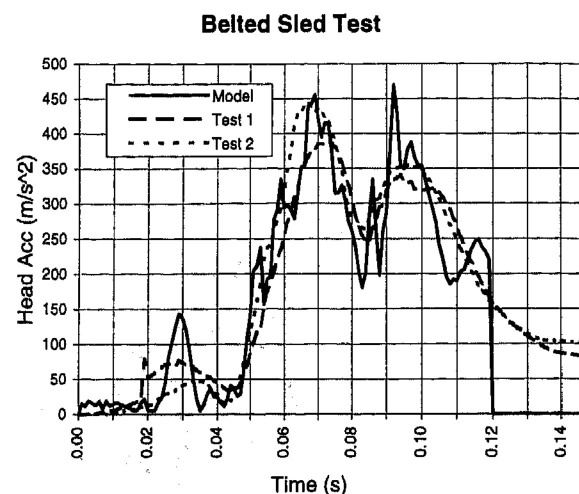


Fig. 32 Head Resultant Acceleration.

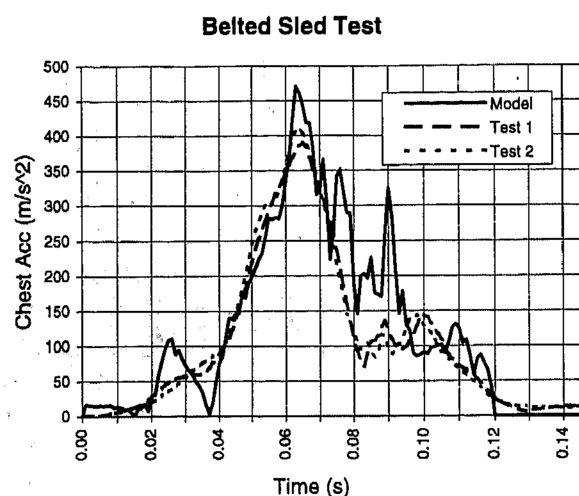


Fig. 33 Chest Resultant Acceleration.

BELTED SLED TEST RESULTS

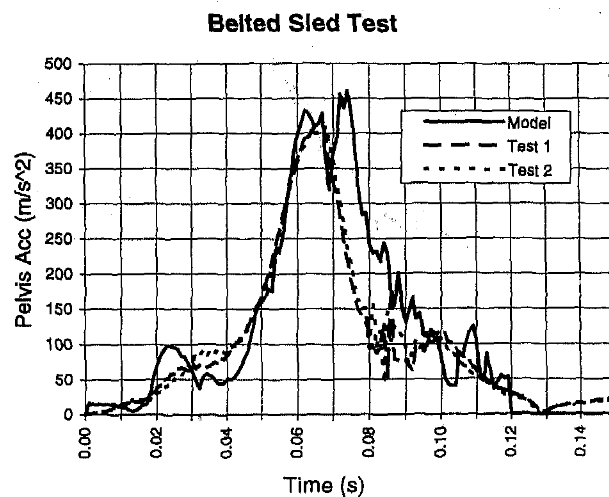


Fig. 34 Pelvis Resultant Acceleration

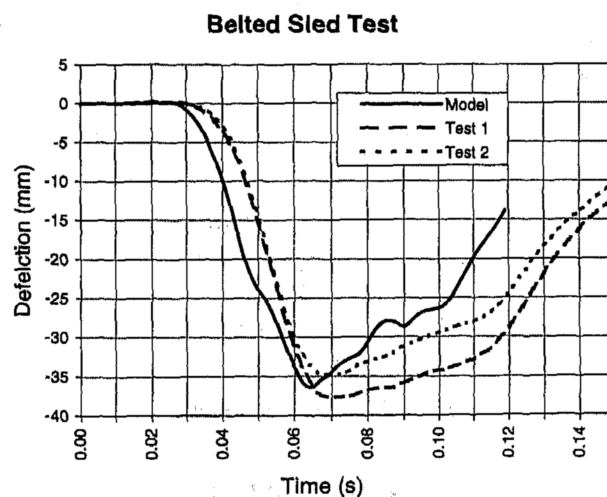


Fig. 35 Chest Compression.

AIRBAG SLED TEST

A second series of sled tests was performed to replicate an unbelted driver side occupant restrained by an airbag and knee bolster at a 30mph impact. As per the belted sled test, the philosophy was to design a test that would remove as many of the variables of the restraint system as possible to focus on the validation of the dummy model.

The seat was the same rigid fixture as used in the belted sled test. The steering wheel was represented by a rigid aluminum disk mounted to a rigid steering column. A six-axis load cell was mounted in the steering column behind the wheel. A rubber block was mounted to the face of the steering wheel disk to provide some compliance to the wheel should the dummy be inadequately restrained by the airbag. Recesses were cut into this block to accept the thumbs of the dummy hands.

The airbag was an existing production airbag, and was unfolded prior to test to avoid the need to have to model the unfolding of the airbag. A pressure tap was placed in the bag in front of the inflator to provide pressure curve data for the airbag model.

The knee bolster was represented by rectangular blocks of polystyrene foam attached to a rigid back plate. The properties for these foam blocks were derived from static compression tests and drop tower tests using spherical and cylindrical drop heads. The toe pan was represented by an inclined rigid plate.

The Hybrid 3 50th dummy had the regular MVSS208 required instrumentation plus upper and lower neck load cells, a six-axis lumbar spine load cell, pelvis accelerometers and upper and lower tibia load cells on one leg. The ankle was the 45 deg hard stop design.

The airbag sled test was repeated twice for the Hybrid 3 50th dummy with good repeatability as indicated in the attached test results.

AIRBAG SLED TEST RESULTS

All analysis and test data is filtered to the SAE classes specified by the Code of Federal Regulations.

The dummy exhibited acceptable kinematics and lifted slightly off the seat at the peak restraint of 90ms. The ankle stops did not bottom out.

The correlation is generally very good for all the measured channels. The lumbar spine shear load and neck moments were also well simulated.

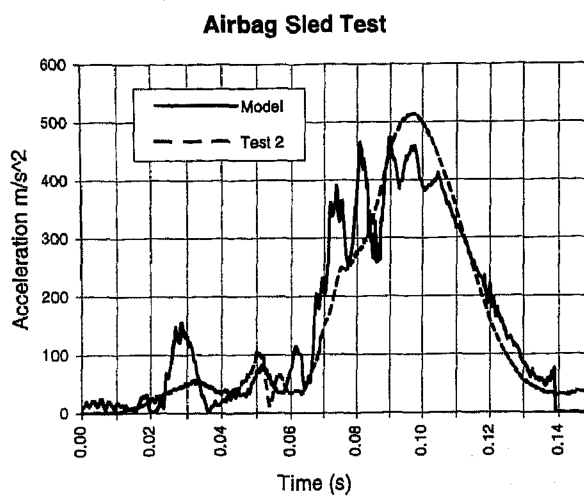


Fig. 36 Head Resultant Acceleration

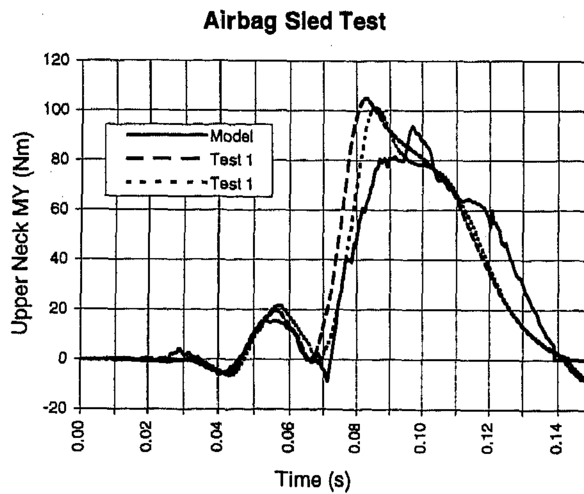


Fig. 37 Upper Neck Moment My

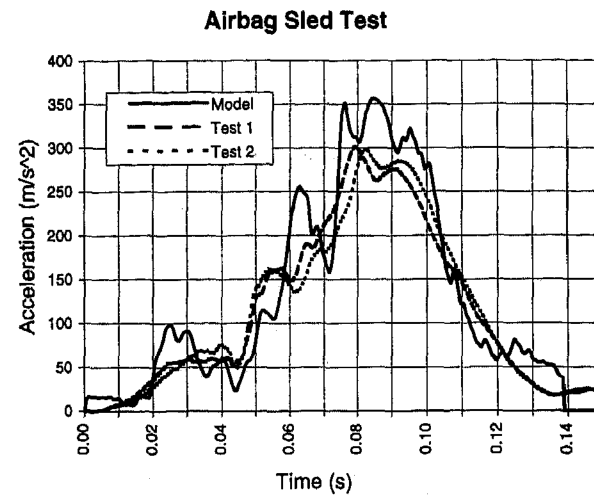


Fig. 38 Chest Resultant Acceleration

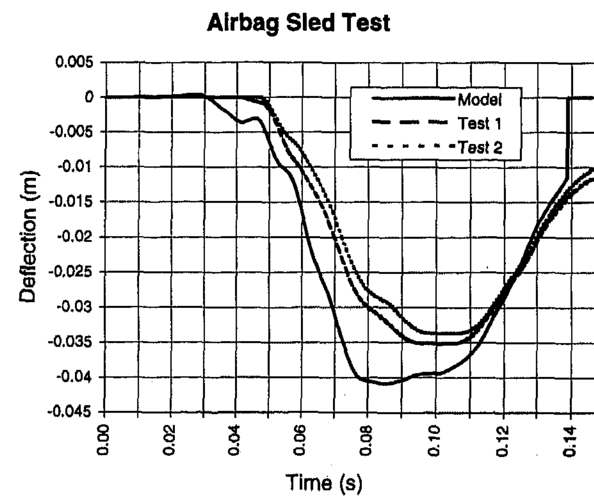


Fig. 39 Chest Deflection

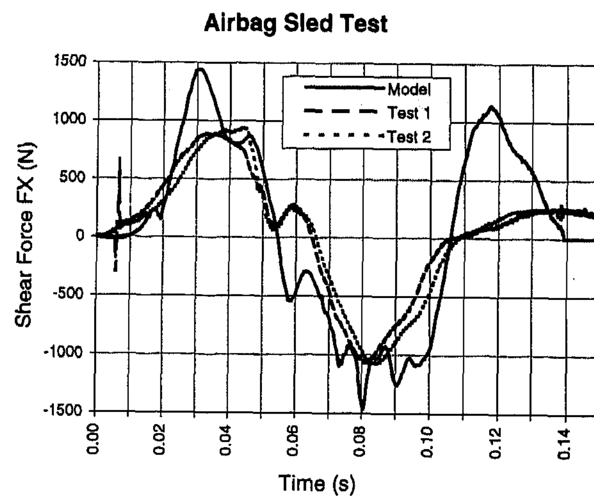


Fig. 40 Lumbar Spine Shear Fx

AIRBAG SLED TEST RESULTS

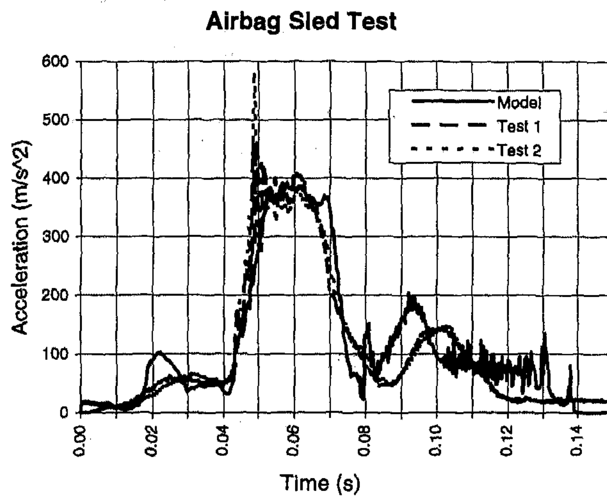


Fig. 41 Pelvis Resultant Acceleration

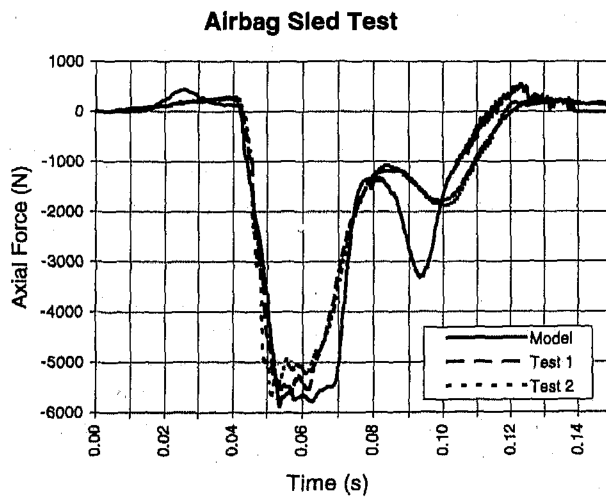


Fig. 42 Femur Axial Force

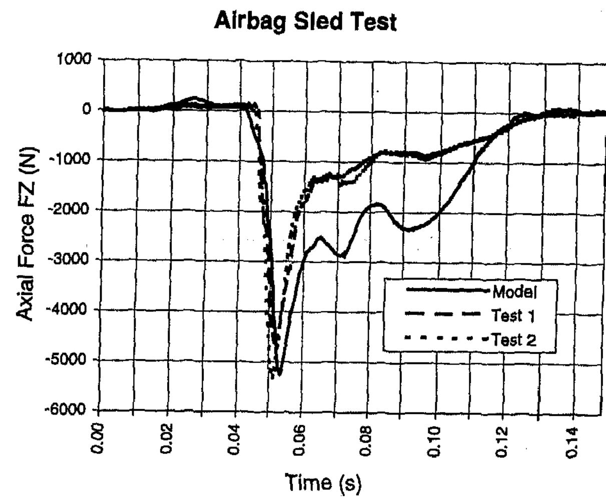


Fig. 43 Lower Tibia Axial Force Fz

CONCLUSIONS

An LS-Dyna model of the 50th percentile Hybrid III dummy has been developed. The development was a close collaborative effort between the dummy manufacturer, First Technology Safety Systems and Ove Arup & Partners International. Primary model development and validation was performed at the component level which has allowed a rigorous approach to model development. The assembled dummy model was also validated against results from pendulum and sled tests.

971045

Limitations of ATB/CVS as an Accident Reconstruction Tool

Michael B. James, Ronald P. Nordhagen, Charles Y. Warner, Douglas "L" Allsop, and Thomas R. Perl
Collision Safety Engineering, Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Occupant simulation models have been used to study trends or specific design changes in "typical" accident modes such as frontal, side, rear, and rollover. This paper explores the usage of the Articulated Total Body Program (ATB) as an accident reconstruction tool. The importance of model validation is discussed. Specific areas of concern such as the contact model, force-deflection data, occupant parameters, restraint system models, head/neck loadings, padding, and intrusion are discussed in the context of accident reconstruction.

INTRODUCTION

The Articulated Total Body (ATB) program is also known as Crash Victim Simulator (CVS); it is the forerunner of another program called DYNAMAN. CVS was originally developed in the early 1970's at Calspan under NHTSA sponsorship. During the 1980's it was enhanced for use by the Air Force, and was re-named Articulated Total Body. Additional improvements and enhancements have continued to be made over the years. In 1989, GESAC, Inc. adapted ATB version IV.2 to run on personal computers. They also developed a user-friendly pre-processor and post-processor. This package is called DYNAMAN and is commercially available. In this paper, ATB refers also to CVS and DYNAMAN.

ATB is a general purpose lumped mass program. It can be used as a three-dimensional occupant simulation program by 1) creating "occupant" body parts of lumped mass ellipsoids connected together by joints, 2) building a "vehicle" by defining contact surfaces and ellipses, and then 3) specifying motion for the "vehicle." Each occupant body part consists of a lumped mass at a given location and a skin of elliptical shape. This is used to define contact between other ellipses and contact planes. The occupant body parts are interconnected by joints with defined properties which restrict the motion of the joint. The program calculates when and where contacts occur between the various ellipsoids & planes and applies forces as a result of those contacts. The occupant is displaced inertially and by the application of these contact forces.

ATB is an initial value simulation program. The input parameters are used to describe the initial occupant kinematics and contact parameters. The output of the first time step becomes the input for the second time step. This process continues through many iterations. As with all initial value programs, small errors in the input values can propagate or grow as the program goes through a large number of iterations.

The ATB program has undergone, and continues to undergo, changes and improvements [Fleck, 1991]. Some of these modifications are proprietary and are not publicly available [Deng, 1990, Prasad, 1993, Grew, 1985]. Most modifications were made in order to achieve reasonable simulations for specific applications. When reviewing the literature it is important to note that not all studies are using the same version of the program. In 1996, and ATB users group was officially formed to help users learn more about ATB, make them aware of new innovative modeling techniques, new applications, peripheral programs, problems and remedies.

The program is relatively complex to use. A typical input file contains 3000 to 7000 parameters. The appropriate values for many of these input parameters are not readily available. There are no recognized "default" or "typical" values. In practice, an existing input file is usually modified to fit the new application rather than creating a new input file from scratch. The input values in the existing files are often used when specific data for the new application are not available, even though the source of the existing data is often not known. Most of these data are not physically measured values, but are values which "seemed to work" in a prior simulation. Reliable input data are very sparse.

DESIGN PROCESS

Occupant simulation programs, such as ATB, can be very powerful tools when integrated into the design process. Once validated, they can be used to evaluate the effects of

parameter changes on the system response and substantially reduce the number of sled tests or full scale crash tests in a design or research project. They can be versatile and economical engineering tools.

For example, British Leyland used a modified version of ATB as a design tool for optimizing the three-point restraint system in a pre-production vehicle [Grew, 1985]. Sled and component tests were run and the appropriate input and output parameters for the model were measured. The input parameters were then adjusted within a reasonable range until the sled test was simulated. Modifications to the restraint system were then made in the model and the corresponding effects were evaluated. Based on the modified simulations, a second sled test was run with a modified restraint system. The inputs were then adjusted to get the best simulation for both sled tests. Further optimization of the model led to additional modifications of the restraint system. Additional sled tests were run to verify the effects of these modifications. Eventually, a well validated computer simulation was developed allowing the design parameters to be better understood. As the researchers pointed out, the tests established and validated the model, while the model helped to interpret the test results and insure that unnecessary testing was avoided [Grew, 1985].

In another research project, the dangers of invalidated modelling were pointed out. There, researchers attempted to develop an occupant simulation model of an unrestrained occupant in a frontal impact (using MVMA 2-D). Difficulties with determining appropriate input parameters led to the conclusion that "it is dangerous to draw conclusions from the model without having convinced oneself that the model matches up well with test results in a situation which is not too much different from the one under consideration" [Forrest, 1984]. In other words, validation is an important part of the application of occupant simulations in the design process.

VALIDATION

The crux of every mathematical modeling endeavor is validation. Validation is the process of establishing how well the predictions of a simulation agree with reality. The usefulness of a computer simulation depends on the extent to which it has been validated. Validation can be a difficult and time-consuming task for a complex simulation program such as ATB. A computer simulation cannot completely model every aspect and detail of a real event because of the innumerable complex physical events taking place. Simplifying assumptions must be made. Therefore, model validation is not a black and white issue; there are levels of confidence associated with each validation effort.

In reviewing the ATB literature, one finds different usages of the term "validation". It has been used to describe the program code as "working," i.e., that the code does not have any programming bugs and produces output that is not random. The more proper use of the term "validation" is demonstrating a specific application of the program properly simulates real physical events such as crash tests or sled tests. Fleck and Segal made this distinction by using the terms verification and validation. They verified that the ATB program ran without crashing and produced output that was reasonably related to the input. However they were unable to accurately model a real event using the ATB program [Fleck, 1982, Segal, 1984].

Since ATB has not been validated in general, it must be uniquely validated for each specific application in order for the results to have scientific meaning. Validation requires well documented crash test or sled test data to (1) establish the ATB input parameters, and (2) provide a basis for judging the reasonableness of ATB program output values. There are difficulties in obtaining appropriate test data for validation because many of the ATB input and output parameters are not routinely measured in crash tests or sled tests. In fact, some of the parameters cannot be directly measured and must be inferred from other measured data.

A true validation protocol requires that the input values be measured independently from the test that is being simulated. For example, input parameters measured from a series of bench or component tests might be compiled to create an input file that represents the actual test to be simulated. This is difficult with complex programs such as ATB, because some inputs cannot be directly measured in component tests. Therefore, such programs are often validated using a pair of full scale tests which are very similar. The first test is used to "calibrate" the model by adjusting the input values within reasonable limits until the measured loads and occupant kinematics from the first test are accurately simulated. The validation simulation is then made by changing only those inputs which represent the differences between the two tests (such as the crash pulse). The second simulation is then compared to the second actual test to determine whether validation has been achieved.

If only trends in "typical" impact modes are to be evaluated, a less stringent validation protocol may be sufficient. For example, a protocol of accurately simulating just one test has been used in some research applications. This procedure can be criticized as being merely a curve fitting project, i.e., showing that the program can generate curves to match the experimental data. There are likely multiple combinations of reasonable input data that could be used with the program to match the experimental curves. Before relying on the results of a study validated in this manner, additional full scale testing ought to be done.

Validation is difficult when reconstructing real world accidents because there is rarely a well documented full scale test with which to compare the simulation output. Sometimes claims of "validation" have been made by matching the injuries in a real world accident to injury threshold values in the simulation. This approach has several problems. First, the program is generally not able to accurately predict injury measures in actual tests. Its predictive power is limited to gross occupant motion. Second, the relationship between dummy injury measures and actual human injuries is not well established for individual exposures. Third, the simulation often requires unreasonable or arbitrary input values in order to predict the injury exposure seen in the accident. Tweaking

variables beyond acceptable ranges so the program matches a known injury is not a scientific endeavor and can result in conclusions that are erroneous and misleading.

Robbins [1989] presented a useful method of quantifying levels of validations in the form of the "Proposed Validation Index" developed by the Human Biomechanics and Simulation Standards Committee of SAE (see Table 1). He noted that for a simulation to be useful as a design parameter trend indicator, a validation level 5 is generally considered sufficient. This level of validation has not been achieved in published efforts using the ATB program [Kaleps, 1982, Smith, 1993, Ma, 1995].

Validation of ATB simulations has traditionally been difficult because of limitations of the program. More sophisticated occupant motion simulation programs, such as MADYMO and finite element impact models, have had much greater success. [MADYMO, 1990 Barbat 1995, Chang, 1994]. As occupant motion simulation programs continue to develop and achieve validation in the research arena, their use will naturally spread to accident reconstruction. However, strict validation will always be a problem when simulating real world accidents without test data for comparison.

In the absence of proper validation, a few keys can be used to evaluate how well a simulation matches an actual accident. First, there should be no obvious errors in the input file, such as improper vehicle geometry, missing contact parameters, occupant motions beyond the capability of the human body, etc. Second, the output must reasonably match the known facts of the accident. Third, the output must not violate the laws of physics. (There are no internal checks within the ATB program to verify consistency with laws of motion or physics.) Fourth, the simulation must actually be modeling what the user intends it to model. When the user focuses on certain output values and adjusts the inputs until those desired outputs are achieved, the simulation oftentimes will be doing things inconsistent with the initial modeling goal. Fifth, one must consider all of the outputs of the simulation: focusing on certain output parameters while ignoring others is bad science.

While it is difficult to create an ATB simulation of a real world accident using reasonable input values that match the physical evidence and comply with the laws of physics, simulations that achieve these criteria may add insights in evaluating real-world accidents, even without formal validation.

VALIDATION ATTEMPTS

Some of the problems experienced while using the ATB program can be illustrated by examining two independent attempts to validate a new data set developed to represent the Hybrid III crash test dummy. One attempt was by Obergefell, et al. [1988], the other was by Khatua, et al. [1988]. Both used frontal sled test data from a series of three tests run by Prasad [Prasad, 90]. Each sled test used two dummies; one Hybrid II (H2) dummy and one Hybrid III (H3) dummy. The sled buck consisted of a plywood seat, toe-board and floorboard, and a 3-point restraint system. The use of plywood surfaces minimized the effect of difficult-to-measure force-deflection functions. The lack of any dashboard, steering wheel, or windshield, eliminated the effect of additional contacts.

Obergefell attempted to validate both the H2 and H3 data sets using a single frontal sled test. This was not a validation in the strict sense because only one sled test was simulated; it was really an exercise in calibrating the model to match the experimental results of the single sled test.

The H2 data set was constructed by Fleck [1982] and consists of 15 ellipsoids. The H3 data set was constructed by Kaleps [1988], and consists of 17 ellipsoids, the same ellipsoids as the 15 in the H2 data set, with the addition of hands. Obergefell modified both the H2 and H3 data sets by adding two additional massless contact ellipsoids, one at the upper torso and one at the lower torso. These additional ellipsoids did not alter the mass or inertial characteristics of the dummy but were added to provide a more realistic geometry for the belts to contact. The seat belt webbing stiffness and friction coefficients were not independently measured. The input parameters were simply adjusted to values which seemed to provide the best fit with the experimental data.

The gross occupant motion was assessed by comparing the high speed film to the simulation's graphical output. Both the H2 and the H3 simulations had occupant motion similar to the test data during the initial part of the event. However, the motion in both simulations began to deviate from the test as the event progressed. The simulation's acceleration data for the head, chest and pelvis, also matched the test data reasonably well early in the event, but then developed substantial deviation. The match of head acceleration was much worse than that of the chest and pelvis acceleration. Neck forces were not reported. Interestingly, the older H2 data set matched the experimental data better than the newer H3 data set, even though the H3 data set was derived using more precise measurements of the dummy.

The inability of the simulations to match the experimental data was attributed to difficulties with the belt model. The authors noted that even small changes in the belt parameters significantly affected the simulated occupant motion. Since the sled test belt forces were the most significant forces acting on the dummy, the belt model was clearly critical to simulating the dummy motion. However, the belt parameters were not independently measured in this exercise; the indication being that the model was simply not capable of matching the experimental data.

Khatua, et al. [1988] attempted to validate the same H3 data set using a pair of sled tests from the same series by Prasad. However, a more rigorous validation protocol was followed; the first sled test was used to calibrate the model and the second test was used for validation. Force-deflection data were independently measured by statically pressing the lower torso into a plywood seat cushion. The lap belt characteristics were measured by pushing the H3 dummy against the lap belt

TABLE 1 - PROPOSED VALIDATION INDEX

Level 0	No agreement between predictions of model and a "reference event."
Level 1	Qualitative Agreement a. Trends of predicted parameters same. b. Kinematics correspond qualitatively. c. Contacts between occupant and vehicle interior are the same in general.
Level 2	HIC and similar indicators predicted by simulation are within 20% of those obtained in reference event.
Level 3	Peak values of important occupant responses limited to a relative error of 20% (20% on vector magnitude, 11.31 degrees on vector direction).
Level 4	Same as above except 5%.
Level 5	Timing of peaks of important vector responses limited to 5% relative.
Level 6	All peaks and valleys in the duration of time-dependent predictions must match the reference event within 10%.
Level 7	Same as 6 except 5%.
Level 8	1% relative error, point by point, over the duration of the reference and predicted event.

in a quasi-static fashion. A generic dynamic amplification factor of 1.25 was applied to the measured data. The shoulder belt characteristics were not directly measured, but were adjusted to provide the best fit with the experimental chest acceleration and shoulder belt load data.

Certain adjustments were required in the H3 data set in order to simulate the "calibration" test. The model predicted substantial bending at the occupant's waist. No such bending was seen in the films. The bending was eliminated by locking the joint between the middle and lower torso. The head and neck motion in the simulation indicated that the neck stiffness parameters in the model were too stiff by a factor of 8. The neck stiffness was reduced to one-eighth of its original value for neck rotations up to an angle of 10 degrees, and then increased to its original value at 20 degrees of rotation. Even with these modifications, the head motion of the calibration test was not accurately simulated because the neck parameters did not include energy absorption (the reason for this is discussed later in this paper). The knee joint values resulted in simulated femur forces that were too low, but attempts to find a combination of parameters that more accurately simulated the test results were unsuccessful.

Once the input parameters had been adjusted in the calibration run to give the best possible match to the test data, the crash pulse was modified to that of the second sled test. The validity of the model was then determined by comparing the second sled test to the validation simulation. The chest accelerations and the shoulder belt forces were in reasonably good agreement, which is to be expected since the shoulder belt parameters had not been independently measured, but had been adjusted to give the best fit in the calibration run. The head accelerations, however, were too high and the lap belt forces were not reasonable. It was concluded that the simulation results did not fully agree with the test results and that the measured stiffness values in the H3 data set for the neck joints, torso joints and knee joints needed to be re-evaluated.

Neither of these attempts to validate the new H3 data set were successful. Obergefell concluded that the data set was probably valid, and the belt model was responsible for the bad results. Khatua concluded that the data set was the problem. It is likely that there are problems with both the data set and the belt model, in addition to other limitations inherent in the program.

The remainder of this paper will explain some of the difficulties in validating the ATB program for even simple frontal sled tests. These limitations exacerbate the difficulty in attempting to simulate a real-world accident.

MODEL LIMITATIONS

With any math model, it is necessary to understand the underlying equations, assumptions, and simplifications in order to evaluate the validity of the output. The more complex the model, the more important this understanding is. This paper does not address all the inner workings of ATB in detail, but presents a description of specific factors which clarify the difficulties in validating simulations for reconstructing accidents.

CONTACT MODEL - The ATB contact model is what "drives" the program. It computes the contact forces which act on the occupant and produce the occupant motion. It is a simplistic model of what is in reality a very complex

event. Even though the ATB program simulates occupant motion in three dimensions, the contact model is essentially one-dimensional.

The user must define a set of "allowable contacts" between the various body segments and vehicle surfaces. Each individual contact definition specifies the force vs. deflection properties which govern that contact. Contact is deemed to occur when an ellipsoid of the occupant intersects a vehicle surface. If a contact is not defined between a specific body segment and vehicle surface, the body segment will pass through that part of the vehicle structure as if it did not exist. Even when the contact has been defined, the ellipsoids and planes do not deform, they still pass through each other, but the geometry of the overlap is used to determine three aspects of the contact: depth of penetration, direction of force application and the point of force application.

Penetration is defined as the perpendicular distance from the plane to the deepest point on the overlapping ellipsoid (Figure 1). The magnitude of the contact force is determined by applying the depth of penetration to a user-defined force-deflection function. For oblique contacts, this is not the actual length of deflection. Therefore, force-deflection data measured in perpendicular tests will not appropriately represent contacts at oblique angles. A single force-deflection function represents the mutual crush of both surfaces. If separate data for the vehicle component and the body segment are available, they must be combined to form a single function for use in the model. Ignoring the compliance of the body segment is equivalent to modeling a rigid occupant, which will result in erroneous occupant accelerations and joint forces.

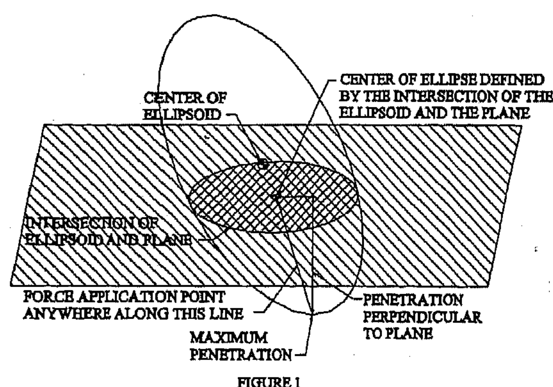


FIGURE 1

The contact force is the resultant of the normal and the friction components. The normal component is always perpendicular to the contact plane. The frictional force is parallel to the plane and directed so as to oppose the ellipsoid's motion. Its magnitude is a function of the normal force. When the magnitude is modeled as being directly proportional to the normal force, the direction of the resultant contact force will be at a constant angle relative to the plane (Figure 2). This does not appropriately describe the pocketing effect which often occurs when body parts contact deformable surfaces and the direction of the contact force changes as the deformation increases.

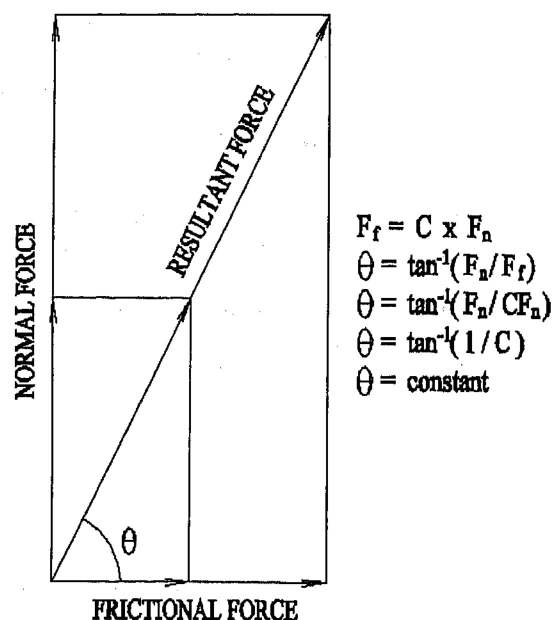


FIGURE 2

Since the ATB program does not model the deformation of the contact surfaces and the resulting pressure distribution, the entire contact force must be applied at a single point. The location of the force application can be defined by the user anywhere along the line connecting the point of maximum penetration on the ellipsoid and the centroid of the ellipse created by the ellipsoid/plane intersection (Figure 1). The location of the force application remains constant for all magnitudes of deflection. It should be chosen based on how much each of the two surfaces deform. For example, if a rigid head form contacts a soft dashboard, the contact force should be applied closer to the surface of the head form (Figure 3a). Whereas, a deformable abdomen contacting a stiff, unyielding plane would be better modeled by choosing the point of force application closer to the surface of the plane (Figure 3b). For contacts where there is deformation of both the plane and the ellipsoid, care must be taken in defining the point of force application, especially when large deformations occur.

The contact model used in ATB has difficulty when ellipsoids contact narrow planes or the edge of larger planes. There are several program options from which the user may choose. The default option results in no contact force unless the center of the ellipse defined by the intersection of the ellipsoid and the plane is within the plane boundary. This can result in very unrealistic results because the contact force can change instantly from a high value to zero as the center of the ellipse moves beyond the edge of the plane. Another option treats the plane as if it were infinitely large as long as any part of the ellipsoid intersects it within the original plane

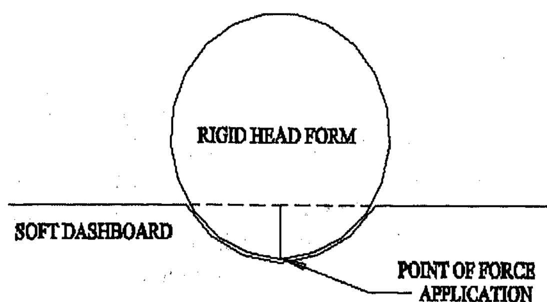


FIGURE 3A

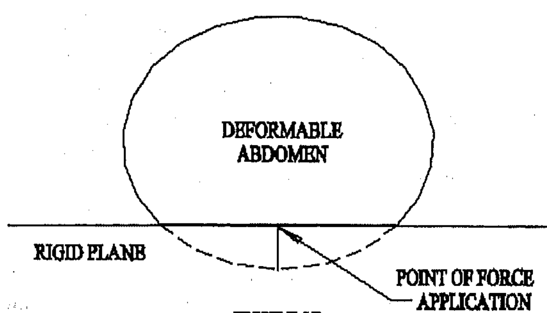


FIGURE 3B

boundaries. The third option reduces the magnitude of the contact force proportionately to the area of the intersection; however, the direction of the contact force is not modified. None of these program options adequately models the true interaction between a body segment, such as the head and narrow vehicle surfaces like the A pillar or steering wheel rim. Creative modeling by the user, however, can be used in some circumstances to overcome this program limitation.

FORCE-DEFLECTION DATA - When defining an ATB input file, every potential occupant segment/vehicle surface contact must be individually defined. If no definition is made, the occupant segment will pass through the plane as if it was not present. Since a single force-deflection function is used to define the mutual deformation of both surfaces, different force-deflection functions are necessary to define the contact between the same vehicle surface and the various occupant segments.

Very few data are available to characterize these force-deflection properties. They are different for each vehicle make and model, and they are difficult to measure. Segal developed a test methodology to statically and dynamically measure properties for several vehicles [Segal 87]. It was hoped that "generic" values could be developed for use in many applications. Unfortunately, the testing indicated a wide range of stiffness values that were very sensitive to slight variations. Interior surfaces of 19 different automobiles were tested using both static and dynamic tests. "Typical" or "average" stiffness characteristic for vehicle interior surfaces could not be determined. They also found that static tests were not representative of dynamic tests, and that there was no consistent amplification factor which could be used to convert static test results to dynamic force levels. Small changes in the direction of loading and the specific location of loading resulted in large variations of measured force-deflection properties.

One cannot expect accurate simulations using generic force-deflection values. Even force-deflection values which have been established for a specific area of a specific vehicle can have substantial variations depending on the location of loading, the angle of loading, and the rate of loading. This extreme variability is one of the most significant problems with using ATB for accident reconstruction. It is not as significant a problem when using ATB as a design tool because the force-deflection values can be modified within reasonable limits until a fit with the test data is achieved.

OCCUPANT PARAMETERS - An ATB occupant is represented by a group of ellipsoids connected together by joints. Each ellipsoid must be defined in terms of size, mass, location of the center of mass, moments of inertia, and joint locations. There are readily accessible sources of data which can be used to compile the ellipsoid data for standard crash test dummies, and also for various size occupants.

As mentioned previously, Fleck [1982] has developed a data set for the Part 572 (Hybrid II) dummy and Kaleps developed a data set for the 50% Hybrid III dummy, with both the seated pelvis and the standing pelvis [Kaleps, 1988]. A program called GEBOD has been developed to create data sets for occupants of various sizes [Cheng, 1994]. The user specifies the occupant's height, weight, or both, and GEBOD will appropriately scale the relevant occupant parameters. The data for adults is based on surveys of Air Force personnel [Grunhofer, 1975, Clauser, 1972] as well as other Air Force studies [McConville, 1980, Young, 1983, Engin, 1987]. The data for children's sizes comes from the Snyder anthropometry for children [Snyder, 1977] and the joint torque data comes from Kaleps [1982].

Recently Ma, et al., developed a partial set of "human" joint parameters [Ma, 1995]. An undocumented set of "human like" joint parameters is included in GEBOD version 4.1.

The joint parameters which are most critical in evaluating occupant injury exposure are also the most difficult to define; namely those characterizing the neck. Head and neck injuries represent the majority of injury in automobile impacts. However, accurately simulating head motion, and predicting head/neck loadings is an area of research that is still developing. The current ATB head/neck model consists of a rigid neck segment connected to the head and the upper torso with a pair of joints. The joints allow rotation between the neck and the adjacent segment, but there is no compliance in the axial or shear directions (Figure 4). Physical neck models, on the other hand, such as the Hybrid II and Hybrid III necks,

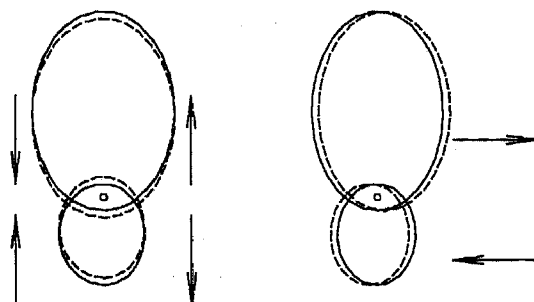


FIGURE 4

are flexible segments with a fixed attachment to the upper torso and limited joint rotation at the head attachment.

The neck of the Hybrid II dummy consists of a single rubber cylinder with a cable running through the center axis. It has bending characteristics which are the same in each direction, i.e., flexion, extension and lateral bending can be modeled with the same characteristic. The human neck, however, does not have the same characteristic in each direction, so the Hybrid III dummy neck was designed to have resistance to extension motion be about half the resistance to flexion motion. It is constructed of six aluminum disks separated by rubber disks, held together by a cable running through the center of the structure. Cuts on the anterior portion of the rubber disks are used to create the different resistance to flexion and extension.

Data sets of the H2 neck have been developed by Paver [1987] and Doherty [1986]. Several attempts have been made to create joint parameters to simulate the H3 neck [Kaleps, 1988, Doherty, 1987, Seemann, 1986, Wismans, 1988]. There is a particular problem, however, with modeling the Hybrid III neck. ATB allows joint characteristics to be defined by either an equation, or by specifying tabular data. When the equation method is used, damping properties can be defined, but only one equation can be used to specify the torque vs. angle characteristic for all directions of bending. When the tabular data method is used different characteristics can be defined for different direction of bending, but no damping can be modeled. The Hybrid II neck can be modeled using the equation method (because it has the same bending characteristic in each direction), and damping properties can be defined. However, the tabular method must be used to model the Hybrid III neck (because it has different bending characteristics) and damping properties cannot be defined. Head motion cannot be accurately simulated without the damping.

The rigid neck model may, with proper evaluation, be used to simulate gross kinematics of head as long as the head does not hit anything. However, better models and more data are necessary before accurate simulations of head contact kinematics and forces can be achieved. A more sophisticated neck model for the ATB program has been developed by Deng [1989]. It more closely approximates the flexible H3 neck by treating each of the six aluminum disks as an individual ellipsoid.

BELT RESTRAINT SYSTEMS - The ATB program has two options for modeling belt restraint systems: The "BELT ROUTINE" and the "HARNESS ROUTINE." The belt routine does not yield reasonable results in most situations, is generally not used, and will not be discussed [Deng, 1989, DYNAMAN User's Manual 2.3.2]. The harness routine is a more sophisticated model, but it is still incapable of accurately simulating occupant belt interactions in many situations.

Each belt in the harness routine is modeled as a series of webbing segments that contact the occupant at user-specified points. The belt length is not a user input but is a function of the occupant's initial position and the belt anchorage locations. Additional webbing can be specified by adding belt slack. Using the harness routine in an accident reconstruction context can be misleading because some of the important input parameters are unknown, or cannot be measured directly, and the simulation output can be very sensitive to small changes in the inputs [Obergefell, 1986].

There are readily available data on seat belt webbing characteristics from standard elongation tests. Unfortunately, these data are not useful for ATB applications because the webbing stress-strain input parameter must account for several additional factors in addition to webbing stretch. For example, Deng [1988] modeled sled tests using independently measured webbing force versus strain characteristics. It was noted that the dummy motion in the sled tests was greater than what would have been expected based solely on the webbing stretch. Additional factors such as anchor deformation, retractor spool-out factors, and occupant compliance must be accounted for in order to correctly simulate the occupant motion.

Shimamura, et al., [1987] used a series of bench tests to independently measure webbing strain, webbing unloading characteristics, retractor spool-out (film spool effect) and movement of webbing through the D ring. All of these effects were then combined into a single belt load versus strain characteristic. The retractor spool-out, or film spool effect, was directly related to both the webbing elongation characteristic and the length of webbing on the reel. These researchers note that it is also important to correctly define the hysteresis characteristics (webbing unloading curve) because the maximum occupant displacement occurs after the maximum belt load is achieved.

Dummy (or occupant) compliance must also be accounted for in the belt model. The ATB code has the capability to independently define the compliance of the occupant, but this feature is infrequently used because it leads to numerical instabilities [Obergefell, 1992]. Modeling the upper torso as two separate ellipsoids connected by a spring-damper system can be used to account for chest compliance (and can also be used to determine chest compression and the VC criteria). The other alternative is to adjust the belt strain

characteristics to include the effect of occupant compliance. The same belt characteristics cannot be used for both the lap belt and the shoulder belt because of the differences in body segment stiffness and webbing length.

One might be tempted to use the actual webbing stiffness values in the model and just add a few inches of belt slack to account for other factors. In practice this introduces additional errors because the occupant gets a running start at a stiff belt, resulting in unrealistically high belt loads which alters the occupant motion. The "softening" factors need to build up over time to realistically simulate the occupant motion.

When using ATB in a design context, it is possible to simply adjust the belt input parameters within a reasonable range until the best fit is achieved with the test data. In an accident reconstruction context, it is not possible to do this because there is no test with which to compare, and the actual occupant motion and webbing loads are not known.

PADDING - Head and neck loadings are difficult to model using ATB not only because of the neck model limitations, but also because of the difficulty ATB has simulating impacts with padded surfaces. Modeling the response of energy absorbing interior padding requires a more sophisticated contact model than that used by the ATB program; the impact algorithm must take into account both the deformation of the padded surface and the complex properties of the padding material.

Padding materials are typically nonlinear and visco-elastic; their dynamic properties are strongly strain rate dependent. As a curved object such as the head impacts a padded surface, three things are happening: 1) the area of involved padding increases as the penetration progresses; 2) the strain rate decreases as the head is slowed by the padding; and 3) the stiffness of the padding increases as the material compresses. Therefore, at each instant there are different conditions for each location within the deforming pad material. The model cannot relate these changing factors without actually modeling the deforming padded surface. It is possible to simulate a specific test using ATB because the strain rate and force-deflection input can be matched, but those inputs are for a specific impact velocity and geometry, and are not valid for general application.

A study by Orringer, et al. [1986], demonstrated the difficulties associated with developing accurate math models of energy absorbing interior padding materials. They developed two complex empirical models of padding materials for ATB. One model had nine parameters and the other model had twenty-one parameters. Even with these elaborate models, they had difficulty matching actual test data.

Finite element analysis, however, has been shown to suitably model head impacts with energy absorbing interior padding materials [Barbat, 1995, Chang, 1994, Chou, 1995, Sherman, 1995]. Substantial improvements in ATB will be achieved as finite element neck models and finite element padded surfaces models are merged with the current program. However, until that occurs ATB is not capable of accurately simulating head impacts with padded surfaces.

INTRUSION - The effects of intrusion on occupant response are receiving increased attention. ATB has the capability to model intruding surfaces, however ascertaining the intrusion experienced in a vehicle impact is a very difficult task. The program does not compute the intrusion motion. The user must specify the position of the intruding surface relative to the vehicle at specific times. This user input defines not only the magnitude or extent of the intrusion, but also the velocity of the intruding surface. There are very few test data with which to evaluate the velocity of intruding surfaces. Intrusion parameters are not commonly measured in crash tests and are difficult to measure using high speed film.

When the intrusion deformation is the result of contact with a fixed object or the ground, the ATB intrusion input parameters must be carefully coordinated with the vehicle motion defined by the crash pulse. Without careful coordination, the intruding surface will simulate unrealistic contact velocities between the occupant and a surface that should have no velocity relative to the ground.

When the intrusion deformation is the result of a car-to-car impact, the timing is critical. There is only a small window of time in which the intruding surface is moving. If occupant contact occurs within that window the contact velocity may be higher, resulting in higher forces and accelerations on the occupant. The simulation output will be very sensitive to slight changes in the inputs. In order to use the program to analyze intrusion, good test data is necessary to determine if the timing of the intrusion motion and the timing of occupant contacts are accurate. When reconstructing an actual accident the detailed information necessary to create an accurate intrusion simulation is almost never available. This does not completely preclude the study of initial gross occupant motion, but greatly hampers the reliability of the simulated occupant contact forces and the subsequent occupant motion due to contact with the intruding surface. As finite element models improve and become more accessible, reliable intrusion motion may become available for use in occupant motion simulations.

CONCLUSIONS

ATB can be a powerful tool in the design process when used in conjunction with a series of vehicle or sled tests. However, it is difficult to develop well validated simulations with only one or two tests. The predictive power of the program is generally limited to gross occupant motion. Factors which limit the accuracy of ATB simulations include the simplicity of the contact model, lack of reliable dynamic force-deflection data, limitations on occupant joint parameters, sensitivity of belt restraint model, and an inadequate neck model.

Adjustments in the input parameters can sometimes be made to simulate specific tests. However, there are limitations to using the ATB program for reconstructing occupant kinematics in actual accidents. Some of these

limitations are due to modeling difficulties, and some are due to the inherent lack of data available with real world accidents. As research and development of occupant simulation programs, such as ATB, continues, the ability to model both test data and real world accidents will improve. Merging finite element codes with the existing ATB code may improve the accuracy of simulations in the future, given proper validation.

BIBLIOGRAPHY

BARBAT, SAEED D., and PRASAD, PRIYA, "Finite Element Modeling of Structural Foam and Head Impact Interaction with Vehicle Upper Interior, 1995, SAE #950885.

CHANG, FU S., HALLQUIST, JOHN O., LU, DAVID X., SHAHIDI, BEHROOZ K., KUDELKO, CONRAD M., and TEKELLY, JOSEPH P., "Finite Element Analysis of Low-Density High-Hysteresis Foam Materials and the Application in the Automotive Industry," 1994, SAE #940908.

CHENG, HUAINING, OBERGEFELL, LOUISE, and RIZER, ANNETTE, "Generator of Body Data (GEBOD) Manual," AL/CF-TR-1994-0051, March 1994.

CHOU, C.C., ZHAO, Y., LIM G.G., PATEL, R.N., SHAHAB, S.A., and PATEL, P.J., "Comparative Analysis of Different Energy Absorbing Materials for Interior Head Impact," 1995, SAE #950332.

CLAUSER, Charles E., TUCKER, Pearl E., McCONVILLE, John T., CHURCHILL, Edmund, LAUBACH, Lloyd L., REARDON, Joan A., "Anthropometric of Air Force Women," Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AMRL-TR-70-5, 1972.

DENG, Yih-Charng, "Analytical Study of the Interaction Between the Seat Belt and Hybrid III Dummy in Sled Tests," 1988 SAE International Congress and Exposition, SAE #880648.

DENG, Y.C., "Anthropomorphic Dummy Neck Modeling and Injury Considerations," Accident Analysis & Prevention 21:85-100, 1989.

DENG, Yih-Charng, "An Improved Belt Model in CAL3D and Its Application," Vehicle Crashworthiness and Occupant Protection in Frontal Collisions, SAE SP-807, 1990, SAE# 900549.

DOHERTY, B.J. and PAVER, J.G., "A Computer Simulation of the Hybrid II Manikin Head-Neck System," Proceedings of the 24th Annual Symposium of the Safe Association, 1986.

DOHERTY, B.J. and PAVER, J.G., "Mathematical Modeling of the Hybrid III Manikin Head-Neck Structure," Proceedings of the International Conference on Mathematical Modeling, 1987.

ENGIN, Ali E. and CHEN, Shuenn-Muh, "Human Joint Articulation and Motion-Resistive Properties," Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AAMRL-TR-87-011, 1987.

FLECK, John T., BUTLER, Frank E., DELEYS, Norman J., "Validation of the Crash Victim Simulator Volume II: Engineering Manual - Part II Validation Effort," August 1982, Final Report, DOT-HS-806-280.

FLECK, John T., "Improvements in the ATB/CVS Body Dynamics Model," 13th International Technical Conference on Experimental Safety Vehicles, S8-W-20, pp.974-976, 1991.

FORREST, Stephen M. and SCHWARZ, Robert, "Design and Development of Modified Production Vehicle for Enhanced Crashworthiness and Fuel Economy - Phase II Final Report - Volume II Technical Report," DTNH22-81-C-07085, p. 187, March 1984.

GREW, N. D. "Applying Computer Techniques in the Design and Development of an Occupant Restraint System," 1985, 10th ESV.

GRUNHOFER, H.J. and KROH, G. "A Review of Anthropometric Data on German Air Force and United States Air Force Flying Personnel 1967-1968," AGARD-AG-205, Advisory Group for Research and Development, 1975.

KALEPS, Ints and MARCUS, Jeffrey H., "Predictions of Child Motion During Panic Braking and Impact," 1982, Stapp Conference, SAE #821166.

KALEPS, Ints and WHITESTONE, Jennifer; "Hybrid III Geometrical and Inertial Properties," SAE #880638.

KALEPS, Ints, White, Richard P. Jr., Beecher, Robert M., WHITESTONE, Jennifer, Obergefell, Louise A., "Measurement of Hybrid III Dummy Properties and Analytical Simulation Data Base Development," Harry G. Armstrong Aerospace Medical Research Labs., Wright-Patterson Air Force Base, AAMRL-TR-88-005, 1988.

KHATUA, Tara, et al.; "ATB Simulation of Hybrid III Dummy in Sled Tests," 1988 SAE International Congress and Exposition, SAE #880646.

MA, Deren, OBERGEFELL, Louise and RIZER, Annette L., "Development of Human Articulating Joint Model Parameters for Dynamics Simulations," 1995 39th Stapp Conference, SAE #952726.

MADYMO User's Manual 2D/3D, Version 4.3 (1990) Department of Injury Prevention, TNO Road-Vehicle Research Institute, April.

McCONVILLE, J.T., CHURCHILL, T.D., KALEPS, I., CLAUSER, C.E., CUZZI, J., "Anthropometric Relationships of Body and Body Segment Moments of Inertia," Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AMRL-TR-80-119, 1980.

OBERGEFELL, Louise, et al., "Prediction of an Occupant's Motion During Rollover Crashes," Thirtieth Stapp Car Crash Conference Proceedings, P-189, Oct. 1986, SAE #861876.

OBERGEFELL, Louise, et al., "Simulations of Rollover Tests," DOT HS 807 372, May 1988.

OBERGEFELL, Louise Ann. "Harness Belt Restraint Modeling," UMI Dissertation Services, 9302405, 1992.

ORRINGER, Oscar, et al., "Crash Padding Research," DOT-HS-806-960, July 1986.

PAVER, J.G., FISHBURNE, B., "The Prediction of Hybrid II Manikin Head-Neck Kinematics and Dynamics," Proceedings of the 25th Annual Symposium of the Safe Association, 1987.

PRASAD, Priyaranjan, "Comparitive Evaluation of the Dynamic Responses of the Hybrid II and the Hybrid III Dummies," Thirty-fourth Stapp Car Crash Conference Proceedings, P-175, Nov. 1990, SAE #902318.

PRASAD, P. and CHOU, C.C., "A Review of Mathematical Occupant Simulation Models," Accidental Injury - Biomechanics & Prevention, by Nahum & Melvin, Chapter 6, P-102, 1993.

ROBBINS, D.H., "Restraint Systems Computer Modeling and Simulation State of the Art and Correlation with Reality," 1989, SAE #891976.

SEEMANN, M.R., MUZZY III, W.H., AND LUSTICK, L.S., "Comparison of Human and Hybrid III Head and Neck Dynamic Response," Proceedings of the 30th Stapp Car Crash Conference, 1986, SAE #861892.

SEGAL, David J., "Side Impact Modeling Using Lumped Mass and CAL-3D CVS Simulations," 1984, SAE #840859.

SEGAL, David J., "Vehicle Component Characterization," MGA Research Corp., Jan. 1987.

SHERMAN, Kenneth C., CLARK, Chris, RAWSON, Janet, and FROST, Colin, "Use of High Strain-Rate Material Data for Predicting Head-Impact Performance of Thermoplastics," 1995, SAE #950331.

SHIMAMURA, Munemasa, OMURAM, Hideo, and ISOBE, Hisaaki, " An Occupant Movement Analysis Using Improved Input Data for MVMA-2D Simulation," 1987, SAE# 870332.

SNYDER, Richard G., SCHNEIDER, L. W., OWINGS, C. L., REYNOLDS, H. M., GOLOMB, D. H., "Anthropometry of Infants, Children, and Youths To Age 18 for Product Safety Design," SAE Technical Paper Series, SP-450, 1977.

WISMANS, J., HERMANS, J.H.A., "MADYMO 3D Simulations of Hybrid III Dummy Sled Tests," 1988 SAE International Congress and Exposition, SAE #880645.

YOUNG, Joseph W., Chander, Richard F., Snow, Clyde C., ROBINETTE, Kathleen M. ZEHNER, Gregory F., LOFBERG, Maureen S., "Anthropometric and Mass Distribution Characteristics of the Adult Female," Office of Aviation Medicine, Federal Aviation Administration, FAA-AM-83-16, 1983.

971052

Control and Application of Intelligent Restraint Systems

John A. Musiol, Lynette M. Norgan-Curtiss, and Michael D. Wilkins
AlliedSignal Safety Restraint Systems

ABSTRACT

Conventional restraint systems designed to meet US FMVSS standards only have one level of operation. The seat belt imparts a restraint force to the occupant reflective of belt stiffness characteristics and the airbag is either inflated or not inflated. The system is tuned to one crash scenario, typically a 30 mph (48 kph) barrier crash with an unbelted 50^{th}%ile MHIII dummy. Situations involving other occupants, crash speeds or belt usage conditions may result in tradeoffs to maintain acceptable results for all conditions. Currently, there is considerable interest in adaptive restraint systems that can detect various crash conditions and adjust the restraint system to provide increased levels of protection. There is also a great deal of interest in systems that can detect an out of position occupant and adjust the airbag deployment to lessen the possibility of deployment induced injuries.

The development of restraint hardware has been on-going for some time and testing shows that these devices can provide enhanced occupant protection. The next logical step is to begin to combine these components with appropriate sensing technologies and control systems that understand the implications of the decision making process. The value of the adaptive restraint system has been demonstrated in its ability to adjust for different crash conditions and a degree of real-time decision making capability has been demonstrated in sled testing. Furthermore, significant reduction in airbag deployment induced injuries are made possible by detecting an out of position occupant and altering airbag inflation parameters.

INTRODUCTION

Airbags installed on automobiles sold in the United States are designed to meet Federal crash test requirements (Federal Motor Vehicle Safety Standard 208). Fundamentally, this crash test standard requires automobile manufacturers to provide protection to an unbelted, average sized male occupant in a 30 mph (48 kph) crash into a rigid barrier, by means of an inflatable restraint [1]. Although the crash energy associated with a 30 mph (48 kph) barrier crash is significantly higher than the majority of collisions that occur on US highways [2], the federal standard requires automakers to design systems for relatively severe conditions. Significant among the airbag system design parameters that require tuning for FMVSS 208 are the speed of deployment and the firmness of the airbag. The airbag must be quite firm to dissipate the energy of an unbelted 50^{th}%ile MHIII dummy at the required crash velocity. However, that degree of stiffness may be greater than is necessary for smaller occupants or for occupants involved in crashes of lesser severity. Additionally, an occupant who is restrained by a seatbelt as well as an airbag, will not require an airbag with the same degree of firmness as would an unbelted occupant. The result is that too much restraint energy may be applied to occupants under certain accident conditions.

During a 30 mph (48 kph) barrier crash, the velocity of the occupant movement within the vehicle interior typically requires that the airbag must be substantially full 30 milliseconds after initiation of deployment [3]. This is necessary in order to allow the airbag time to provide adequate space for the absorption the occupant's kinetic energy. This deployment speed dictates the need for very high inflation flow rates that translate into very high forces in close proximity to the airbag. While these forces pose no threat at distances involving normally seated occupants, they may impart injurious forces to occupants who happen to be positioned too closely to the deploying module. Real world crashes have a high degree of randomness. Despite designer's best efforts to predict and test for various crash scenarios, some crash situations may result in an unbelted occupant being too close to the module upon deployment. The problem is particularly acute when small children are involved, since their lower mass increases the possibility of injuries. [4]

Restraint systems of the future will incorporate adjustable components that will be tuned for various factors such as occupant weight, crash severity, seat

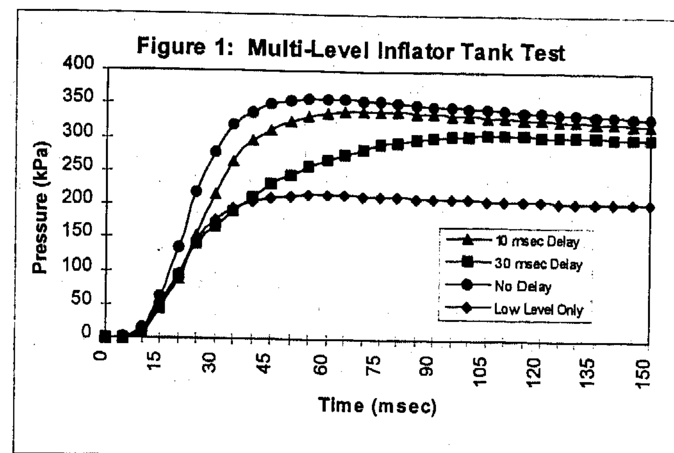
belt usage and occupant position within the passenger compartment [5]. Appropriate sensors must be employed to detect these conditions and a control system must be developed to use the information to produce an appropriate response. This entire assembly of sensors and adjustable restraints can be considered to be part of an advanced "Adaptive Restraint System". This paper presents the state of development of this system and presents the results of testing and analysis.

DESCRIPTION OF PROPOSED SYSTEM

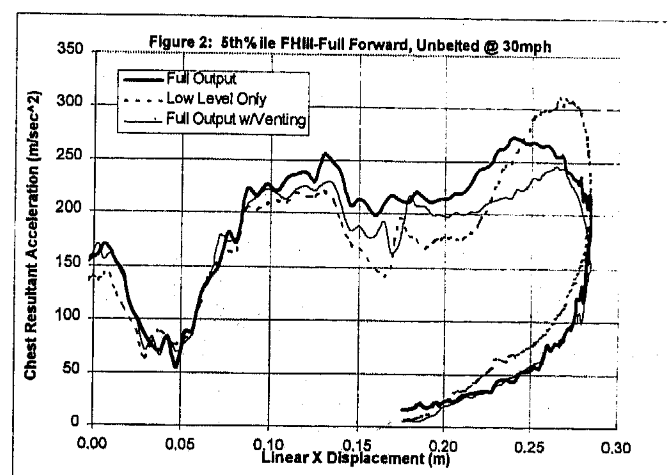
Typical restraint system performance can be improved when one considers the variety of actual crashes that occur on our highways. For example, the vast majority of seatbelts on the road today are simple retractable bands that hold an occupant in place. In a crash, the tension in the belt increases nearly linearly with the amount of stretch imparted to it by the occupant. Thus, the tension in the belt is ever increasing, translating into increasing force levels imparted to the chest of the occupant. Since injuries are related to peak forces and accelerations sustained by occupants, this type of restraint presents possibilities for improvement. "Constant force" seatbelt systems are under development that limit the amount of tension in the belt (and therefore the force placed on the occupant) by allowing compliance at a pre-determined constant force level. Additionally these seat belt systems utilize electrically actuated high-output pre-tensioners to impart a pre-load to the belt to further increase its effectiveness. [6]

The constant-force seatbelt and airbag together create a system that can be finely tuned to provide nearly optimal restraint performance in a given crash situation [6]. However, this still represents a system that is optimized for only one particular crash scenario. Various size occupants require different restraint force levels. A system tuned for an average male at 35 mph (56 kph) will be more firm than is desirable for a smaller, lighter occupant. Similarly, a system tuned for the average person may have too little restraining capability to keep a heavier person from impacting the windshield or instrument panel. The optimal restraint force changes as other factors such as crash severity and belt usage vary. Ideally, restraint systems should be able to adjust for these factors.

To provide optimal restraint, multi-level restraint devices should be capable of adjusting restraint forces, to particular situations. For instance, a constant force retractor is under development that offers three levels of load limiting, adjustable in real time in response to the conditions present at the time of collision. Likewise, two methods will be presented to adjust the restraint force provided by an airbag. The first is a multi-level inflator to alter, not only the volume of gas introduced into the airbag, but also the velocity at which this gas is delivered. This output is varied by the manner in which the inflator's two initiators are actuated. For a low level deployment, only one squib is fired. For a high level deployment, both squibs are fired simultaneously. If an inflation level between the high and low is desired, a delay can be programmed between the actuation of the two squibs to finely tune the output. This is shown in Figure 1. The second method of adjusting an airbag's restraint force is to vary the amount of gas that is vented out of the bag during occupant restraint.



For normally seated occupants, it is desirable to introduce a full airbag early in the event to initiate occupant restraint as soon as possible utilizing all available space for energy absorption. [7]



This suggests that so long as an occupant is not in close proximity to the airbag module, the inflatable restraint should be actuated at high inflation levels. However, the particular conditions present at the time of the crash may dictate that different restraint forces be applied depending upon those conditions. In order to adjust the restraint force provided by an airbag for normally seated occupants, the amount of gas vented out of the airbag will be adjusted by means of vent openings in the airbag housing. These openings are electrically actuated at the time of the crash to control airbag stiffness during the crash event. Figure 2 illustrates the chest acceleration levels as a result of adjusting both the venting characteristics and inflator output as predicted by

MADYMO simulation. A study was set up with a PAB system designed to FMVSS 208 requirements, except that a 5th%ile FHIII replaced the 50th%ile MHIII in our model. As can be seen from the figure, if the inflation level is lessened in response to the lower mass of the occupant, the peak accelerations actually increase. When the inflator output is left unchanged and the venting is increased, the chest accelerations are decreased as compared to the baseline system.

This venting system cannot replace a multi-level inflator. For occupants who might otherwise be injured by the high deployment speed of an airbag, the multi-level inflator's ability to decrease flow rates can have a significant effect on injury criteria. Figure 3a-c shows the results of static out of position testing with a 5th%ile FHIII at distances of 25, 150 and 250 mm from the PAB module. The tests were run with an FMVSS 208 inflator and also with a low level inflator to quantify improvements in injury criteria. As can be seen from the figure, reducing inflator output may significantly decrease the potential for deployment-related injuries.

Figure 3a: HIC

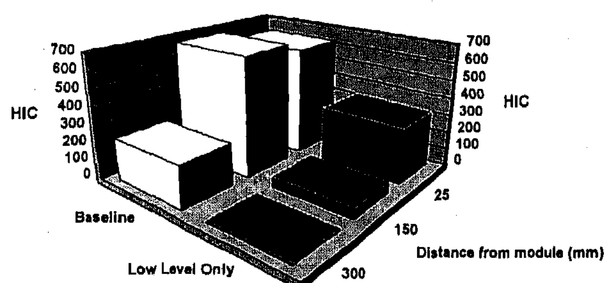


Figure 3b: Chest g's

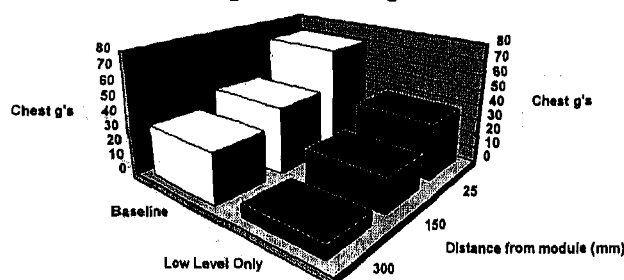
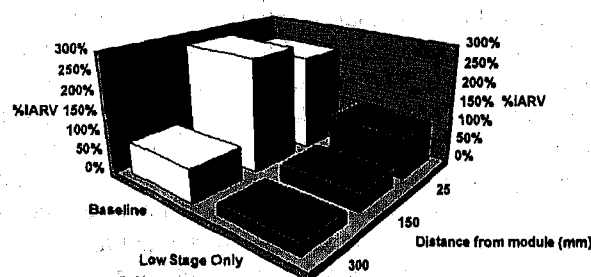


Figure 3c: %IARV-Neck Tension



Figures 3a-c: Static Test Results For a 5%ile FHIII

In summary, a sensor system will need to detect the following inputs:

1) Seat belt usage

2) Occupant proximity to airbag

3) Occupant weight

4) Crash severity

A control system in turn will take these sensor inputs and utilize them to adjust:

1) Airbag fill velocity

2) Gas volume introduced into the airbag

3) Airbag venting area

4) Seat belt load limiting.

Accurate sensing of these various factors is critical for an effective system. The following sensors are being developed to control the adaptive restraint system presented.

Occupant Position - The proximity of an occupant to the airbag module is a very important factor in determining correct airbag deployment forces. An electric field occupant presence and proximity system is under development. This system consists of a series of copper foil sheets embedded into the instrument panel and oriented in such a way as to act as a capacitor. Rather than a traditional capacitor, however, where the electric field flows between two plates, the orientation of these plates acts to project an electric field outward, toward the occupant. Since humans have enough moisture in their bodies to act as a dielectric, the presence of a person in the electric field will change the capacitance of the sensor. The magnitude of change in capacitance varies with the proximity of the person to the sensor. Therefore, measuring the change in capacitance can determine the proximity of an occupant to the module. A particular benefit to electric field technology is its ability to discriminate between humans and most inanimate objects which are poor dielectrics.

The electric field also can penetrate objects such as newspapers that may block ultrasonic or infrared beams utilized in other occupant sensors. Perhaps the most important benefit to the electric field occupant sensor under development is the system's ability to react in under 2 milliseconds. This response time is fast enough to allow the sensor to detect an occupant being thrown forward in the passenger compartment due to panic braking or a long crash pulse/late fire time event.

Occupant Weight - Several companies have weight classification sensors under development. Typically,

these sensors are positioned in the seat cushion. For purposes of this analysis and the physical sled testing presented in a later section of this paper, four load cells were positioned at the seat mounting locations to measure the weight of the occupant in the seat.

Seatbelt Usage - Seatbelt buckle switches have long been used to determine if the belt has been fastened and to actuate an audible reminder. While these switches can be used during the early development of adaptive restraint systems, they are not adequate for production level adaptive restraint systems. Hall Effect sensors will be necessary to provide adequate reliability levels when used as an integral part of a production level adaptive restraint system.

Crash Severity - A current, state of the art single point sensor system includes accelerometers inside the electronic control unit that measure the reaction of the passenger compartment and a microprocessor to analyze these inputs and make the fire/no-fire decision. Virtually all major single point sensor manufacturers are developing systems that will not only determine if an airbag is needed, but also will be able to discriminate the level of severity of the crash.

Integrating all of these variables is a fairly complex problem. Control of each restraint factor depends upon all of the sensed conditions of crash severity, belt usage, occupant weight and position. Moreover, those factors are also interdependent upon each other. For example, the amount of seat belt load limiting required is dependent upon the stiffness of the airbag. A computer algorithm was developed to analyze these factors and determine the appropriate settings for adjustable restraint components.

COMPUTER MODELING:

MADYMO crash simulation was used to develop appropriate levels to which adjustable restraint factors should be tuned for various crash conditions. A model was assembled to simulate a passenger in a generic vehicle environment and a generic pulse. The model was correlated to a 35 mph (56 kph) crash with a belted occupant using the results of HYGE sled testing. The following factors were modeled:

Occupant Sizes:

5th%ile FHIII, 50th%ile MHIII, and 95th%ile MHIII

Restraint system:

Airbag/Seatbelt with Pretensioner and Airbag only

Occupant Position:

Full Forward, Mid-position, and Full-Forward

Crash severity:

14 mph (22 kph), 18 mph, 30 mph (48 kph), 35 mph (56 kph)

For the purposes of this study, crash severity was defined as impact velocity into a rigid barrier at 0°. Simulations were run with various combinations of the above factors and several combinations of airbag venting and seatbelt load limiting. For the initial study, only normally seated occupants were investigated. Previous analysis determined that regulating airbag venting characteristics was more effective than regulating inflator flow characteristics when adjusting restraint forces for normally seated occupants. Consequently, only seatbelt load limiting and venting were varied. Table 1 illustrates the results of the modeling for one of the cases, (50th%ile MHIII, mid-position at 30 mph (48 kph) barrier crash). This process was repeated for other cases until enough information was available to determine optimal restraint characteristics for the anticipated situations. That information, in turn, was used to produce a control algorithm.

Table 1:

Run No.	Bag Porosity	Load Limiter, kN	HIC	3MS - Chest G's	Chest Def. (mm)
1	0xStd. Por.	0.9	727	45.55	35.8
2	0xStd. Por.	1.3	706.1	46.67	35.6
3	0xStd. Por.	1.8	659.8	41.91	35.6
4	0xStd. Por.	2.2	661.8	45.78	36.8
5	0xStd. Por.	2.7	688.8	42.51	36.3
6	1xStd. Por.	0.9	484.4	40.06	26.7
7	1xStd. Por.	1.3	490	39.03	27.7
8	1xStd. Por.	1.8	476.1	39.71	28.2
9	1xStd. Por.	2.2	482.3	41.91	28.0
10	1xStd. Por.	2.7	494.8	40.84	28.6
11	2xStd. por.	0.9	348.9	31.22	26.0
12	2xStd. por.	1.3	383.4	29.15	25.1
13	2xStd. por.	1.8	401.6	31.19	24.6
14	2xStd. por.	2.2	411	32.75	26.4
15	2xStd. por.	2.7	399.2	36.36	26.0
16	3xStd. Por.	0.9	409.6	29.27	26.0
17	3xStd. Por.	1.3	379.6	28.02	26.7
18	3xStd. Por.	1.8	432.9	31.45	25.4
19	3xStd. Por.	2.2	429	30.64	25.0
20	3xStd. Por.	2.7	422.8	32.86	25.7
21	4xStd. Por.	0.9	416.5	29.85	26.9
22	4xStd. Por.	1.3	418.3	30.89	26.1
23	4xStd. Por.	1.8	429.7	31.37	24.8
24	4xStd. Por.	2.2	434.6	33.22	24.1
25	4xStd. Por.	2.7	400.8	36.23	25.3

Table 1: Modeling Results For a 50%ile MHIII-Midposition using a 4-inch Pretensioner at a 30 mph Barrier Crash

CONTROL ALGORITHM:

The entire system that can sense various aspects of a crash and tune the restraint system accordingly is referred to as an "Adaptive Restraint System". A system schematic is presented in Figure 4. Once the various inputs are known, there must be a decision making process to determine appropriate settings for the restraint devices. If only one input is taken into account, occupant weight for example, the restraint system tailoring can be a simple linear adjustment from low-level for a light occupant, to high level for heavy occupants.

However, taking several factors into account increases the complexity of the decision making process.

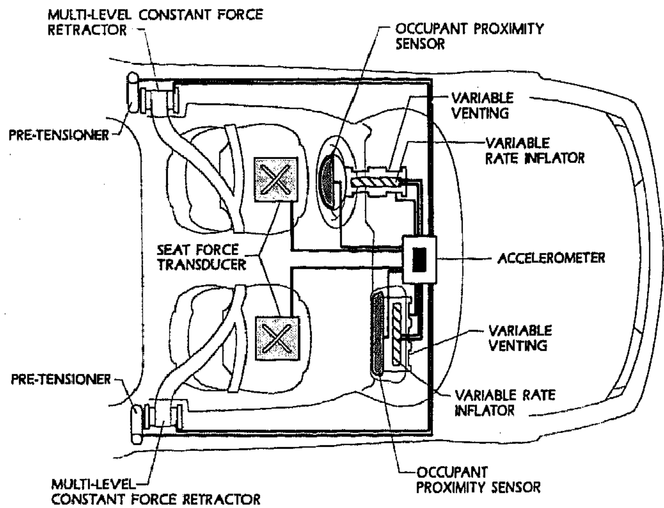


Figure 4: System Schematic

The first step in validating the algorithm was to pick various crash scenarios with different factors as defined. Those conditions were then input into a computer program with the algorithm to see what restraint adjustment levels would be chosen. MADYMO simulations were performed on each scenario to compare predicted injury criteria utilizing restraint settings recommended by the algorithm with predicted injury criteria utilizing a baseline system. In this case, the baseline system was a current, restraint system that included a single level load limiting seat belt system and airbag. This number will be used in a mathematical formula to set levels of the adaptive restraint components. See Table 2 for one possible configuration of these factors.

Table 2:

Factor	Magnitude
A-Crash severity	1-mild crash, 10-severe crash
B-Occupant Mass	1-light person, 10-heavy
C-Seatbelt Use	1-belted, 2-unbelted
D-Occupant to IP proximity	0-touching, 10-normally seated

The entire system would be controlled by an electronic module that monitors all relevant factors such as those in Table 2 and chooses the appropriate airbag and seatbelt settings. The first decision to be made is whether the crash is severe enough to fire the pretensioner and set the load limiting level of the constant force retractor. The second decision is whether the crash is severe enough to fire the airbag. At that point, all the factors must be taken into account and the control system must manage the adjustment of the restraint system. An algorithm based on a set of mathematical equations was derived to calculate restraint adjustment levels. This algorithm operates on the assumption that there is a condition where all factors are maximized, calling for the maximum restraint force to be applied. For the airbag inflator, the highest output occurs when the occupant weight, the occupant distance from the airbag, and crash severity are at maximum levels and the occupant is not wearing a seat belt. For airbag venting, the maximum restraint force (and therefore the minimum venting) occurs when the occupant weight, the occupant distance from the airbag, and crash severity are at maximum levels and the occupant is not wearing a seat belt. For the constant force retractor, the maximum restraint force occurs when the occupant weight and the crash severity are at maximum levels but occupant distance from the airbag is minimum.

A separate control equation is used for each of the restraint components. The equation compares the state of the actual input factor such as crash severity, occupant weight, etc. to a maximum value and this difference is used to modify the restraint system from the maximum. Weighting factors are then assigned to each input, giving more weight to those factors that will have a greater influence on the particular restraint component setting. For example, a large weighting factor is applied to the occupant proximity sensor signal when it is used to control airbag inflation. This results in a significant decrease in airbag inflation velocity which, in turn, should lessen the possibility of deployment-related injuries.

The equation approach allows for a continuous response to the input parameters as opposed to a look up table that has been proposed for these systems. The MADYMO study presented earlier was used to determine appropriate weighting factors for each restraint component. The control of this system is based on the assignment of a number (from 1 to 10, for example) to each of the inputs to represent its magnitude. As illustrated in Table 3, there was significant improvement in injury criteria for all of the crash situations investigated.

Table 3: MADYMO Data

					Improved		Baseline	
Run No.	Occupant	Belt	Load Limit (kN)	Vent (mm^2)	HIC	Chest g's	HIC	Chest g's
1	5%ile HIII	Y	1.5	+4961	47	22.90	74	20.93
2	95%ile HIII	Y	1.5	+4961	28.5	15.16	116.5	10.98
3	95%ile HIII	Y	4	+1936	956.2	85.25	1304.4	43.14
4	5%ile HIII	Y	1.5	+4961	222.9	55.78	622.3	43.40
5	50%ile HIII	N	NA	+1936	653	48.83	652.5	41.37
6	50%ile HIII	N	NA	+4961	12.6	23.20	77.3	11.55
7	50%ile HIII	Y	1.5	+3025	354.8	4.27	482	34.14
8	50%ile HIII	Y	1.5	+4961	25.5	1.69	71.5	14.20
9	50%ile HIII	Y	2.5	+3025	572.5	5.23	794	47.29
10	5%ile HIII	Y	1.5	+4961	99.5	4.76	248	37.68
11	95%ile HIII	Y	4	+3025	411.4	5.83	738	30.87
12	95%ile HIII	N	NA	0	1339	7.41	1339	7.41
13	5%ile HIII	N	NA	+3025	191.9	73.86	415.9	55.81
14	95%ile HIII	N	NA	+3025	90	22.87	117	18.56
15	5%ile HIII	N	NA	+4961	32.3	25.79	68.4	19.22

*-Belted cases include 2.2kN load limiter on torso belt.

Table 3: Adaptive Restraint Modeling

SLED TEST VALIDATION

A sled test was performed to further validate the adaptive restraint concept. The main purpose of this testing was to determine if the various conditions could be sensed and the systems actuated, all in the time frame at which a crash event occurs. This particular test condition was designed to determine if the system could recognize an occupant that was thrown out of position due to the crash forces in a late fire event and respond by lowering the output of the inflator.

The electric field occupant sensor has been shown to have an extremely rapid response time. Figure 5 illustrates the output of the electric field occupant sensor during a previous sled test. In this case, the output was not used for decision making, but was simply monitored to compare with the actual dummy trajectory. As can be seen by Figure 5, the occupant sensor tracked the head location quite closely, once the head came within the range of the sensor. It should be noted that this test was performed with an earlier version of the sensor having less range than current versions of the device.

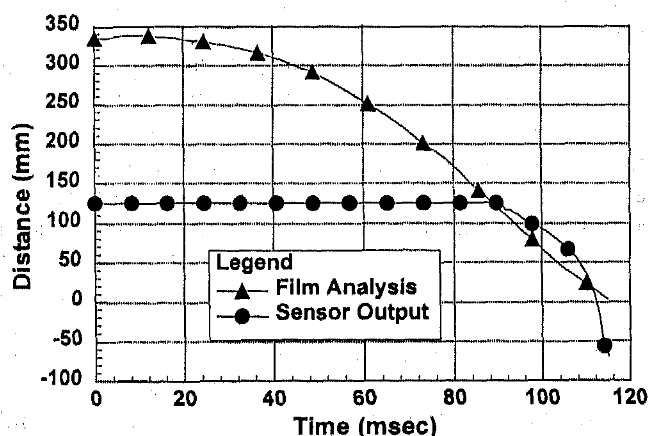
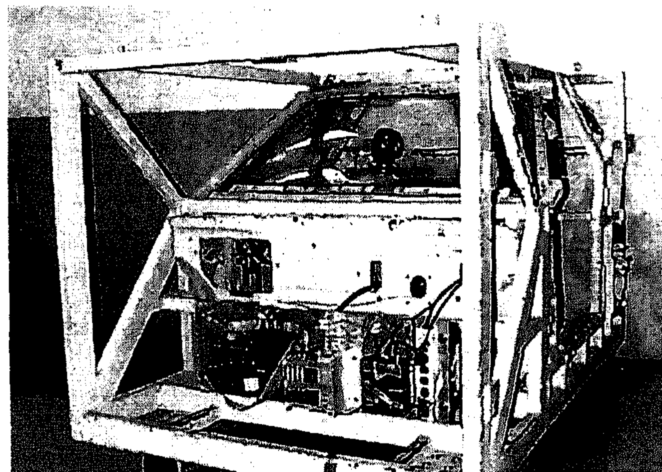


Figure 5: Comparison of Occupant Position for Capacitive Occupant Sensor vs. Film Analysis (22.4 km/hr velocity hyge sled test.)

The simulated crash test was conducted on a 9-inch HYGE sled. The generic test fixture, or "buck," consisted of sheet metal reinforced with rectangular steel tube. A 5th%ile FHIII Anthropomorphic Test Dummy (ATD) was situated without a safety belt in the right front passenger seat in a full-forward position. The airbag module consisted of a multi-level inflator, nylon airbag, steel housing with variable venting capabilities and an electric field occupant sensor located within the airbag door. The forward sensing range of the occupant sensor was approximately 14 inches (356 mm). Therefore, the nose of the ATD was positioned slightly beyond that distance so that the occupant would not be detected until it moved into range of the sensor during the simulated crash event. The control electronics were mounted on the sled buck and included a microprocessor to run the control algorithm. During the event, the control electronics would determine airbag firing time and vary the output of the dual-level inflator based on the input from the occupant sensor. The control system also controlled actuation of the airbag variable venting system, although the tailoring of the inflation rate was expected to have the greatest effect on airbag performance in this out of position test. The sled buck with microprocessor and related components are shown in Figure 6.


Figure 6: Test Setup

The Sled was fired at 20 mph (32.2 kph) with a pulse that represented an angular collision. This pulse was chosen because of its relatively late fire time. To ensure that the occupant was quite close to the airbag during deployment, the fire time was further increased to 97 msec.

No external signals were sent to the sled buck for control of the restraint devices. The output of the sled carriage accelerometer was routed to the adaptive restraint control electronics which used that signal to determine the initiation of the crash event. The initial fire time of the first squib was programmed into the microprocessor prior to the test, but the second squib fire time, as well as the decision to actuate the venting devices, were all controlled by the algorithm. The timing of the control system's decision-making process for this test is shown in Figure 7. As can be seen, the algorithm chose a delay time of 28.3 msec between the firing of the first and second squibs based on the proximity of the occupant's head to airbag door. Because of the relatively low weight of the occupant and moderate crash severity, the system chose to open all of the airbag vents as well.

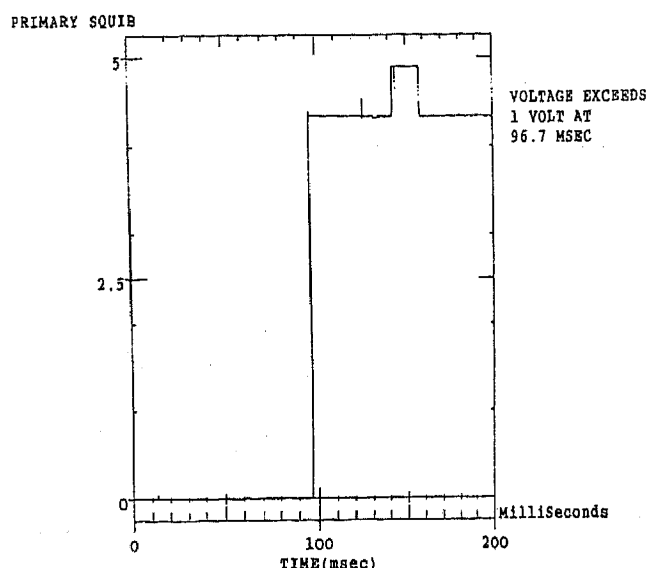


Figure 7a: Primary Squib

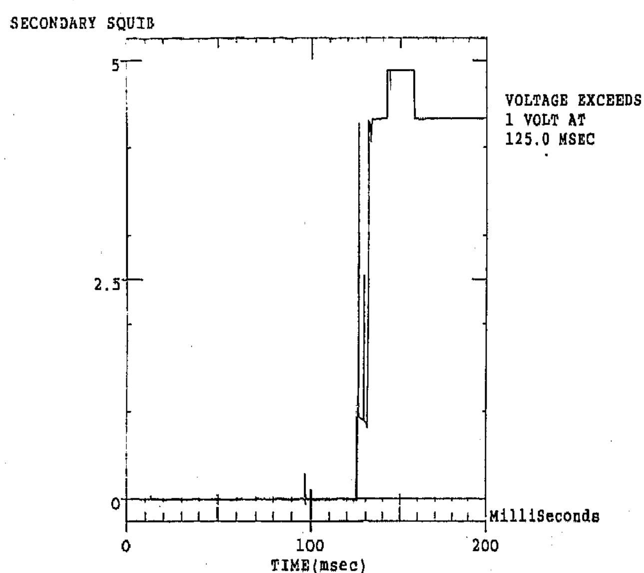


Figure 7b: Secondary Squib

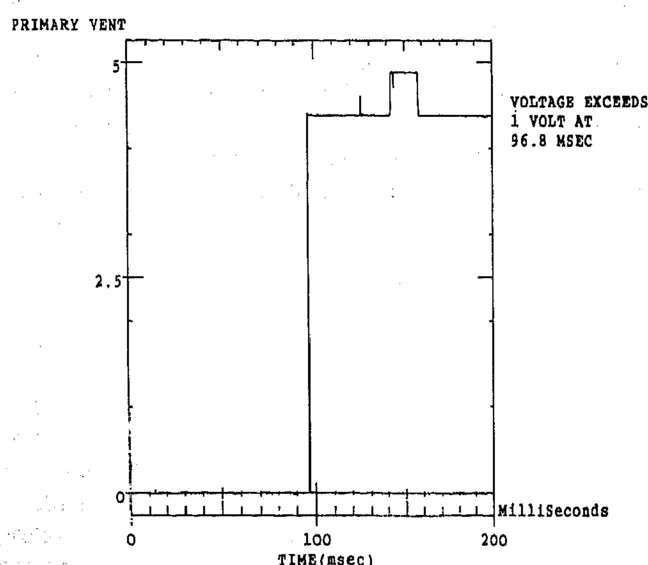


Figure 7c: Primary Vent

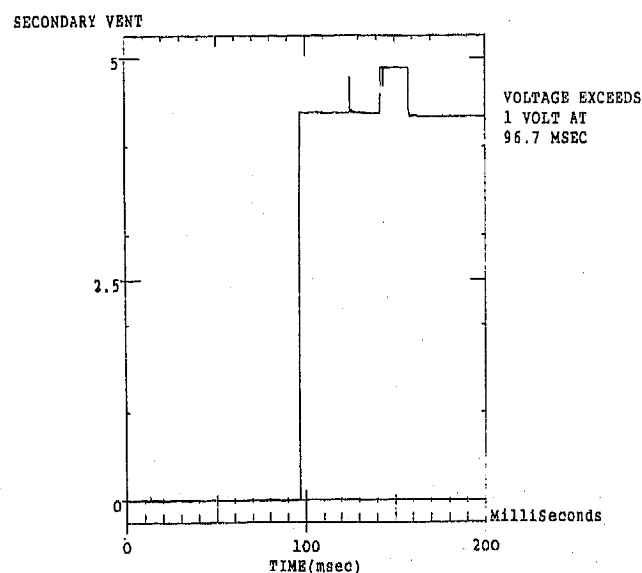


Figure 7d: Secondary Vent

CONCLUSION

Although much work remains to be done in the area of adaptive restraints, initial efforts appear to be quite promising. The value of the adaptive restraint system concept has been demonstrated through computer simulation. Initial sled tests show that the electronics can react quickly enough to be effective. Considerable work remains to test and tune all of the parameters to specific vehicles. Given all of the variables involved with real world conditions, and the limits of presently-available sensing technology, it is probably not reasonable to expect precise restraint system optimization for all of the situations which are likely to be encountered. It is the opinion of the authors that the goal of the adaptive system should be simply to demonstrate significant performance improvements over current, non-adjustable systems. Future work will involve a substantial number of sled tests to demonstrate the benefits of such a system in many more situations, and ultimately, the application of this system to an actual production vehicle.

ACKNOWLEDGMENTS

The authors would like to express their gratitude and appreciation to Shari Lyn, Kiran Sambutar and Jeff Sikorski for their contributions in the preparation of this paper.

REFERENCES

1. "Code of Federal Regulations", Title 49, Part 571.208, Section S5 (Occupant Protection), October 1995,

2. "National Accident Sampling System - 1995 Crashworthiness Data System", U.S. Department of Transportation, National Highway Traffic Safety Administration.

3. Goch, S., Krause, T., Gillespie, A., "Inflatable Restraint System Design Considerations", SAE 901122, P-233, Society of Automotive Engineers, Warrendale, PA, 1990

4. Winston, F.K., Reed, R., "Air Bags and Children: Results of a National Highway Traffic Safety Administration Special Investigation into Actual Crashes", SAE 962438, P-305, 40th Stapp Car Crash Conference Proceedings. November, 1996

5. Miller, H.J., Patrishkoff, D., Maripudi, V., "Injury Reduction with Intelligent Restraint Systems", Proceedings from "Bag and Belt '96", 4th International Akzo Nobel Symposium on Automotive Occupant Restraint Systems, April 1996.

6. Miller, H.J., "Occupant Performance with Constant Force Restraint Systems", SAE 960662, Society of Automotive Engineers, Warrendale, PA, 1996

7. Bonello, K. "Occupant Energy Management Technique for Restraint System Analysis and Design - Understanding the Physics of the System", SAE 922083.

971055

Evaluation of Angular Displacement Measurement Techniques for Tracking the Motion of Anthropomorphic Test Devices

Jacqueline Marshall
Ford Motor Co.

Dennis A. Guenther
The Ohio State Univ.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The measurement of angular rotation has many applications in crash testing, particularly in tracking the motion of crash dummies. There are currently a few devices for determining angular rotation. These include accelerometer arrays, magnetohydrodynamic (MHD) sensors, potentiometers, and high speed films. However, there are problems associated with all of these methods. Systron Donner has developed a new device called a "Quartz Rate Sensor" or "QRS". The QRS utilizes a piezoelectric chip which produces a DC voltage proportional to the rate of rotation of the sensor about its sensitive axis. Angular displacement can then be determined from a simple integration. Results of preliminary tests performed at The U.S. Department of Transportation's Vehicle Research and Test Center suggest that the QRS's yield very accurate results.

BACKGROUND

The need to measure the angular displacement of an ATD has become increasingly vital to injury assessment. Traditional injury assessment techniques tend to rely on translational head acceleration, engineering quantity effectively measurable by translational accelerometers. However, modern injury assessment techniques are based more on angular quantities rather than translational, and the instrumentation required to measure these quantities is not as readily available.

A study was conducted by Viano and Arepally (16) to consider the adequacy of injury criteria and tolerance levels specified by FMVSS 208, the standard that measures the safety performance of restraint systems and supplements. They suggested that additional criteria be added to the standard to represent fundamentally important aspects of a vehicle occupant's kinematic response. These additional criteria include: the change in torso angle (used to evaluate injury to the rib cage), rotation and rebound of the chest and pelvis (necessary to study the effects of submarining), and the bending of the neck.

Submarining occurs when the pelvis of a belted passenger slides underneath the lap belt, causing the load of the impact to fall across the abdomen instead of the hips. This phenomena, which may result in serious injury, is not accounted for in traditional injury assessment techniques. Both Freeman and Bacon (4) and Viano and Arepally (16) agree that monitoring pelvis angle may provide a means of quantifying submarining. Freeman and Bacon, using a three-dimensional high speed video system to measure occupant motion, performed extensive tests to evaluate how changes in seat belt anchorage and seat adjustments affected whether or not submarining occurred.

The measurement of angular displacement is also necessary for predicting neck injury. The widely used Mertz and Patrick neck injury criteria relates neck injury to the bending moment of the head developed about the occipital condyle and the angle of the neck at the first thoracic vertebra. This quantity could be evaluated using an existing upper neck load cell along with angular displacement data. An additional neck injury assessment technique, created by Gadd, relies on the angle between vertebra to determine an injury threshold. This criterion could be evaluated directly with angular displacement measurement (12).

A new ATD device, the Cervical Omni-Directional Bending Apparatus (COBRA), was recently evaluated at the Vehicle Research and Test Center. The COBRA, which fits inside the neck of an ATD, is designed to characterize the shape response of the neck to an impact. The angle of the head relative to the top of the neck, however, must be determined through additional angular motion sensors to completely define the motion of the head/neck complex.

An additional potential need for accurate angular displacement data may be for brain injury evaluation. Brain injury is traditionally predicted from head acceleration data. Many researchers believe that injury to brain tissue is more closely related to the relative displacement between the brain and the skull than it is to head acceleration. A brain compliance model is under investigation at General Motors to assess brain injury risks in the Hybrid III ATD. This model consists of a lumped-mass model of a brain attached to the skull with a series of springs and dampers (15). If this brain

compliance model becomes available, the displacement of the brain mass and the skull could be compared to develop a new brain injury criterion.

While the need for measuring angular displacement accurately has been well established, the means for doing so has not. There are currently a few devices used to determine angular displacement. These include accelerometer arrays, high speed films, potentiometers, and magnetohydrodynamic (MHD) sensors. All of these methods are evaluated in (7). A new angular velocity sensor manufactured by Systron Donner, the "Quartz Rate Sensor", is included in this paper.

Perhaps the most widely used method of determining angular displacement is the linear accelerometer array. These exist in a variety of configurations which differ mainly in the placement and number of redundant accelerometers used. At least nine linear accelerometers must be used to obtain angular motion in three dimensions. Viano (15) cites several potential causes of error in the angular acceleration data. These include transverse (cross-axis) sensitivity, mismatch of accelerometer pairs, misalignment of accelerometer pairs, signal noise, and numerical analysis. When angular acceleration is integrated to obtain angular velocity, then integrated again to obtain angular displacement, the error is compounded.

Angular accelerometers do exist, but they are much less common than the linear accelerometers. The angular devices are seldom used because they are notorious for producing very noisy signals.

High speed film or video provides another means of measuring angular displacement. This system uses a video camera or cameras capable of high resolution at minuscule time intervals to track specific targets. The film is then digitized to obtain the displacement of the targets. A primary problem with the use of film occurs when the camera's view of the moving target is obstructed by flailing dummy arms, air bags, or the automotive structure. A second problem occurs when the target is not perpendicular to the line of sight of the camera.

Rotary potentiometers provide a direct measurement of angular displacement, and are generally believed to be very accurate. NHTSA commonly uses angular pots for calibrating ATD's. The primary deficiencies of potentiometers are their size and mounting restrictions. The pots can not be mounted inside a dummy because they must be rigidly attached to two members experiencing relative displacement. Pots mounted on the outside of an ATD can not be used in most crash tests because they will be damaged by any direct impact to the sensor. In addition, the rotary pot's ability to withstand deflection at high accelerations is limited.

A relatively new angular velocity sensor, the MHD, has been used in recent years. Several organizations, including the NHTSA, the Naval Biodynamics Lab, Systems Research Laboratories and the Biodynamics and Biocommunications Division at Wright Patterson Air Force Base have evaluated the sensor in detail. While the sensor's packaging and post-processing requirements seem appealing, all evaluations have shown considerable error in many instances which has not yet been explained. In addition, many researchers express uncertainty over the post-processing requirements.

The QRS relies on piezoelectricity to determine angular velocity. A block diagram of the QRS is illustrated in Figure 1. The double-ended tuning fork and its supporting structure are all fabricated chemically from a single wafer of monocrystalline piezoelectric quartz. Two drive tines are

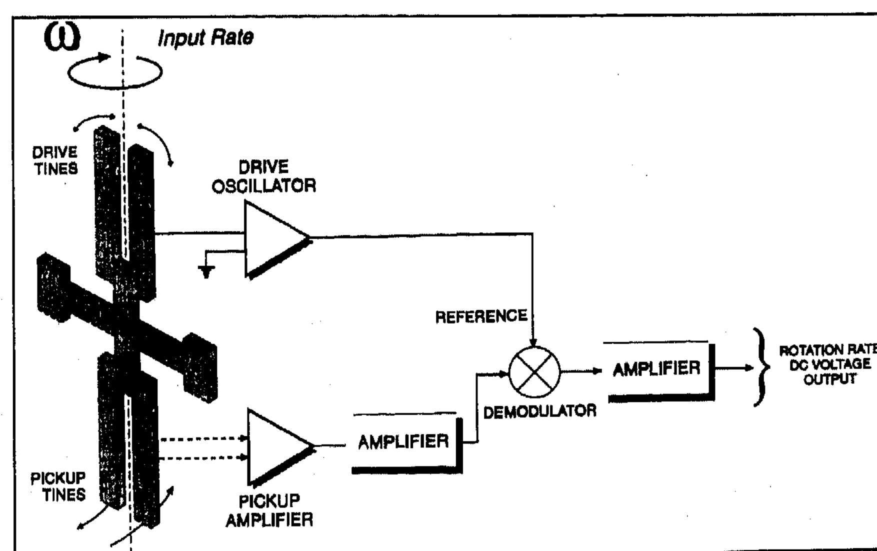


Figure 1 Quartz Rate Sensor Block Diagram (modified from Systron Donner, 1994)

driven at a precise amplitude by an oscillator circuit. This amplitude causes the tines to move toward and away from each other at a high frequency. Each tine has a Coriolis force acting on it of:

$$\overline{F} = 2m\overline{\omega_i} \times \overline{V_r} \tag{1}$$

where m is the mass of the tine, $\overline{\omega_i}$ is the input angular rate, and $\overline{V_r}$ is the instantaneous radial velocity.

The Coriolis forces resulting from the drive tines act perpendicular to the plane of the fork assembly in opposite directions. This produces an oscillating torque proportional to the input rotational rate. The pickup tines respond to the oscillating torque by moving towards and away from the plane of the fork, producing a signal at the pickup amplifier. This signal is amplified and then demodulated into a DC signal, which is proportional to the angular velocity of the sensor.

The use of a single piece of piezoelectric quartz simplifies the sensor, resulting in excellent stability. The QRS will produce an output signal only in response to angular motion about the tuning fork's axis of symmetry. This results in a very low cross-axis sensitivity. Systron Donner reports that the device has no cross-axis sensitivity.

METHODS

Test Procedure

Low-frequency testing was performed using the pendulum impact simulator at VRTC. The pendulum, shown in Figure 2, is typically fitted with a head and neck at its base. The pendulum can then be raised to specific heights and dropped. As the pendulum reaches the vertical position, it impacts a fixed block of aluminum honeycomb. The impact of the pendulum with the honeycomb simulates a step input similar to that of an actual crash. The honeycomb also acts to reduce rebounding and vibration of the pendulum upon impact.

Measurement techniques evaluated during this test include the QRS, the MHD sensor, and a 2-accelerometer array. The dummy head/neck configuration typically used on the end of the pendulum was replaced by a steel block. The block was carefully machined to allow proper positioning of each sensor. After the sensors were attached, the block was rigidly mounted to the base of the pendulum.

The QRS sensors were first secured on a T-shaped QRS tri-axial mounting block recommended by Systron Donner. The mounting block was then bolted to the bottom of the test fixture block so that each sensor's sensitive axis was correctly aligned. Only the data from the sensor measuring velocity about the y-axis were used in this test, with the y-axis being lateral, the z-axis vertical and the x-axis longitudinal. The QRS's measuring velocity about the x-axis was included to study cross-axis sensitivity.

The MHD used in this test was an ARS-09 model in the tri-axial cube configuration. This cube is designed to be directly mounted to the flat surface of the block. Four holes were carefully machined to provide proper alignment. The tri-axial cube contains three sensors such that it is capable of measuring velocity about all three axes. However, the velocity about the y-axis was the only that sensor that was wired.

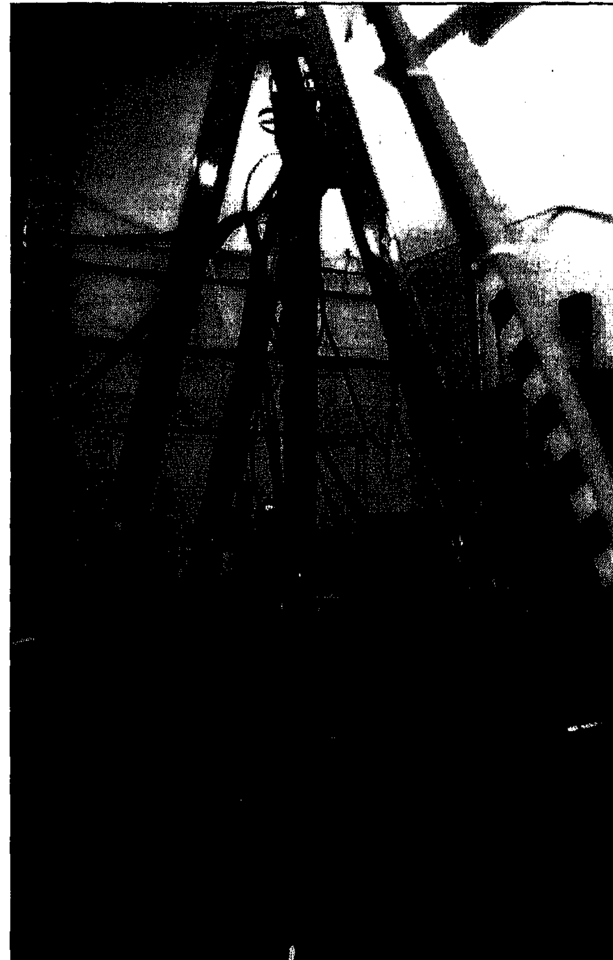


Figure 2 Pendulum Impact Simulator

Test Results

Displacement data from the linear accelerometers produced very poor results. It was determined that the operating range of the sensors was too high for the conditions experienced in this testing series, and the measured signals were of the same order of magnitude as the noise. Although accelerometers with more appropriate specifications are available on the market, they were not available at VRTC at the time the testing took place. The item of primary interest in most crash tests is the peak value of the quantity being measured, whether it be acceleration, velocity, force or displacement. Because the peak displacement values are known for this particular test series, it was not necessary to calculate theoretically exact angular displacement curves. The initial and final values and the general shape of the curves are known. The angular position curve should start

from zero degrees, increase parabolically, and reach the drop angle at impact.

For clarity, one representative displacement plot for each drop angle is included in this section. The plots obtained from the full test series are included in Reference 7. Figure 3 illustrates the displacement obtained from the QRS and MHD sensor when the pendulum was released from a 30° incline. Displacement at drop angles of 75°, 90°, and 110° are given in Reference 7.

The results from the QRS appear superior to the results from the MHD sensor. In each case, the QRS data exhibit the expected shape for the drop. The curve shows that the pendulum bounces back slightly at impact and rests near the expected angle. The MHD data exhibit a somewhat similar shape, but consistently exhibit displacement data that is higher than the actual value.

The peak displacement measured by both sensors is compiled in Table 1 for all of the tests run, along with the

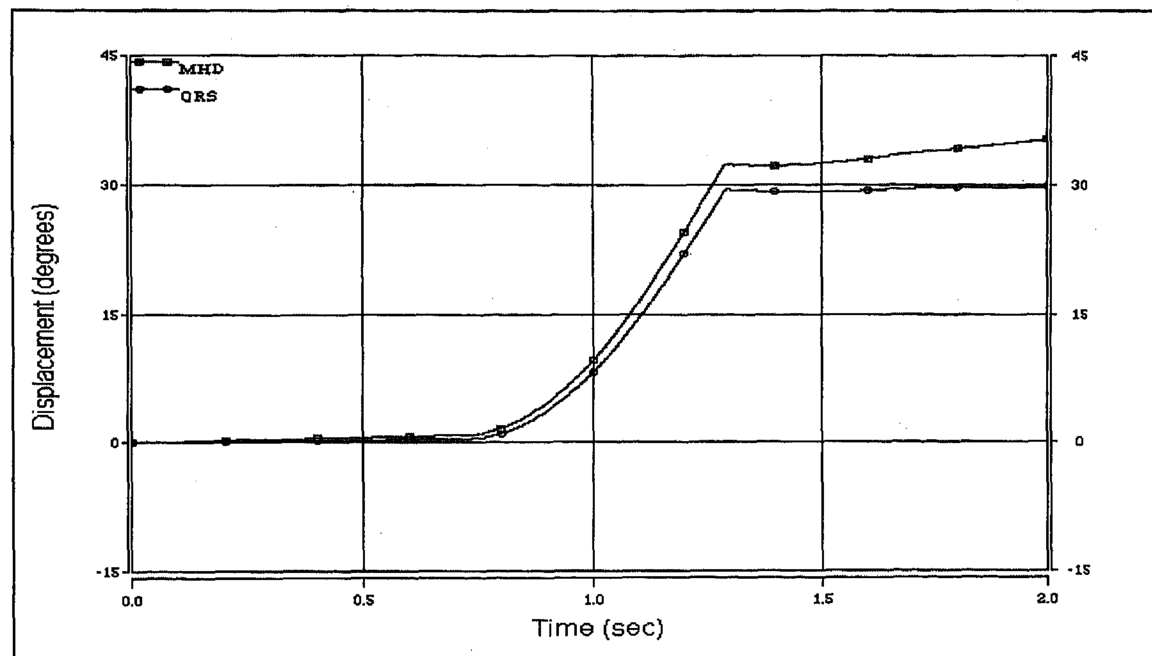


Figure 3 Angular Rotation Measurements for 30° Pendulum Drop

Table 1 Peak Displacement Error for Low-Frequency Test Series

Test Number	Pendulum Displacement (deg)	Peak QRS Displacement (deg)	Error (%)	Peak MHD Displacement (deg)	Error (%)
135	30	29.45	1.83	32.34	7.81
136	30	29.72	0.95	32.57	8.57
137	30	29.79	0.71	32.65	8.82
147	75	76.13	1.50	83.04	10.71
148	75	76.13	1.50	81.93	9.25
149	75	76.48	1.98	81.73	8.98
141	90	89.38	0.69	97.05	7.83
142	90	89.35	0.72	94.83	5.37
143	90	89.14	0.95	95.47	6.08
144	110	110.58	0.53	117.82	7.10
145	110	110.91	0.83	118.88	8.07
146	110	110.49	0.45	119.04	8.22

error as compared to the expected result. The table shows that the QRS exhibited an error of less than 2% for all of the tests run. The MHD, however, produced errors no better than 5.37%, and as high as 10.71%.

Discussion of Low-Frequency Test Results

The results of this test series indicate that the QRS is a viable alternative to traditional angular measurement sensors in the low-frequency environment. The calculated error of between 0.45% and 1.98% is acceptable for the vast majority of crash test applications. Additionally, the post-processing required to achieve displacement from the raw data was much simpler than that required for the other two methods evaluated here.

One disadvantage of the QRS is its size. It is much larger than the accelerometers tested, and slightly larger than the MHD sensors tested. The T-shaped block required to mount the sensor adds to the bulk of the instrumentation. For most applications, the QRS and its mounting kit are sufficiently compact. However, some applications may have space or weight limitations that may prevent its use. Figure 4 illustrates a size comparison between the tri-axial QRS, MHD, and accelerometer instrumentation. Note that the QRS and MHD configurations shown here are sufficient to measure rotation in three dimensions. Three additional accelerometer cubes, with a total of six additional accelerometers, would be necessary for three-dimensional calculations.

The MHD sensor did not achieve impressive results in this test series. Large error values were calculated for the entire range of drop angles. An additional disadvantage is the cumbersome post-processing required by the compensating routine.

Head/Neck Pendulum Testing Without Forehead Impact

A Hybrid III head and neck were rigidly attached, upside down, to the end of the pendulum arm. Figure 5 illustrates this configuration. The pendulum is released from a preset drop angle and impacts the aluminum honeycomb when vertical. As the pendulum arm impacts the honeycomb, the head and neck rotate freely about the base of the pendulum.

Measurement techniques evaluated during this series include the QRS, the MHD sensor, an accelerometer array, rotary potentiometers, and high-speed video. Figure 6 illustrates the mounting of the head instrumentation. This special edition Hybrid III head is designed to accommodate four precisely machined and positioned accelerometer mounting blocks which allow easy instrumentation of a nine-accelerometer array. Endevco 7264 linear accelerometers were mounted at the cg on the x- and z-axes, and at the top of the head on the x-axis. Also included in the head was a GSE three-axis neck transducer, designed specifically to fit in the base of the Hybrid III skull. This transducer measures

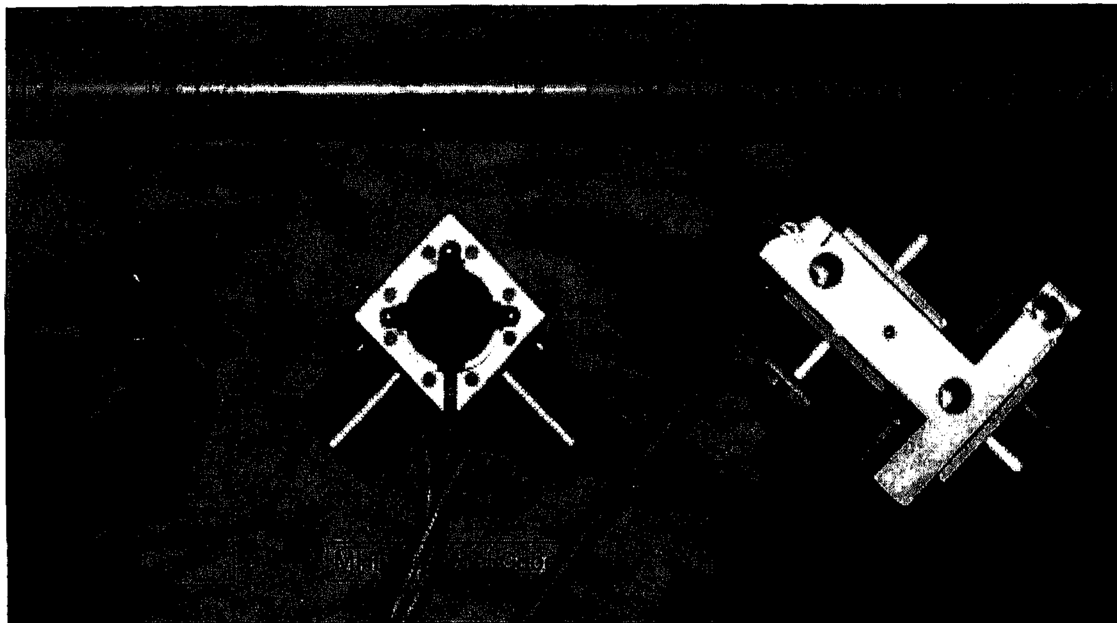


Figure 4 Size Comparison of Various Sensors

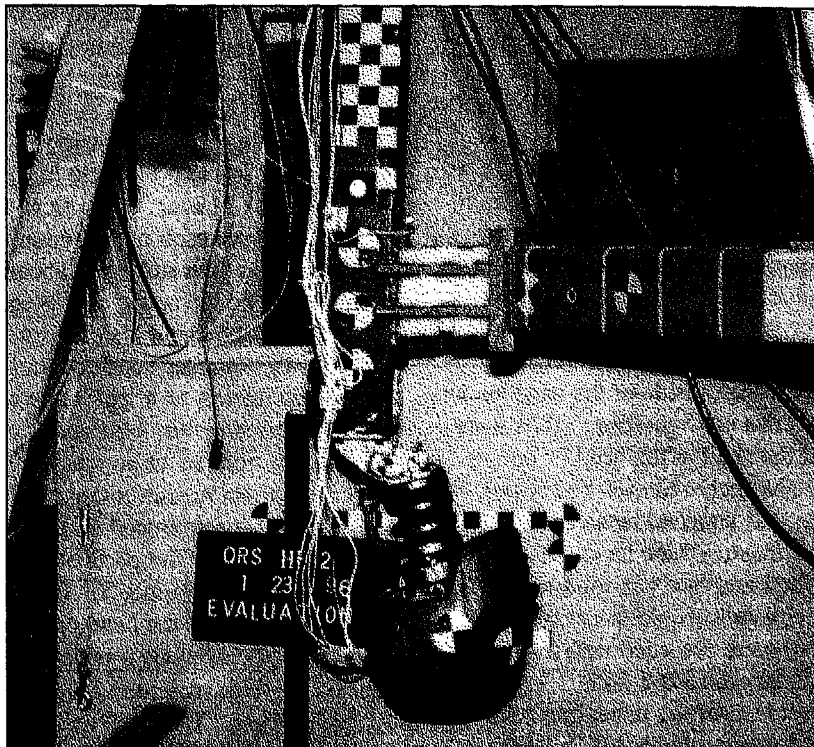


Figure 5 Hybrid III Head and Neck Attached to Pendulum

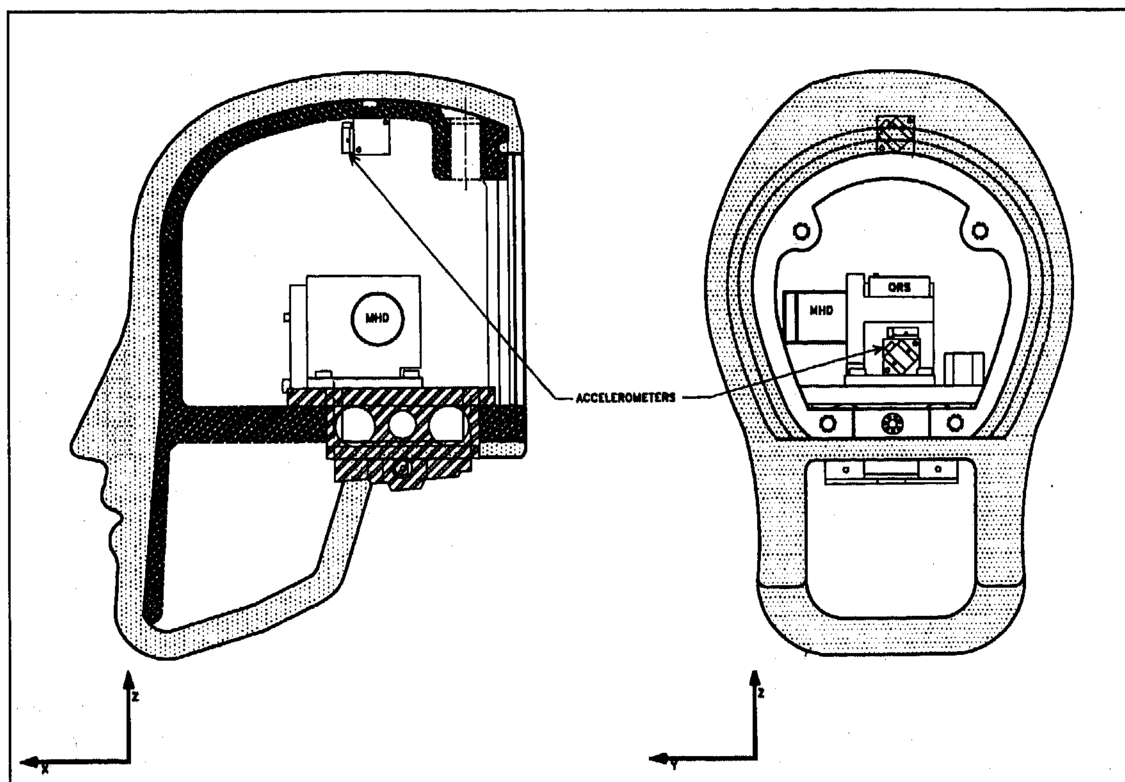


Figure 6 Internal Head Instrumentation (modified from Denton)

forces in the x- and z-axes, and the moment about the y-axis.

VRTC technicians created a special fixture for this test by modifying a Systron Donner tri-axis mounting bracket. A single ARS-01 MHD sensor and a QRS11-03600-100 were mounted to the fixture, and the fixture was mounted to the base of the inner skull. The fixture was carefully created such that each sensor's sensitive axis would be fixed along the appropriate axis. Assuming the dummy skull undergoes rigid-body motion, all points on the skull will experience equivalent rotational velocity and acceleration. It was therefore unnecessary for the sensitive axes of the rotational sensors to align with the head cg. Additionally, potentiometers were mounted to the outside of the head/neck assembly. Figure 7 illustrates this configuration. Tabs were welded to both the outside of the skull at the head cg and to the base of the neck. These tabs anchored each rotary pot. A linear sliding potentiometer was attached between the two so that the head potentiometer measured the head angle relative to the neck angle. The linear pot was used for a convenient sliding axis; linear data were not necessary for angular calculations. The tests were recorded with a nac high-speed video system. An endevco 2262-200 accelerometer measured the rotation of the pendulum arm.

Test data were recorded using the VRTC's DSP Technologies data acquisition system with all circuit parameters as specified by the sensor manufactures. Data were sampled at a rate of 10 kHz. All sensor data were filtered to 1650 Hz in accordance with SAE Class 1000.

To obtain varying impact velocities, the pendulum was dropped from heights of 60°, 90°, and 110°, with 90° being a drop from the horizontal position. Two trials were conducted at each height to ensure repeatability.

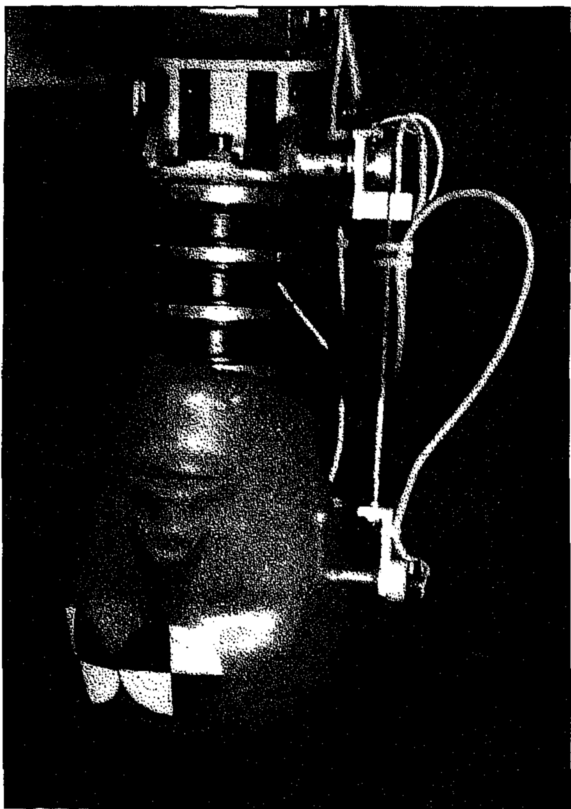


Figure 7 Potentiometer Mounting

Test Results and Discussion

A major obstacle in the analysis of these test results is the lack of exact theoretical data by which to gage the performance each measurement method. A theoretical simulation with sufficient accuracy would be quite complex and beyond the scope of this study. It was therefore decided to gage each method's performance against the result from the potentiometer instrumentation. NHTSA accepts potentiometers as a legitimate method for ATD calibration. When used within their specifications, rotary potentiometers produce very accurate results. In addition, potentiometers take a direct measurement of angular displacement so that error is not introduced from any post-processing requirements. The plots obtained for the full test series are included in Reference 7. Figure 8 illustrates the displacement of the head cg after the honeycomb impact when the pendulum was dropped from 60°.

One may note several interesting trends in the plotted data. The QRS data follow the potentiometer data very closely in both shape and magnitude. The video data also adhere well with the pot data, but the curve is quite choppy. This choppiness is due mostly to limitations on the resolution of the digitizing process. The accelerometer data are very close to the potentiometer curve up until the maximum displacement. After this point, however, the curve begins to taper off. This occurrence is not unexpected since the accelerometer data undergoes multiple integrations. Because integration is a cumulative process, it is likely that the error will increase as time increases. The compensated MHD data appear to follow the shape initially, but then taper off before the peak is reached. Although this test is shorter than the specified time for which compensation is required, using the compensating routine on the data did improve the results. However, the peak results are still low in all cases. The difference seems to become worse as time progresses.

Table 2 shows a compilation of the peak displacement values calculated from each measurement method along with the percent difference as compared to the potentiometer data. This table also demonstrates some interesting trends. While the difference in the QRS data is generally small, the value increases as the drop angle increases. The MHD data exhibit differences as large as 7.27%. The differences in the MHD data decrease as drop angle increases. Both the accelerometer array data and the video data exhibit small differences for all drop heights.

A power spectral density (PSD) analysis was performed on each test in this series to determine the frequency content of each signal. Figure 9 illustrates the power spectrum for a pendulum test dropped from 110°. PSD plots (see Ref. 7) illustrate the relative quantity of a signal that exists at a particular frequency. Peaks on the PSD plot occur at the dominant frequency or frequencies which

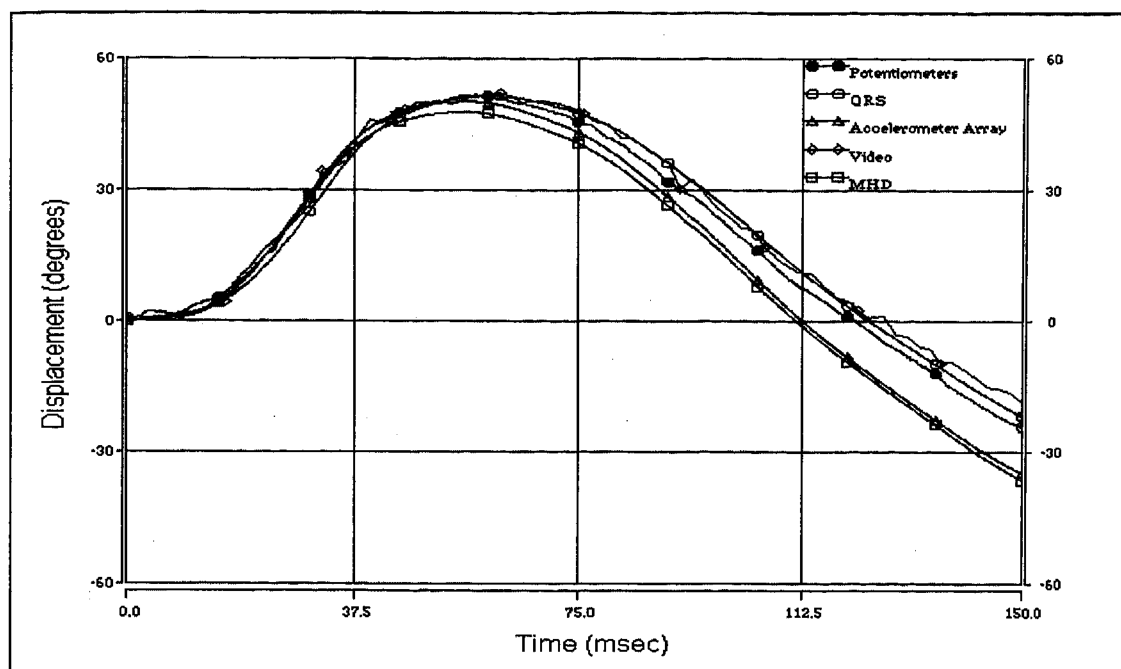


Figure 8 Rotation Measurements for Head/Neck Pendulum Drop from 60°

Table 2 Peak Displacement Values for Head/Neck Pendulum Test Series

Test Number	Drop Height	Potenti-ometers	QRS	Difference (%)	MHD	Difference (%)	Acceler-ometers	Difference (%)	Video	Difference (%)
019	60°	49.39	48.99	0.81	45.80	7.27	47.16	4.51	50.21	1.66
020	60°	51.40	50.86	1.05	47.79	7.02	50.16	2.41	52.20	1.56
021	90°	67.18	69.00	2.71	63.53	5.43	65.88	1.93	71.13	5.88
022	90°	68.12	70.10	2.91	64.48	5.34	66.40	2.52	71.53	5.00
023	110°	71.23	74.24	4.23	68.09	4.41	72.15	1.29	75.12	5.46
024	110°	71.85	74.53	3.73	68.36	4.86	72.16	0.43	75.54	5.14

are in a particular signal. The value of power magnitude is a relative number and can not be compared directly from test to test.

The PSD results are very similar for all drop heights. The PSD plots show that most of each signal is concentrated around 4 and 5 Hz. There is no significant frequency content above 7 Hz, and essentially zero content exists above 20 or 30 Hz.

In this particular head/neck application, the QRS appears to perform well. Results from all six tests agree with the potentiometer results within 4.23%. Additionally, a single integration of the test data provided the displacement results.

The reasons for the percent differences increasing slightly as the drop angle increases have not been determined. The PSD analysis demonstrates that all of the tests were run well within the frequency range of the QRS sensor. It is possible that the difference is due to higher inaccuracy in the potentiometers in this range. However, the slight increase is not significant enough here to cause alarm.

Although we were able to mount the QRS sensors in the Hybrid III head, the physical characteristics of the QRS are less than ideal for this application. Measuring the rotation about all three axes would require the use of three QRS's mounted on a tri-axial mounting kit. By itself, the kit fits fairly easily inside the skull with some minor modifications to the existing hardware. However, the kit takes up most of the skull's space, making it difficult to include other instrumentation in the head. A more significant problem may be that the total weight of three QRS's and the mounting kit is 320 grams (11.3 ounces) before cabling is added. ATD heads are precisely constructed and calibrated to model representative human weights and cg locations. Heavy instrumentation in the head alters both the weight and cg location. Most ATD heads have ballast which can be moved, added and subtracted to adjust for the weight and cg

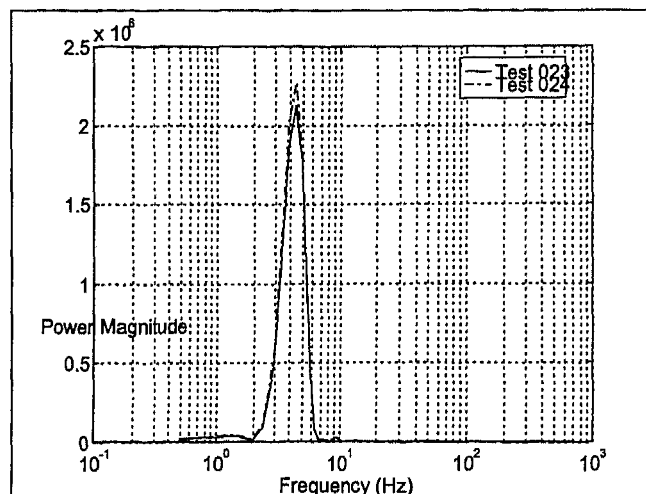


Figure 9 Power Spectral Density of Head/Neck Pendulum Tests from 110°

location, but this is a trial and error process that usually takes considerable time. This modification may or may not be a problem, depending on the nature of the test being performed.

Somewhat larger discrepancies were noted between the MHD and potentiometer measurements. These discrepancies ranged from 4.41% to 7.27%. These results are surprisingly poor based on several other publications (11, 13, 6, 18) which reported moderately better results in similar experiments. None of the publications detail the post-processing of MHD data, including whether or not a compensating routine was used, and what variable inputs may have been used. The Laughlin and Willems reports compared angular velocity measurements, not displacement, to angular velocity calculated from accelerometer arrays. The error may be smaller in the velocity measurement, and then magnified when the data are integrated to obtain displacement.

ATA has rated the bandwidth for the ARS-01 model used for this test at 0.3 to 1000 Hz. All significant frequency content was found from PSD analysis to be within this range. It is conceivable that the performance of the MHD sensor declines towards the low end of its bandwidth.

For the tests discussed here, the peak displacement value occurred sooner than 100 milliseconds. It was assumed that the compensating routine would not improve results because of this short duration. However, the routine did improve results significantly. This is most likely due to the fact that the MHD sensor is in motion for several hundred milliseconds while the pendulum is falling. This motion occurs before the initial test event, the honeycomb impact, occurs. Any pre-event motion must be added to the test duration when making decisions regarding use of the compensating routine.

The MHD sensor has some advantages over other methods. When the compensating routine used in the low-frequency test series is not needed, displacement results are obtained by a simple integration. Also, at 200 grams (7.05 ounces), a tri-axial MHD cube is lighter than the QRS and less likely to significantly affect the cg location in the head. Finally, unlike the QRS's, the MHD sensors can be used singly without the mounting cube, making them much more compact in applications requiring measurement about a single axis.

The peak displacement values calculated from the accelerometer data agreed fairly well with the potentiometer data. Discrepancies ranged from 0.43% to 4.51%. These differences are within the allowable range for most crash test applications. However, the trend of the accelerometer data to increasingly stray from the pot data as the test progresses should be mentioned. In tests of a longer duration, the accelerometers will produce poorer results because of the accumulating integration error.

One reason accelerometer arrays are so commonly used for head instrumentation is because the mounting fixtures and related hardware are readily available. The Hybrid III head, as well as most ATD components, is available with built-in accelerometer mounting utility. Although the number of sensors required is high, the mounting hardware usually does not have to be created or modified.

The high-speed video results are reasonably good. Peak displacement differences range from 1.56% to 5.88%. The video curves tend to follow the potentiometer curves throughout the duration of the test. However, in these particular tests the head motion was constrained on the X-Z plane. The video targets were not prone to a parallax effect nearly as much as they would be on an unbelted ATD in an automobile crash. In addition, this test was set up so that no obstructions were present, although they may necessarily appear in other tests.

As mentioned earlier, this test series presented an ideal environment for potentiometer usage. The type of destructive modification necessary to affix potentiometers to

the ATD is not always acceptable. Nor is the obtrusiveness of the configuration tolerable in all experiments. These limitations are particularly true in crash and sled tests. The mechanism used here measures rotation about the y-axis only. If measurement of rotation about any other axis were required, finding points to anchor the rotary pots to would be much more complex. A potentiometer configuration that measures rotation about all three axes simultaneously would be extremely complex and obtrusive.

Head/Neck Pendulum Testing with Direct Forehead Impact

This test series is very similar to the head/neck pendulum testing previously described. The difference is that after the initial honeycomb impact, the head swings around and strikes a steering wheel. The steering wheel was instrumented with a six-axis load cell. Only three of the axes were recorded for this test: the normal and tangential force, and the moment about the y-axis. A special fixture, illustrated by Figure 10, was designed to attach the steering wheel to the load cell and hold the wheel at an adjustable distance from the head. The steering wheel force data were collected to determine how the instruments respond when impact to a rigid object occurs in a crash or sled test, such as to the steering wheel or A-pillar.

The pendulum impact simulator and the Hybrid III head and neck are identical to that discussed previously. As before, measurement techniques evaluated during this series include the QRS, the MHD sensor, rotary potentiometers, and high-speed video. Accelerometer data was not collected from this test due to channel limitations. The data acquisition process is also identical to the process described previously.

Two pendulum drops were conducted at 60°, and three were conducted at 90°. Higher drops were not attempted for fear of damaging the ATD head and its instrumentation during the steering wheel impact.

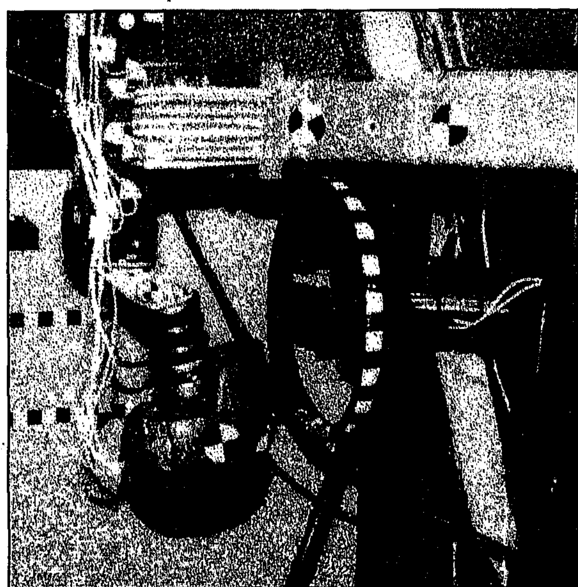


Figure 10 Steering Wheel Impact Simulator

Test Results and Discussion

As in the previous test, results from this test series are compared against the results from the potentiometer instrumentation. Since the results vary from test to test, even when the tests were conducted at the same pendulum drop angle, all five of the tests in this series are included and discussed in this section.

Figure 11 illustrates the head displacement after the honeycomb impact for test 004, which represent a drop angle of 60°. Figures 12 illustrates the head displacement for test 007, which was a pendulum drop from 90°. Table 3 explicates the peak displacement values calculated from each measurement method along with the percent difference as compared to the potentiometer data.

Greater variation exists in this test series than in the previous two. The QRS agrees well in both shape and magnitude for all tests except Number 007. For this particular test, the QRS measures 9.48% higher than the potentiometers. The MHD shows percent differences ranging from 3.10% to 9.59%. The video results demonstrate significant differences, as high as 29.25%. The video data is also very choppy. There is probably significant error in the high-speed video data. As the forehead hit the steering wheel, one of the digitized targets became difficult to track. This is particularly true at the peak displacement which occurs during the steering wheel impact.

A power spectral density (PSD) analysis was performed on each test in this series to determine the frequency content of each signal. Figure 13 illustrates the power spectrum for pendulum tests dropped from 60°. The PSD results are very similar for both drop heights. The PSD plots show that all of the signals exhibit peaks at 5 Hz. There is no significant frequency content above 7 Hz, and essentially zero content exists above 20 Hz or below 2 Hz.

Size, weight, and post-processing considerations are identical here to those in the previous test, and will not be repeated here. However, the performance characteristics will be discussed in detail.

There is significant variability in the peak displacement between the first pendulum test from 90° and the third 90° test. The same steering wheel was used for all the tests in this series. It is very likely that the first 90° test, number 007, hit the steering wheel hard enough to deform the wheel and change its stiffness. In the second test, the forehead hit a softened wheel and was able to travel further. This effect was even more significant on the third test.

The QRS appears to perform well in all of the trials except Test Number 007. Tests 004, 006, 008, and 009 exhibit differences between 2.33% and 4.49%, and follow the shape of the potentiometer curve throughout its range of motion. It is unclear why the result of test 007 is so poor compared to the other tests. The PSD analysis of the entire test series shows no significant difference between the frequency concentration in this particular test and the rest of the test series.

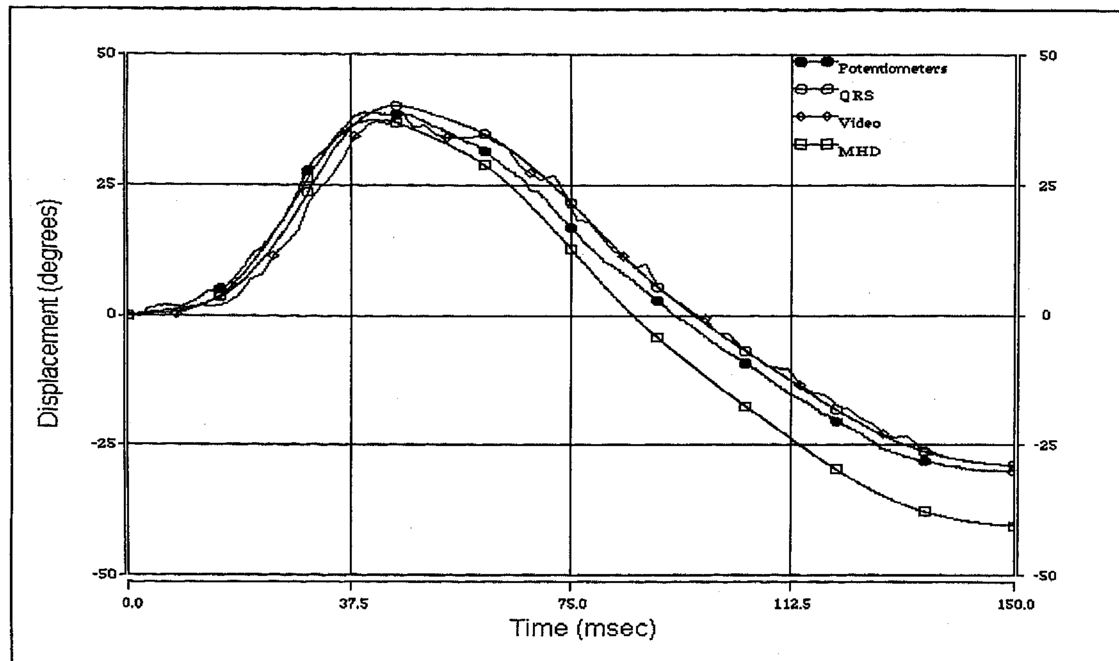


Figure 11 Rotation Measurement for 60° Drop, Test Number 004

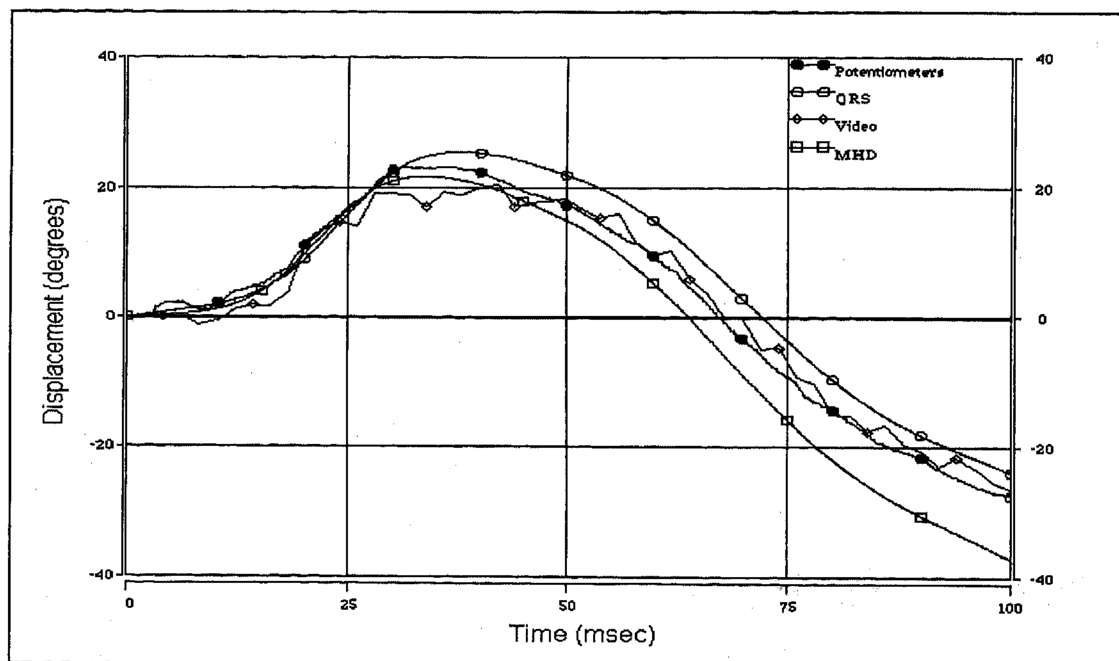


Figure 12 Rotation Measurement for 90° Drop, Test Number 007

The MHD data were examined in their compensated and uncompensated forms, and again compensation improved the data. Therefore, the compensated data are presented here. The discrepancies found for tests 004 and 006 were relatively small. The discrepancies were a bit higher for tests 007 through 009.

The video data here showed very poor results. Following any one of the displacement curves, one notices a large deviation from the potentiometer data right around the peak. This deviation decreases after the peak and then seems to catch back up several seconds after the peak. Apparently, the digitizer loses the target as the forehead strikes the steering wheel and then finds it again after the head rebounds. This problem could be corrected in this case by

Impact Test Series

Test Number	Drop Height	Potenti-ometers	QRS	Difference (%)	MHD	Difference (%)	Video	Difference (%)
004	60°	38.72°	40.12°	3.62	37.52°	3.10	38.91°	.49
006	60°	39.02°	39.93°	2.33	37.21°	4.64	36.17°	7.30
007	90°	23.20°	25.40°	9.48	21.72°	6.38	20.26°	12.67
008	90°	30.96°	32.11°	3.71	27.99°	9.59	*	--
009	90°	38.12°	39.83°	4.49	35.81°	6.06	26.97°	29.25

* Film data was corrupt for this test

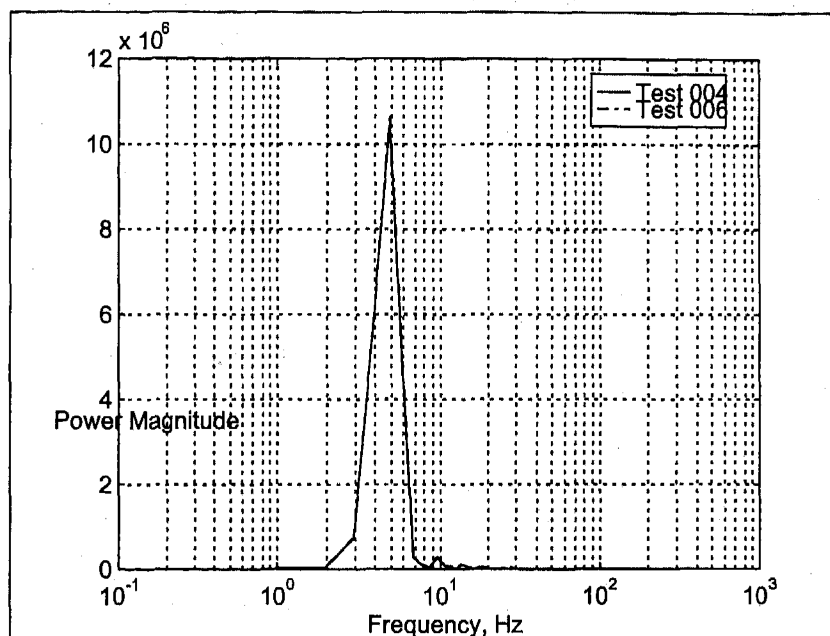


Figure 13 Power Spectral Density of Head/Neck Pendulum with Steering Wheel Impact Tests from 60°

repositioning the targets so that they are in better view of the video camera.

CONCLUSION

Evaluation of Measurement Techniques

Five different measurement techniques were evaluated for this study. These techniques include direct measurement from rotary potentiometers, digitized video analysis, and indirect measurements from QRS's, MHD sensors, and translational accelerometers. Each technique has unique characteristics which make them suitable for different applications.

Quartz Rate Sensor Evaluation

The QRS has proven to be a viable alternative for measuring displacement in many typical ATD tests. Test results indicate that the QRS provides accurate results in both low and medium frequency environments. Unlike some other methods, only one sensor is required to measure rotation about each axis, which simplifies the test configuration and data acquisition.

There are some characteristics which may make the QRS undesirable for a particular test. Systron Donner specifies the QRS's bandwidth as "greater than 60 Hz", with a -3 dB point at 75 Hz. Signals from tests such as direct windshield impacts, which may contain frequency content above 60 Hz, should not be recorded with this sensor. Tests with strict weight limitations may not allow the 320 gram sensor. Additionally, the size of the QRS limits it to use in the larger body components such as the skull, chest, or pelvis.

Ideal applications for the QRS include measurement of the angle of the chest and pelvis. The QRS is also suitable

for head measurements in low and medium frequency tests. The QRS is also ideal when one needs to reduce the number of data channels required due to limited channel availability.

MHD Sensor Evaluation

The MHD has limited uses in ATD applications. The MHD showed decent accuracy in the both of the head/neck tests, but is marginal for the low frequency pendulum block test. There is the possibility of additional "user error" in MHD test results because of potential misuse of the ill-defined compensating routine.

The primary advantage of the MHD is that it has very high frequency capabilities, much higher than the QRS's or potentiometers. Like the QRS, only one MHD sensor is required for obtaining data about each axis. For measurements in three dimensions, the MHD is available in a compact tri-axial cube.

Ideal applications for the MHD include any measurement of motion which contains extremely high frequencies for which more accurate sensors are not available. These applications may include examination of some non-biofidelic responses in an ATD.

Accelerometer Evaluation

Accelerometers have long been used for ATD response measurements. The data presented in this study show good accuracy from the accelerometers in the medium frequency range. Because accelerometers have been widely used in the past, the fixtures required to mount them are already in place inside many ATD components. This availability makes them convenient to use. In addition, an accelerometer can be added by drilling two small holes, which can be considered a minimal alteration to the dummy.

The use of translational accelerometers to measure angular rotation also has some drawbacks. Typically nine accelerometers are used to measure rotation about three axes. The MHD and QRS methods require only three sensors. With many data acquisition systems, accelerometer usage will limit the number of channels available for other purposes. The post-processing required to obtain rotational displacement need also be mentioned. The multiple integration process accumulates error in the displacement data. Longer tests will continue to accumulate error as the test progresses.

Accelerometer arrays are ideal for measuring angular rotation when a test requires the array for other purposes such as acceleration data. For an application such as this, no additional instrumentation would then be required for the displacement calculation. Tests in which excessive instrumentation is not an issue could make use of an accelerometer array.

Rotary Potentiometer Evaluation

Rotary potentiometers do provide accurate results, but their mounting requirements restrict their use significantly. Obtaining rotation about a single axis requires two potentiometers to be rigidly attached to the ATD with a sliding axis between the two. This configuration must be mounted externally on the ATD. Necessary modifications require permanent alteration of the ATD component. The sensors and their sliding member can not be subjected to any direct impact. Therefore, potentiometer use is restricted to tests in which motion is very constrained. Ideal uses for rotary potentiometers include constrained tests such as the head/neck pendulum tests, as described in this study.

High-Speed Video Evaluation

High-speed video is very useful in ATD testing, but should not be used as a primary means of obtaining angular displacement data. The video's accuracy ranged from excellent to terrible in the medium frequency tests. When used in ideal settings with optimum lighting with motion constrained on a small single plane, the video is very accurate. However, few applications are this ideal. Many potential problems may arise from video analysis, including the parallax effect described earlier and obstructed views of targets. The resolution of the digitizing process is also limited, which results in choppy data.

Despite its limitations for providing accurate displacement data, high-speed video is a very useful tool. The digitizing process is simple and inexpensive, and provides a quick ballpark check on other data. More importantly, video provides a visual record of a test. When oddities surface in sensor data, video data can provide important clues about their origin. For instance, unusual data spikes or broken body components may be the result of an unexpected collision between a body part and a test fixture or another body part.

ACKNOWLEDGMENTS

The authors would like to thank all of the many people who assisted in this project. The authors are grateful to the National Highway Traffic Safety Administration for sponsoring this research. Finally, all discussions and conclusions in this paper represent the opinion of the authors and not necessarily those of the NHTSA, TRCO, or The Ohio State University.

REFERENCES

1. Burke, W. Vincent. Evaluation of a Prototype Magnetohydrodynamic Angular Acceleration Transducer. US DOT HS 808186, Final Report, 1993.

2. Denton, Robert A, Inc. Unique Solutions to Measurement Problems. Rochester Hills, MI.

3. Elias, Jeff, Howard Pritz and Roger Saul. Accelerometer Equivalency: Determination of the Dynamic Environment for Accelerometers Used in Dummy Impact Testing. Event Report VRTC-83-0303, US DOT Vehicle Research and Test Center. East Liberty, OH: 1995.

4. Freeman, C. M. and D.G.C. Bacon. The 3-Dimensional Trajectories of Dummy Car Occupants

Restrained by Seat Belts in Crash Simulations. Proceedings of the 32nd Stapp Car Crash Conference. Warrendale, PA: 1988.

5. Garrett, Ted. "Evaluation of Kinematic Measurement Devices for the Cervical Spine". Event Report, US DOT Vehicle Research and Test Center. East Liberty, OH: 1995.

6. Laughlin, Darren R. A Magnetohydrodynamic Angular Motion Sensor for Anthropomorphic Test Device Instrumentation. Proceedings of the 33rd Stapp Car Crash Conference. Washington, D.C.: 1989.

7. Marshall, J.A., Evaluation of Angular Displacement Measurement Techniques for Tracking the Motion of Anthropomorphic Test Devices, M.Sc. Thesis, The Ohio State University, 1996.

8. Marguiles, Susan Sheps and Lawrence E. Thibault. A Proposed Tolerance Criterion for Diffuse Axonal Injury in Man. Journal Biomechanics, Vol. 25, No. 8: 1991.

9. Melvin, John W. and T. Rex Shee. Facial Injury Assesment Techniques. 12th International Technical Conference of Experimental Safety Vehicles, US DOT National Highway Traffic Safety Administration, Goteborg, Sweeden.: 1989.

10. Ommaya, Ayub K. "Biomechanics of Head Injury: Experimental Aspects." The Biomechanics of Trauma. East Norwalk, CT: Appelton-Century-Crofts, 1985.

11. Saul, R., B. Tanner and D. Zuby. Hybrid III Neck Instrumentation Measurement Requirements. US DOT National Highway Traffic Safety Administration Vehicle Research and Test Center Final Report, 1991.

12. Saul, R., Brian Tanner and Thomas F. MacLaughlin. Hybrid III Neck Hyper Extension Determination. Event Report 860034-03, US DOT Vehicle Research and Test Center. East Liberty, OH: 1990.

13. Shipley, Buford W., Jr. Donna Jo Baughn and Howard B. Pritz. An Evaluation of Sensor Systems for Measuring Rotational Motion. SAFE Symposium. Las Vegas: 1993.

14. Systron Donner. GyroChip. Concord, California: 1994.

15. Viano, David C. Biomechanics of Head Injury--Toward a Theory Linking Head Dynamic Motion, Brain Tissue Deformation and Neutral Trauma. Proceedings of the 32nd Stapp Car Crash Conference. Warrendale, PA: 1988.

16. Viano, David C. and Sudhakar Arepally. Assessing the Safety Performance of Occupant Restraint Systems. Proceedings of the 34th Stapp Car Crash Conference. Orlando: 1990.

17. Viano, David C., et al. Measurement of Head Dynamics and Facial Contact Forces In the Hybrid III Dummy. Proceedings of the 30th Stapp Car Crash Conference. San Diego: 1986.

18. Willems, Gilbert C. and David R. Knouse. A Detailed Evaluation of the ATA Angular Motion Sensor in Realistic Simulated Crash Environments. Proceedings of the 35th Stapp Car Crash Conference. San Diego: 1991.

971056

Topsy - A Modular Chassis Parameter Measurement System

Edward J. Heitzman and Edward F. Hietzman
Heitz Chassis Lab., Inc.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

Topsy is a modular system designed for occasional measurements of all composite inertial, kinematic, and compliance parameters which appear in today's sophisticated mathematical simulation models. The emphasis in its design is not "production-line" testing, but research: the operator can observe the vehicle's reaction to developing forces; and as insights develop he can modify the test program as he goes along. Test errors are minimized by running all tests under servo control with continuous display of all data and recording of all known error sources, while the operators closely monitor the test and the developing data.

The test facility is organized around an "infrastructure" consisting of baseplates, vehicle locating fixtures, scales, hydraulic and pneumatic power sources, interchangeable valve assemblies and actuators, and transducers. These are organized along with specialized devices into the several forms required for the different tests.

When not in use, Topsy can be disassembled and stored on shelves and in cabinets.

This paper describes the overall system and the methodology, components, and instrumentation used in the various measurements.

INTRODUCTION

Like its namesake, the little slave girl in *Uncle Tom's Cabin*, Topsy "was never born, but just grew". Archival records indicate that it was being used for center of gravity measurements on various small cars in 1982 and on a Chevrolet Suburban in 1988; structural shake of a Mercedes convertible in 1986; Audi ride steer and Mazda lateral acceleration weight transfer in 1987; etc. Over the years it has grown in a process driven by anticipation of customer test requests and the practice of determining after each test how it could have been done better. Topsy has developed into a modular system because of the way it has grown: whenever possible the equipment on hand is revised, modified, or redesigned to perform new tasks or to do old tasks better. In some cases new and old approaches are used interchangeably until one proves its superiority, then the "loser" is discarded. In construction or acquisition of any new equipment, strong consideration is given to possible multiple uses.

The methodology for measurement of CG height, roll moment of inertia, and roll/yaw product have evolved from the methods used by Cornell Aeronautical Laboratory in measurements on a 1953 Buick, as described in unpublished memoranda written in 1956. Compliance methodologies were inspired by the "Nedly-Wilson" Vehicle Parameters Facility built at Chevrolet in 1971 [1]. More recently, development has been guided by SAE J1574 [2].

Topsy has now reached a state of maturity at which its developers feel that it is worth reporting on. We will first describe the "infrastructure", and then how these components are organized with special-purpose devices for the different tests: whole-vehicle inertia; compliance and steering; kinematics; shocks and struts.

INFRASTRUCTURE

BASEPLATES - The various test devices utilize three baseplates: the laboratory floor; a vertically-mounted aluminum I-beam; and the Frame/elevator assembly.

The lab floor is concrete overlaid with industrial tile. Threaded steel inserts are installed in the concrete wherever necessary to locate fixtures. When not in use the holes are filled by flat-head stainless steel capscrews. When desired hole locations change the old holes are covered with new tile.

The 150 mm vertical I-beam is permanently attached between the floor, the cinderblock lab wall, and

several ceiling joists of the adjacent electronics laboratory room. It serves as the baseplate for measuring springs, struts, and shock absorbers, and as the reaction member in sidepull testing.

All inertial, kinematics and compliance measurements are made on the Frame/Elevator assembly which is shown in lowered position in Figure 1. This assembly forms a testing baseplate which is assembled when needed. It can be raised and lowered to a convenient working height, and can be pitched for a "drive-on" capability.

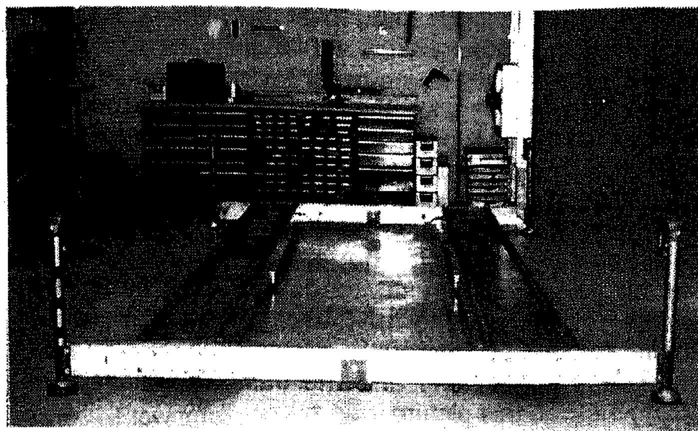


Figure 1: Frame/elevator assembly in down position

FRAME DESIGN - The Frame consists of two transverse beams and two longitudinal siderail beams. The transverse beams are rectangular aluminum tubes, reinforced by aluminum plates welded on the upper and lower surfaces, and by welded-on plates where attachment is made to the siderails. The siderails are 380 mm wide welded assemblies, 4090 mm long, made up of three rectangular tubes capped by welded-on 3 mm aluminum plates and 13 mm thick endplates. Galvanized "Unistrut" steel channels with rolled-over edges are attached to top and bottom edges and along the bottom centerline of the longitudinal beams. These are bolted at 152 mm intervals to flush-mounted "Rivnut" threaded inserts. The Unistrut rails add considerably to bending stiffness, and with the many available clamps and fittings provide a convenient system for attaching equipment to the Frame.

The basic moment of inertia of each siderail is estimated to be 1500 cm^4 of aluminum. The Unistruts add 500 cm^4 of steel. The composite moment of inertia is therefore approximately 3000 cm^4 of aluminum. The effective moment of inertia of a current transverse beam is about 625 cm^4 of aluminum, which will soon be doubled by welding on another upper and lower cap. Bending deflections are generally low because the distance from vehicle wheel centers to support members is short. The mass of each side rail is 98 kg and that of each transverse beams is 15 kg without elevators.

The attachment between a siderail and transverse beam is through 12 bolts at each end, with holes tapped into the siderail ends. Spacing between siderail centerlines is adjustable from 1270 to 1702 mm to accommodate vehicles of various trackwidths. Wider trackwidths will require welding on more plate and drilling more holes in the transverse beams, or welding up longer ones. The side beam length of 4090 mm accommodates vehicles of up to 3300 mm wheelbase.

Ramps attach to the end of the siderails. With the Frame Elevator Assembly at maximum pitch angle, a vehicle can be driven onto the frame assembly.

The Frame has secondary use as a general-purpose work table to which components can be clamped for testing. When its space is needed for other projects, it is disassembled and stored on a shelf above the laboratory entrance doors. Assembly takes one man about one hour.

FRAME ELEVATORS - The Elevator Assembly consists of four motorized ball- screw linear actuators, attached vertically to the end rails at each corner of the Frame. Each actuator is capable of over 9000 N and has a stroke of 460 mm. The actuator ends are fitted to rubber or polyethylene feet through spherical bearings.

The actuator motors are connected through the inside of the Frame beams and a single cable to a hand-held control module, where they are controlled, singly or together, by toggle switches. The elevators provide a table height range of 215 to 675 mm, and a pitch range of ±6 degrees.

Each elevator actuator has a mass of 19 kg. They are stored in a closet when not in use.

VEHICLE LOCATING JACKS

SCISSORS JACK - The two scissors jacks, one of which is shown in Figure 2, were designed to lock the vehicle to the Frame at selected ride heights, for inertial and compliance measurements. When raised or lowered by turning the Acme-threaded screws with a wrench or air wrench they remain horizontal, so that front and rear ride heights can be adjusted without worry about introducing roll. They are laterally stiff, since parallelogramming requires bending the 19 mm diameter screw into a sine wave. Nevertheless, during early lateral compliance testing, in which displacement transducers were referenced to the Frame rather than the vehicle body, it became necessary to put adjustable threaded spacers under each threaded block to reduce the lateral compliance resulting from this sine wave bending. With the spacers installed, lateral deflections were reduced to 0.2 mm at 4100 N sideforce. The mass of each scissors jack is 19 kg. When not in use they are stored vertically in the "crossmember corner" of the laboratory.

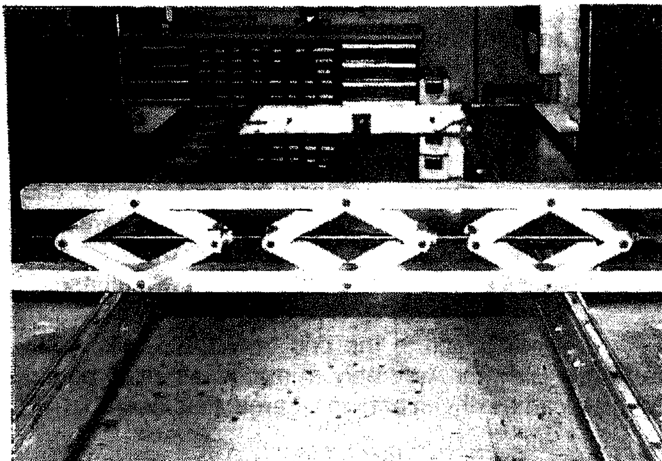


Figure 2: Scissors jack assembly

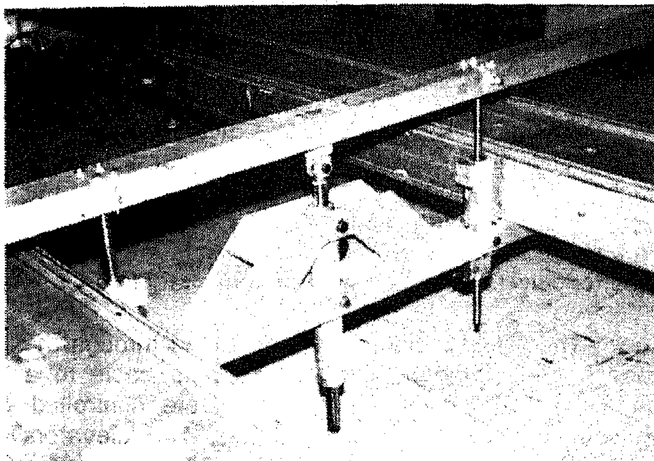


Figure 3: Ride/Roll crossmember mounted below Frame, with servo cylinders

RIDE/ROLL CROSSMEMBER - The "Ride/Roll" crossmembers, one of which is shown in Figure 3, were designed to exercise the vehicle suspension in kinematics testing, but can also be used in compliance tests. The center restraint for this crossmember is a 32 mm diameter steel shaft running in linear ball bearings in a cylindrical housing, which is attached to the lower crossmember "bridge truss" by shaft collar-trunnions. The cylindrical housing is adjusted up or down to optimize the loading on the vertical shaft. The inner shaft attaches by a spherical bearing to a clevis welded to a 38 mm diameter screw, which is threaded into a block in the upper crossmember. The location of the spherical bearing defines the point about which the vehicle rolls, provided that the shaft is vertically locked by tightened shaft collars. That point can be adjusted by screwing the clevis into or out of the crossmember. With the vertical shaft free, the vehicle finds its own roll center, which can be computed from the vertical shaft motion or by other means.

The ride/roll crossmembers can be assembled "upside down" or "rightside up", and they can be attached to either the bottom of the Frame as in Figure 3 or to the top side. One of the four configurations is chosen for a particular test according to test vehicle road clearance and ease of installation.

The vertical displacement and roll angle of the upper crossmember is controlled by any of several trunnion-mounted linear actuators. For compliance testing slow-moving electric motor ball screw actuators are used, or hydraulic cylinders are locked vertically by shaft collars. For kinematics and frequency response testing, a variety of hydraulic actuators are used, depending on force, stroke, and piston velocity requirements.

Removal of an actuator requires only pulling out a spring-loaded hitch pin from the upper spherical rod-end, and removing two trunnion bolts.

The housing for the vertical slider is normally located rigidly at top and bottom for lateral stiffness. However, for measurement of roll/yaw product in dynamic testing one end is located as shown in Figure 4. Here the housing is pivoted about its lower mount, but the upper end is free to move laterally except for the load cell restraint. Fore/aft loads are taken by the ball bearing, which tends to roll on either the front or rear surface of the channel with a clearance of about 0.2 mm. The vertical slider and the truss sides have been enlarged since this picture was taken.

The lower "truss" crossmember has a mass of 15.6 kg, and that of the upper crossmember is 11.6 kg. They store in "crossmember corner". The vertical bearing and shaft assembly has a mass of 5.3 kg, and stores in a closet.

ATTACHMENT TO VEHICLE - Where possible the Scissors or Ride/Roll crossmembers bolt to "rivnuts" installed in the vehicle frame or underbody. (After testing the rivnuts are plugged with setscrews.) For passenger cars, they are more often clamped to rocker pinch welds, using machinists vises modified for air-wrench installation. The vises bolt to or are clamped to the locating crossmembers.

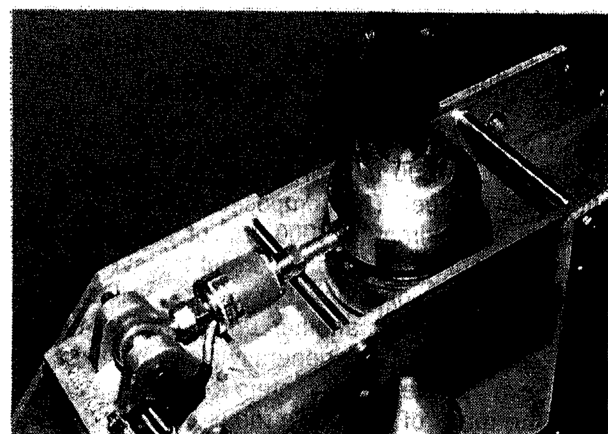


Figure 4: Side force measurement in roll dynamic testing

ACTUATORS - An array of electric motor and raulic actuators are used interchangeably, with a ındard" trunnion mount. For essentially static itioning jobs, Saginaw ball-screw actuators are used, er with Saginaw motors or smaller Dayton motors ı higher-ratio gearing. With one exception, motors 90 volts, run off rectified 115 Vac and controlled by ɟle switches or by operational amplifier-driven relays. the pull-point locators in sidepull testing, the ɹators use 12 volt motors driven by general-purpose ʌo amplifiers.

Hydraulic cylinders or servo cylinders are used ɜnever variable force or motion is needed. Four sizes hydraulic cylinders are used, depending on force, ıke, and matching of piston velocity requirements to ted pump capacity. Piston areas (on the rod side) are ıroximately 1.3, 2.6, 6.5, and 12.9 cm^2 respectively.

The three smaller size cylinders use the same ıic design, which may be configured as single-ended, equal-area/double ended, by selection of end-caps I piston shafts. The photo in Figure 5 shows the same nder in three configurations: a single-ended, end unt in the foreground; with double-ended, trunnion unt and a double-ended, end mount in the rear.

Flow into and out of the cylinder is controlled by nifold blocks. The input end may be fitted with a ıple ported block or with servo valve manifold blocks. ɜ manifold blocks can be fitted with servo valves, or ı red or blue "dummy blocks". Blue dummy blocks ɾe the same flow direction as the simple porting cks, and red dummies have porting of the opposite ɜction. Cylinders are easily changed from parallel to ɔosing directions by switching dummies.

The cylinder end opposite to the porting block ; provision for a geared potentiometer, which is ɾen by a rack gear running in a groove milled into a ner of the cylinder. The potentiometer provides direct dback without intervening compliance or backlash ɜn the cylinder is used as a position servo.

The "big" cylinder is similar in design but is not fitted with a servovalve. The cylinder body is simply a section of thick-wall extruded tubing, which is drilled and tapped for end caps and fittings, then honed. Because it does not have to be bored, it is easy to make any length cylinder.

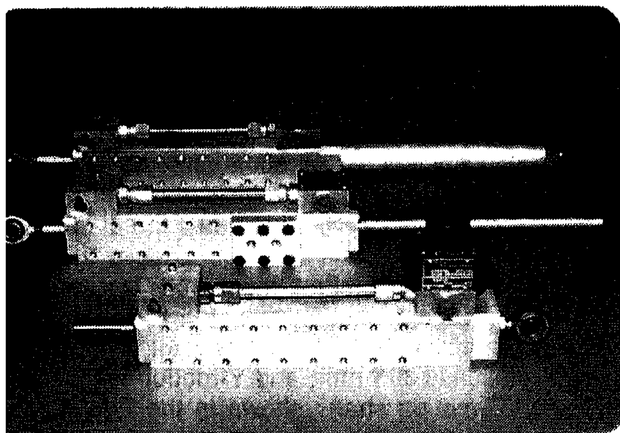


Figure 5: Hydraulic cylinder configurations

All cylinders use O-rings with less squeeze than that normally specified, for low friction.

The "bad news" about hydraulics is given by the approximate equation:

$$cm^3 / s = \frac{kW \times 10^6}{kPa}$$

At 10 000 kPa, a 4 kW pump supplies an average flow of about 400 cm^3/s. Since a single 6.5 cm^2 cylinder with a 5 cm peak-to-peak stroke at 1 Hz requires an average flow of 65 cm^3/s, the smallest diameter cylinder that will do the job is the one to be used in dynamic testing.

SERVO VALVES - Flow control servovalves are used for position servos, and pressure control servovalves for force generation. The pressure control valves are mounted in assemblies containing a filter, valve, and differential pressure transducer. All servo valves are driven by standard operational amplifiers.

HYDRAULIC SUPPLIES - Two similar hydraulic supplies in roller cabinets are employed. Both are shown in the photo of Figure 6. One has a variable displacement pump driven directly by a 1.2 kW, 115 volt motor; the second one has an identical pump belt-driven by a 4.4 kW, 230 volt motor. The pump in the 4.4 kW cabinet can run at 10 000 or 20 000 kPa by changing the compensator spring and the belt pulleys. Both cabinets have the necessary filters, reservoir, accumulator, etc. The smaller one is fitted with two servovalves with outputs fed through solenoid valves for locking cylinders in place, for the sidepull system. The larger one is used in kinematics and compliance testing.

The hydraulic pump cabinets connect by flexible hoses to a floor-mounted distribution system, which for kinematics/compliance testing runs lengthwise under the Frame. The distribution system consists of parallel stainless steel lines with accumulators on each end, and

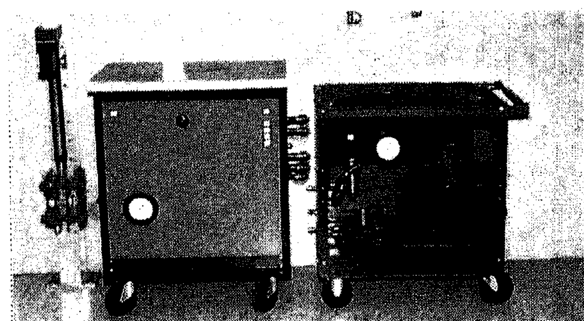


Figure 6: Hydraulic cabinets

servovalve manifolds or outlets where required. Relatively large accumulators are used to provide flow capacity well in excess of pump capacity for periods of several seconds. The accumulators are charged by a motor-driven 20 000 kPa air compressor.

WHEEL-LOAD SCALES AND BEARING PADS- Four strain-gage platform scales are used for vehicle weight and individual wheel loads. Each scale platform is 381 mm square and 100 mm high. An accuracy of ±0.02 percent is specified with the load located anywhere on the platform. Maximum load is 8900 N per scale. Accuracy certification is to NTEP Class III, one part in 6000.

The scales are not "drive-on": the vehicle must be raised and then lowered onto the scales.

For measurements of suspension kinematics and compliances, each wheel scale is fitted with ball-bearing low-friction pad assemblies. Each assembly consists of upper and lower hardened steel plates, separated by an array of hardened steel balls. Each upper plate is fitted with mounting ears for a wheelpad forcer assembly, and has a serrated steel friction surface.

POSITION TRANSDUCERS - Two types of position transducers are used. Spring-loaded rectilinear potentiometers are used for small, well-defined deflections such as lateral vehicle motion or deflections of the Frame in CG tests. For all other work we now use "home-made" string pots or string encoders. A string device (Figure 7) consists of a spring motor with its output cable wrapped in a single turn around a pulley mounted on a pot or encoder shaft. Spring motors come in several tension levels, and are inexpensive and reliable. Since the "chassis" is just a piece of flat stock it can be modified for a given application and replaced for the next. Precise encoder scaling is accomplished by selecting pulses per revolution and then machining a pulley to the correct diameter.

Encoders are used wherever possible because they have good resolution with infinite "stroke", and are zeroed anywhere in that stroke with a pushbutton switch. Potentiometers are used only when position data should not be lost when power is turned off.

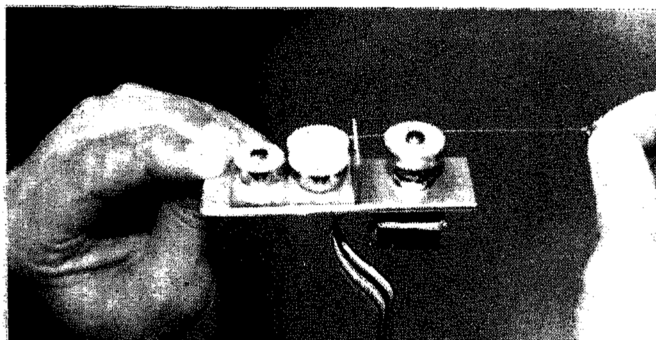


Figure 7: Small string encoder

ACCELEROMETERS/INCLINOMETERS

Force-balance servo accelerometers, either Systron-Donner Model 4310 or Sundstrand Model 303B, at 0.25, 0.5, or 1.0 g full scale, are used as inclinometers. These units have linearities of 0.05 percent full-scale, and resolutions of 0.0006 degrees. A Systron-Donner force-balance servo angular accelerometer with ±10 radians/second full scale, 0.1 percent linearity, and 0.005 percent resolution is used in angular inertia testing.

STEERING WHEEL - A "second steering wheel" instrumented to measure angle and torque is used in compliance and steering system measurements.

WHEELHUB TRANSDUCER PLATE - Wheelhub transducer plates shown in Figure 8 are designed for steering ratio and geometry, kinematics, and compliance measurements. Adapter plates are attached to lugnut extensions through four or five slots which guarantee accurate centering. The resulting spin axis extension is a threaded shaft containing a ball bearing hanger for a plumb bob or a tire deflection encoder, and a lockable attachment to the wheelhub transducer plate. This plate contains camber and caster inclinometers, a bubble level, a rearward-facing laser pointer for steer angle zeroing, a ±45 degree, 0.05 degree resolution "coarse" steer angle encoder, and attachment points for the fine steer angle encoders required for kinematics and compliance measurements.

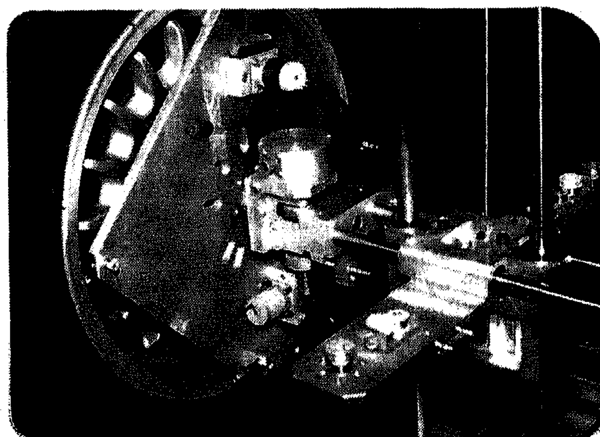


Figure 8: Wheelhub transducer plate

BALL-SLIDE FOLLOWER MECHANISM - The ball-slide mechanism shown in Figure 9 was designed to measure wheelcenter motion in kinematics and compliance testing, and vehicle bumper motion in sidepull testing. It consists of three orthogonal, dual-shaft, linear ball-bearing sliders, in X-Z-Y configuration, with motion measured by spring-motor encoders. Range of motion as determined by lengths of the slider shafts is arbitrary. At present, X-motion is 460 mm, Z-motion is 305 mm, and Y-motion is 230 mm, but the short Z-motion shafts shown in the figure are used for closet storage. Transducer range is arbitrary since the encoders can be zeroed anywhere. The encoders

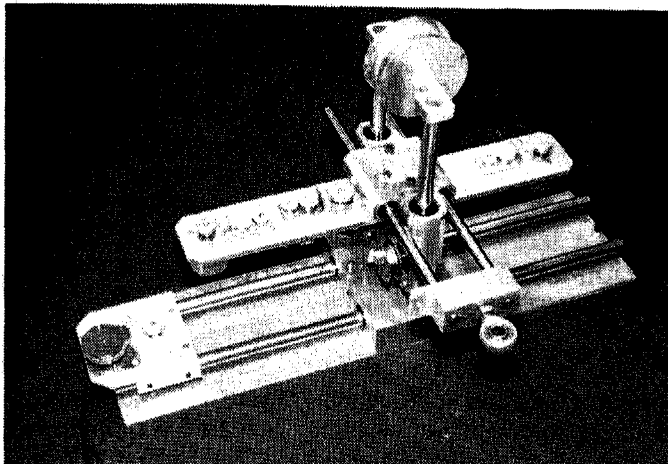


Figure 9: Ball-slide follower mechanism

installed for X-Z-Y motion have 256 pulses per revolution and 20.7 mm diameter pulleys, which with the 12-bit data acquisition conditioners result in ±127 mm full-scale with 0.06 mm resolution at 4X, and ±508 mm full-scale with 0.25 mm resolution at 1X. Force levels are 4 N in the X-axis and 1.6 N in the Y-axis, determined by the spring motors. In the Z-axis the weight of the Y-axis portion and the pull-down Z spring motor are counterbalanced by a vertical spring motor.

For compliance and kinematics testing small steer angles are measured by two additional encoders, spaced 356 mm apart to match attachment points on the wheelhub transducer plate, as shown in Figure 8. These steer angle encoders have 1,000 pulses per revolution and pulleys chosen for a 4X full-scale of ±4 degrees with 0.002 degrees resolution, and a 1X full scale of ±16 degrees with resolution of 0.008 degrees. As the slider mechanism follows the X and Z wheel motions, no corrections for such motions are needed.

DATA ACQUISITION SYSTEM - The proprietary data acquisition system shown in Figure 10 allows close monitoring of a test in progress, and useful review of completed test programs. All testing is videotaped, with raw transducer data recorded in serial binary form on one or more lines at the top of the picture. Voice, noises, and transducer outputs are recorded in synchronism continuously during the entire test on two-hour tapes, so that anomalies that might occur during or between test runs can later be investigated. Throughout the test, data is displayed as columns of numerics, twelve channels to a column, overlayed on the video picture, for visual monitoring. The displayed overlay does not have to be recorded, as it can be recreated and recorded at any time. All processed data is referenced to hours, minutes, and seconds of tape time, so that final plotted data can be referenced to the video. Figure 11 shows a typical compliance test screen with numeric overlay.

Transducer data is conditioned in general-purpose, strain gage, shaft encoder, or frequency input amplifiers; and in tenth-order 14.2 Hz



Figure 10: Data acquisition console

Butterworth active filters. All signal conditioners have manual and autozero capability, and on-board calibration (E-Cal, R-Cal, Excitation Off, Fixed-Count, or crystal oscillator) appropriate to their function. The filtered data is sampled 60 times per second, digitized to 12 bit accuracy and resolution, and stored with 12 channels on one horizontal line of the video recorder. Each data line appears as a dot-dash line across the top of the visible video picture. Figure 12 shows a screen with 24 channels of binary data on two video lines.

The system is designed to use a maximum of nine "data lines" with 12 channels per line, for a total of 108 10-Hz data channels sampled at 60 Hz. However, it is also possible to devote some data lines to "high-speed" use, with two data channels sampled at 480 Hz or one data channel sampled at 960 Hz.

Interface between the Pentium-PC computer and either the data acquisition system or the replayed



Figure 11: Screen with numeric overlay

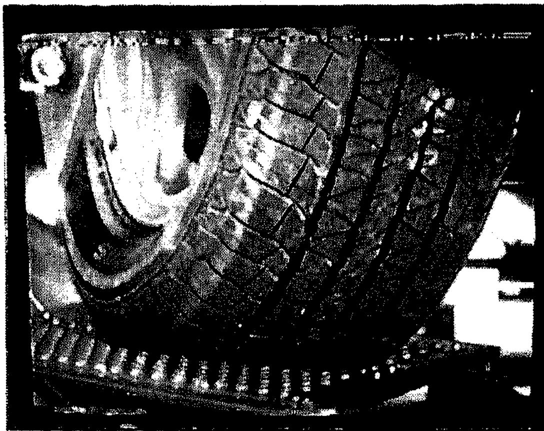


Figure 12: Screen with only binary data

tape uses a "RAM Buffer" module. Through this module, the computer sees the data being produced or reproduced as RAM which is updated with twelve 12-bit data samples, (or high speed equivalents) in 53 microsecond bursts occurring sixty times per second.

CALIBRATION OF INSTRUMENTATION - Wheel scales are calibrated by recording the placing of 24 NBS Precision F (±0.002 percent) test weights (each 50 pounds, approximately 222 N) on the scale, one at a time, then transferring them one at a time to the next scale, etc., plotting the results, and computing the best straight line through the data for each scale.

The CG moment arm load cell and roll inertia springs are calibrated in the same way: by putting the weights on the Frame against a Unistrut rail one at a time, then transferring them one at a time to the same location on the other side and plotting the result.

The small tension/compression load cells are calibrated by screwing several in series with a master load cell used only for that purpose, and loading the composite with a hydraulic cylinder and a triangle wave generator. Individually they are single-point checked with precision test weights. Sidepull load cells are single-point checked by hanging precision 2500 pound and 5000 pound test weights (approximately 11 000 N and 22 000 N), in visits by the New Jersey State Weights and Measures truck.

All load cells generally use identical 9 meter cables, both in calibration and in use. During calibration each load cell has its R-Cal checked against its own history for its output with a Micro-Measurements ±0.02 percent R-Cal resistor. Ordinary one percent resistors are then installed for checking during a test program. If for some reason a non-standard cable is used for a test the precision R-Cal can be used to adjust excitation level to compensate for cable wire resistance.

Amplifiers are checked before each use with an Electro-Services Corp "Transducer Simulator", in 0.5 millivolt/volt increments to 3.0 millivolts/volt.

Calibration of inclinometers are checked periodically on a sine table with machinist's gageblocks.

Before each use string encoders and pots are checked over 450 mm travel by a Starret metal ruler, and rectilinear pots are checked by a vernier caliper.

EQUIPMENT FOR WHOLE-VEHICLE INERTIA PROPERTIES

For measurement of CG height, the Frame sits on strain-gaged knife-edges (Figure 13), with roll motion forced by a ball-screw linear actuator equipped with a 2225 N load cell (Figure 14). The actuator attaches between a transverse torque arm bolted to Unistrut rails on both sidebeams, and the lab ceiling. The actuator motor and position potentiometer connect through a ceiling plug to a wall-mounted speed control box and to the data acquisition system. For CG measurement the actuator forces constant-speed roll motion between reversing limit stops set at ±12.5 degrees.

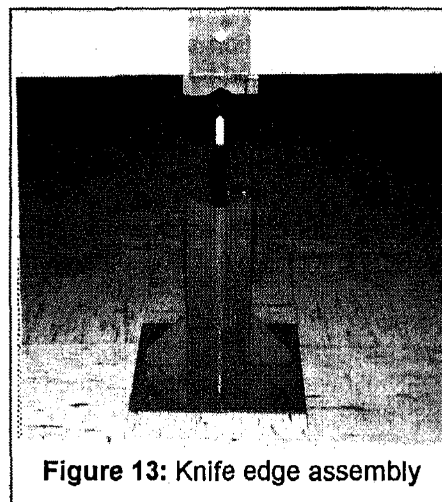


Figure 13: Knife edge assembly

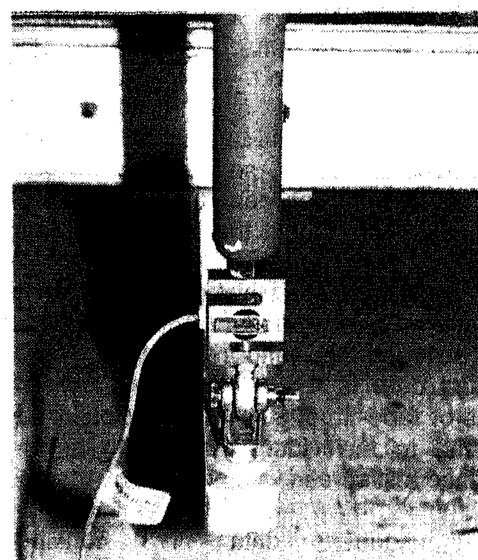


Figure 14: Torque arm with actuator load cell for CG height

An unloaded, rectangular tube, "longitudinal reference beam" is installed between the front and rear knife-edge pads. Transverse channel sections are clamped to the Frame lower-surface Unistrut rails below the front and rear wheelcenters to position vertically-mounted rectilinear potentiometers with their spring-loaded shafts impinging on the longitudinal reference beam, to measure Frame deflection. Vertical string potentiometers are installed between the reference beam and vehicle at front and rear axle locations to monitor ride height during a test.

For roll moment of inertia calibrated coil springs are bolted between the Frame ends and the floor. The Frame rocks back and forth at a frequency of approximately 1 Hz, determined by the spring rates and the inertia of the vehicle-plus-Frame. The knife-edge strain gages measure side forces which result from roll/yaw coupling. The oscillation period is measured by the angular accelerometer.

Pitch moment of inertia is measured in the same way, with the knife-edges relocated to the midpoint of the side beams.

For yaw moment of inertia the Frame sits on a "Yaw Cradle". The "cradle" (Figure 15) is a frame made up of welded rectangular aluminum tubing bolted to the lab floor through a 500 mm Rotek ball bearing. The Rotek bearing was chosen because it has bolt flanges on both inner and outer races, and does not require extreme flatness in mounting surfaces.

The cradle inner and outer races are connected through a torsion bar for angular restraint. The torsion bar is a strain-gaged BMW axle shaft, complete with constant-velocity universal joints. Test frequency is about 0.33 Hz.

When not in use the Yaw Cradle stores vertically, chained to a storage cabinet to ensure against tipping.

Figure 16 shows the storage system for everything except the Frame: calibration weights, scales, turntables, knife edges, Yaw Cradle, and "crossmember corner" are external, and everything else fits inside the cabinets.

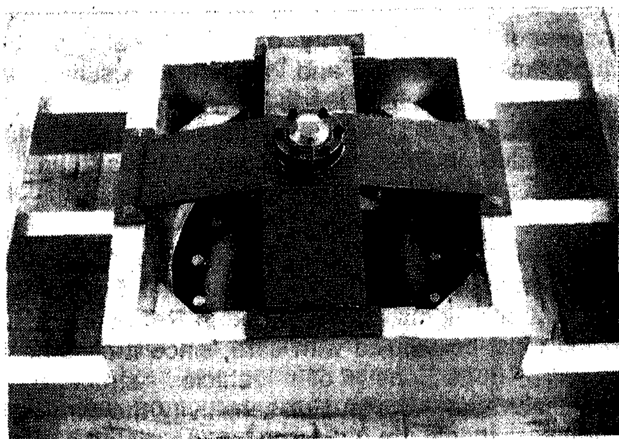


Figure 15: Yaw Cradle



Figure 16: Equipment storage

MEASUREMENT PROCEDURES FOR WHOLE-CAR INERTIA MEASUREMENTS

VEHICLE INSTALLATION - The vehicle is cleaned to remove accumulated dirt and its fuel tank is filled to overflowing; then it is weighed and the longitudinal and lateral location of its CG is computed. The longitudinal reference beam is installed, and the Frame-deflection potentiometers are attached below the calculated wheel center locations. No-load Frame deflection is recorded, and then the vehicle is driven on. The scissors jacks are attached to the vehicle, by bolting to Rivnuts installed in the vehicle underbody or frame or by clamping to rocker weld flanges. The vehicle is rolled to locate its longitudinal CG at the Frame midpoint, then rollers are placed under the scissors jacks. The vehicle is lifted slightly and centered laterally, then it is lowered and the rollers are removed. The scissors jacks are bolted to the Unistrut rails and the vehicle is set to the specified ride height.

Pots or encoders are installed to measure lateral deflection under load and any ride height changes. Supplementary, monitoring vertical stringpots are installed between the Frame and the rocker midpoints, the radiator support crossmember and the rearmost available point on the vehicle structure. All transducers are zeroed.

If moments of inertia are to be measured, scales are placed under the four corners of the Frame elevators, and barbell weights are attached to the Frame ends to adjust the longitudinal Frame CG location to its midpoint, compensating for equipment location. The vehicle is now located so that its CG and that of the Frame are directly over the Frame center.

CG HEIGHT - The Frame is lowered onto the knife edges, the vertical actuator is installed, and then the elevators are totally retracted.

The vertical actuator is set to cycle through its ±12.5 degrees range at its "fast rate" of 1 minute per cycle, while the data are observed. Any excessive deflections, or shifts as motion goes through center, require correction. The front and rear inclinometers must read the same, and they must agree with the potentiometer in the vertical actuator. The video recorder is turned on, and the actuator is then set to its 3 minutes per cycle "test rate" for at least six additional cycles. If cycle-to-cycle transducer readings are identical in each cycle, the test run is over.

If desired, the ride height is then reset, and the procedure is repeated.

After all vehicle-on inertial testing is completed the location and heights of the scissors jacks is noted, and they are removed. The vehicle is removed, and the vehicle-off readings of the Frame deflection potentiometers are recorded.

The scissors jacks are re-installed on the Frame, the gain of the load cell signal conditioner is switched to a 10x higher sensitivity, all vibration-creating devices (overhead fan, any rotating machinery) are turned off, and the measuring process is repeated for computation of the Frame tare.

Data processing, as described in Appendix 1, consists of plotting the restoring force and vehicle lateral deflection versus the tangent of the tilt angle over 3 cycles, and computing the slopes for the vehicle-plus-Frame and for the Frame alone. The computation is summarized in the following equation:

$$H_v = \frac{R}{W_v}\left[\frac{dF_{V+F}}{d\tan\theta} - \frac{dF_F}{d\tan\theta}\right] - \frac{dY_s}{d\tan\theta}$$

where

H_v = height of the total vehicle CG above the knife edges

R = torque arm length

W_v = weight of the total vehicle

θ = roll angle

$dF/d\tan\theta$ = slope of the force/angle relationship for the vehicle-plus Frame and for Frame alone

$dY_s/d\tan\theta$ = slope of the lateral shift/tilt angle relationship of the vehicle CG

Typical data plots are shown in Figure 17. Forces are corrected before plotting as discussed in Appendix 1. A single slope is shown for $dF_{V+F}/d\tan\theta$ in Figure 17; however, an incompletely-filled fuel tank will cause a small, but noticeable local slope increase through the zero tilt angle point. For this reason we also compute individual slopes for positive and negative tilt angles, eliminating the center ±0.5 degree.

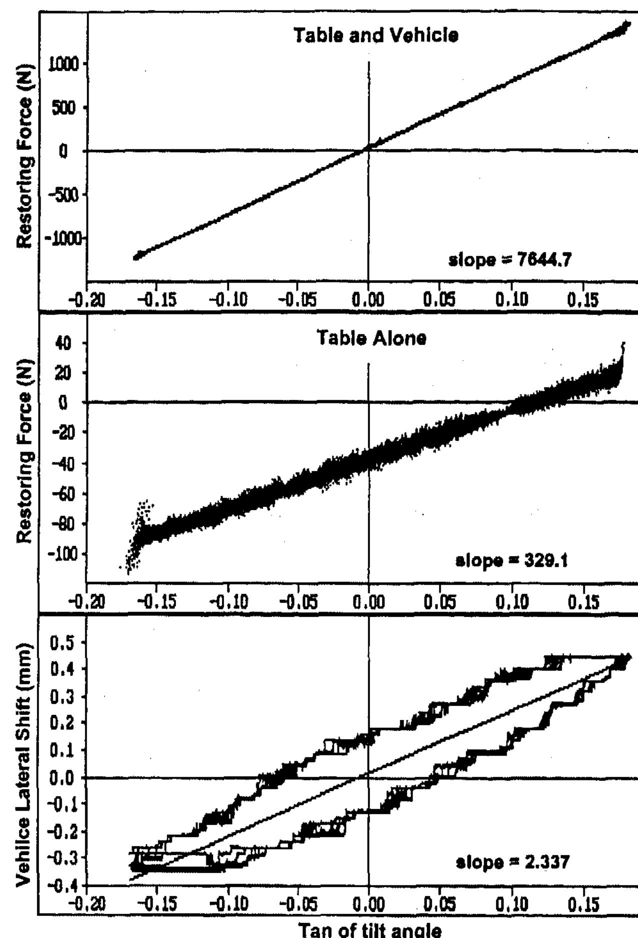


Figure 17: CG Height Test

The height of the "road surface" above the knife edges, including the measured Frame deflections, is then subtracted to obtain the total-vehicle CG height above the road surface.

CG height is usually measured at curb weight, and at additional trim heights to encompass loading conditions. We prefer not to test at different loads, because of the complexity of locating and securing "equivalent dummies" and adjusting resulting ride heights. We believe that loaded CG height can better be computed from the spring deflections due to added load, the individual heights of any applied load, and the vehicle CG measured at different ride heights.

The CG heights of sprung and unsprung masses can be separated by measuring the total CG height at different trim heights, as described in Reference 3. This method must be utilized with care, since it assumes no resulting deflections of vehicle structure or engine/subframe mounts. For sensitivity the trim height differences should be as large as possible: however,

too-large differences can create nonlinear displacements of heavy components which in some vehicles may create significant errors. Especially, unloading to the point where independent suspension rebound stops are engaged should be avoided. Since it is easy to re-measure at different trim heights, we feel it advisable to measure at three trim heights encompassing the loading range, plus one about 2.5 cm in rebound, and plot the result to evaluate the trends. Since it is also easy to add stringpots or encoders, it is advisable to monitor deflections of the engine cradle and the body structure, and make corrections if necessary.

ROLL MOMENT OF INERTIA AND ROLL/YAW PRODUCT

THEORY - In the test setup, the pitch axis is considered as being perpendicular to the plane of symmetry. The system is then one of two degrees of freedom for which the equations of motion are:

$$I_{xx}\ddot{\phi} - I_{xz}\ddot{\Phi} = L$$
$$-I_{xz}\ddot{\phi} + I_{zz}\ddot{\Phi} = N$$

where

I_{xx} = moment of inertia about the roll axis

I_{zz} = moment of inertia about the yaw axis

I_{xz} = product of inertia

$\ddot{\phi}$ = angular acceleration in roll, deg/sec^2

$\ddot{\Phi}$ = angular acceleration in yaw, deg/sec^2

L = Moment about roll axis

N = moment about yaw axis

The knife edges prevent any yaw displacement; therefore the equations of motion reduce to:

$$I_{xx}\ddot{\phi} = L$$
$$-I_{xz}\ddot{\phi} = N$$

ROLL MOMENT OF INERTIA - The Frame, with the vehicle installed and centered as described above, is lowered onto the knife edges, and calibrated coil springs are bolted to the corners of the Frame and to the laboratory floor. The angular accelerometer is attached to the vehicle. A string encoder is installed to measure roll amplitude. The knife edge strain gages are connected. The elevators are completely retracted. The recorder is turned on, and the Frame is oscillated by hand to about ±6 degrees amplitude and released. Settling time for the resulting oscillation is around 150 periods of about one second each. Transducer data is fed to the computer for immediate generation of stripcharts. These are checked to identify any glitches or inconsistencies. If everything looks good the vehicle test is over.

The test is repeated for the Frame and vehicle restraint equipment alone.

The moment of inertia of the total vehicle and Frame about the knife edges is found by equating the moment exerted by the springs to the angular acceleration. The moment of inertia of the sprung mass about a horizontal axis through its CG is then obtained by subtracting the moments of inertia about the knife edges of the Frame and the various unsprung masses, and then correcting this result to the sprung mass CG by the Parallel Axis Theorem.

$$Ikk_T = (4Ka^2 - W_T h_T)(P/2\pi)^2$$

$$Ixx = Ikk_T - Ikk_F - Ikk_{UM} - (W_S/g)h_S^2$$

where

Ixx = Moment of inertia of sprung mass about a horizontal axis through its CG

Ikk = Moments of inertia about knife edges of vehicle plus Frame, Frame alone, and Unsprung Masses

a = Lateral distance from knife edge to centerline of spring

K = Rate of each spring

h_T = height of Vehicle & Frame CG above knife edges

h_S = height of vehicle sprung mass above knife edges

W_T = total weight on knife edges

P = Period of oscillation

ROLL/YAW PRODUCT OF INERTIA

When the frame is oscillating about the horizontal knife edge roll axis, the lateral acceleration of the mass CG produces a side force on the knife edges, and the angular acceleration produces a yawing moment which is reacted by the knife edges. The product of inertia is found from the yaw moment by the following equation:

$$Ixz = (T/\phi_0)(P/2\pi)^2 - Ixz_{um} - (W_s/g)(x_{st}h_s)$$

where

Ixz = product of inertia about longitudinal and vertical axes through the cg of the sprung mass

T = moment reacted by the knife edges

ϕ_0 = roll displacement, in radians

$P =$ oscillation period, in seconds

Ixz_{um} = estimated product of inertia of unsprung mass with respect to the knife edge axis and the vertical axis through the CG of the test configuration

$W_s =$ weight of sprung mass

$g =$ gravitational acceleration

$x_{st} =$ longitudinal distance of sprung mass cg from cg of test configuration

$h_s =$ height of cg of sprung mass over knife edge axis

Note that the Frame is symmetrical, and so does not contribute to roll/yaw moment.

PITCH MOMENT OF INERTIA - The knife edges are located under the midpoints of the side beams, and the angular accelerometer is installed to measure pitch motion. Otherwise the procedure is identical to roll.

YAW MOMENT OF INERTIA - The elevators are raised to full height, the longitudinal reference beam is removed, and the yaw cradle is slid underneath and bolted to the floor. The Frame is lowered onto the Cradle, centered by mating wedges on the Cradle and Frame, and the elevators are fully retracted. The Frame is bolted to the Cradle. The yaw accelerometer is attached to the vehicle, a linear accelerometer is attached to the end of the Frame, the cradle's torsion bar strain gage is connected, a yaw amplitude encoder is installed, and the recorder is turned on. The Frame is oscillated manually in yaw to about ±10 degrees and released. (Settling time is 10 - 12 cycle periods of about 3 seconds each). Transducer data is fed to the computer for immediate generation of stripcharts. These are checked for glitches or inconsistencies in torsion bar or accelerometer output signals, and for vehicle movement with respect to the Frame. Corrections are made if necessary. Five or six data runs are then made. After removal of the vehicle, the test is repeated for the Frame alone, to obtain the tare value.

The moment of inertia is computed from the amplitude and frequency of the oscillation and the torsional spring rate, or from the angular acceleration and the spring torque.

DRIVETRAIN INERTIA - Simple pendulums consisting of 23 kg cylindrical weights on 610 mm rods, attached to the wheelhubs are used to measure rotating inertia.

EQUIPMENT FOR KINEMATICS TESTING - The Ride/Roll vehicle location crossmembers are used, with hydraulic cylinders in double-ended/equal area configuration.

The front Ride/Roll crossmember is rigidly attached to the Frame. The rear crossmember is attached to the Frame through spherical bearings, which permit a 5-degree fore/aft rocking motion to accommodate "cosine foreshortening" when the vehicle is pitched.

The vehicle sits on the bearing-equipped scales, restrained only by the two ball-bearing vertical sliding shafts. Those shafts are unrestrained in ride, and may be unrestrained or locked in roll.

If unrestrained in roll the vehicle "finds its own" roll centers, which can be located by measuring the vertical displacement of the restraint shafts, or by wheelpad displacement or force equilibrium methods. If used, each vertical restraint defines a point on the roll axis, which can be located from 25 to 100 mm below the bottom of the rocker panels, and to the ground if spacers are added.

Wheel adapter plates are attached to lugnut extensions, then the transducer plates are added and leveled. The transducer plates carry camber and caster inclinometers, and rearward-facing lasers for setting wheels in the straight-ahead position. They mate to the ball-slide wheel follower mechanism.

The ball-slide wheel followers are attached to the vehicle, on beams clamped to the Ride/Roll upper crossmembers. Tire deflections are measured separately, by an encoder attached between a ball bearing on the transducer plate axis, and the wheelpad.

For parallel ride motion the cylinders are plumbed in parallel, with all dummy blocks blue, with the two front cylinders driven by one pressure control valve and the two rear cylinders by another. The front valve is driven by a triangle-wave generator, and the rear nulls an inclinometer to keep the vehicle level. For roll the dummy valve blocks on one side are changed to red, and both valves are driven by the triangle generator for equal front and rear moments.

KINEMATICS TEST PROCEDURE - The Ride/Roll Crossmembers are installed, scales fitted with ball-bearing pads are located under each wheel, and the various displacement transducers are attached. The instrumented "second steering wheel" is installed, and the steering wheel angle is set for straight-ahead road wheels and fixed. The brakes are locked ON.

The vehicle sprung mass is exercised by hydraulic servos through several cycles of jounce, pitch, and roll motions from compressed bump stops to wheel-off, at about one cycle per minute. Throughout each run measurement is made of wheel loads; steer angle; displacements of wheel centers, tires and shock absorbers; wheelpad rotation and lateral translation; and anything else that may be of interest. Test runs are normally made with anti-roll bars both connected and

disconnected. All data is recorded for each run on the test videotape.

In data processing all data is plotted against time and inspected for smoothness and consistency. Three cycles of all desired data are plotted against the selected motion variable. Third-order least-square equations are fit to data plots except spring rates, where segmented first-order fits are usually required.

If roll centers are defined as the center of rolling motion in the absence of side forces, their heights can be computed from wheelpad displacements in ride or in roll [2], or from the vertical motion of unlocked lateral restraint shaft.

Figure 18 shows typical results for steer angle, camber, and wheelcenter Y-deflection. Figure 19 shows typical results for a roll test.

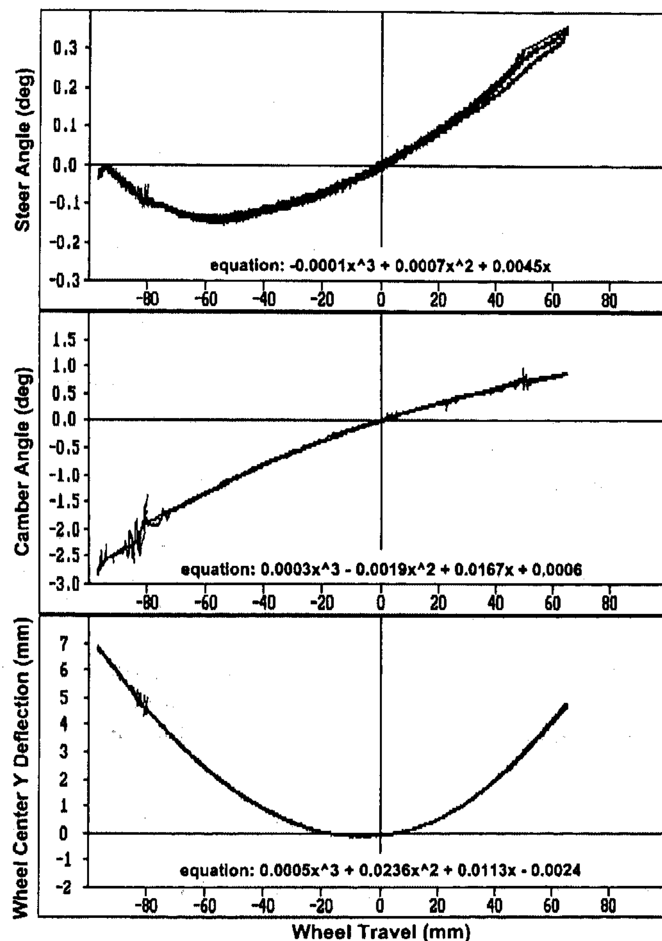


Figure 18: Ride Kinematics Test

EQUIPMENT FOR COMPLIANCE MEASUREMENTS - Either the Ride/Roll or Scissors Jack crossmembers can be used. If the Ride/Roll crossmembers are used, the hydraulic cylinders are locked in position by rod-mounted shaft collars to prevent vertical motion.

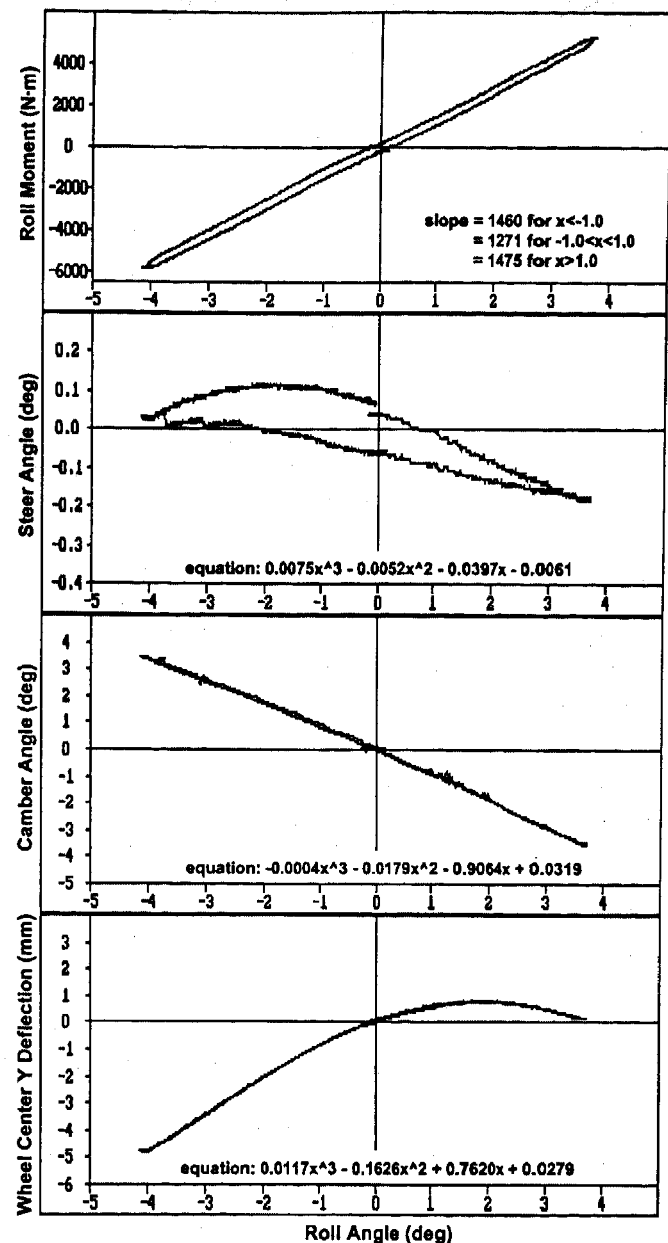


Figure 19: Roll Kinematics Test

The vehicle sits on the wheelscale-bearing plates. A set of four servo cylinders fitted with load cells are mounted on two transverse Unistrut crossmembers, and attached to mounting ears on the bearing pads, as shown in Figure 20. The four cylinders are driven by a single pressure-control servovalve, so that their force outputs are all identical. A small amount of 40 Hz dither is added to the servovalve signal to eliminate friction. Red and blue dummy valve blocks determine the direction of force output from each cylinder, to create "aiding" or "opposing" forces or moments. The servovalves are driven by a triangle wave generator.

LATERAL FORCE AND MOMENT COMPLIANCES - The vehicle is secured to the frame by either the Ride/Roll or the Scissors jack method at the desired trim height. Transducer plates and ball-slide wheel followers are used as in kinematics

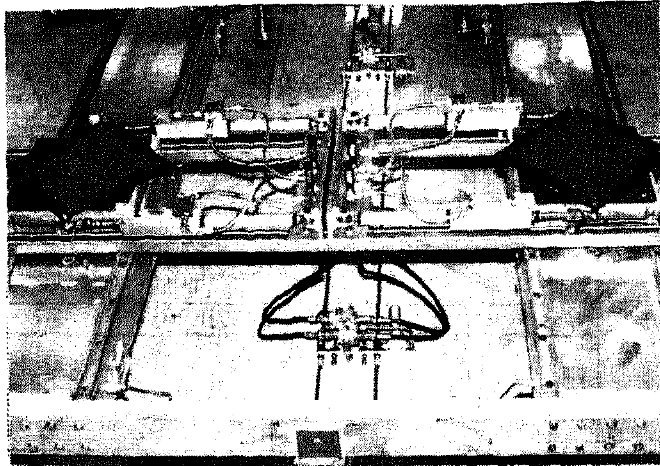


Figure 20: Wheelpad forcer assembly for compliance measurements

measurements. The wheel followers are attached to the vehicle. The force-measuring "second steering wheel" is attached to the vehicle handwheel and fixed by struts and suction cups to the vehicle windshield. The brakes are locked.

Tires are carefully centered on the wheelscale bearings, using a plumb bob or vertical laser for longitudinal and lateral location. A check run is made in "toe in/toe out" configuration to insure that sideforce produced by torque is negligible. The plumb bob or laser point is marked on the wheelpad surface, and the servo cylinder shaft extensions are noted, to restore position after wheelslip.

Test runs are made with aligning torques aiding and opposing, and sideforce aiding and opposing. In each configuration runs are made at incrementally increasing peak force levels, up to wheel-slip. Each run consists of at least four cycles at a rate of one cycle per minute.

When wheelslip occurs, the servos are used to power the tire back into place.

Steer angle and X,Y,Z wheelcenter displacements are measured from the ball slide follower. Camber and caster changes are measured by the accelerometer-inclinometers on the wheelhub transducer plates. Force and moment inputs are measured by load cells on each of the forcing servo cylinders. The cylinder differential pressure is measured as a check of measured forces.

Force and pad displacement inputs are plotted against time and inspected for quality and force/moment crosstalk. Three cycles of all desired data are plotted against input forces or moments, and third-order least-squares equations are fitted to the plots to obtain slopes. Typical plots, for an independent rear suspension measured in "sideforces opposing" configuration, are shown in Figure 21.

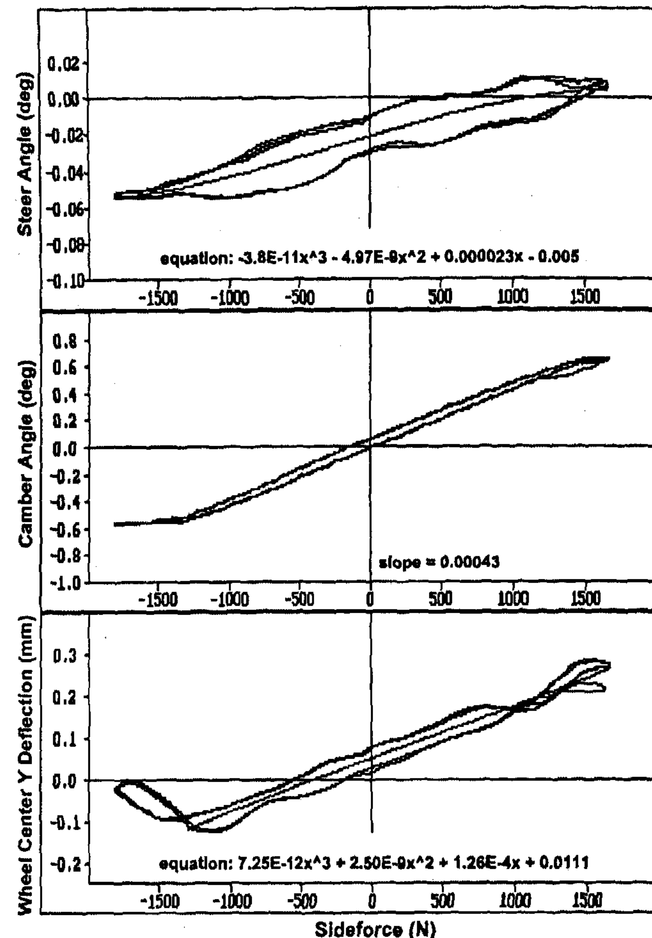


Figure 21: Typical compliance data

If roll centers are defined as points about which forces balance at zero roll angle, their heights are computed from the ratio of vertical load change to applied horizontal force in the sideforce-opposing test.

LONGITUDINAL FORCE AND MOMENT COMPLIANCES - The combined wheelpad- forcer servos are rotated 90 degrees to generate longitudinal forces. Longitudinal forces opposing at each pad produce the same toe in/out and steer moments as in the lateral case. All-forces-aiding measures longitudinal compliance, and one wheel aiding with the opposite opposing simulates unbalanced braking or thrust.

Data plotting is the same as in the lateral case.

STEERING SYSTEM MEASUREMENTS

COMPLIANCE - The compliance of the entire steering system including gearbox and column is measured in the sideforces and aligning torques aiding configuration. Sideforces and aligning torques opposing tend to balance at the steering gear, and so do not load the column. With power steering the only column torque is that required to twist a small torsion bar input to the valve: therefore with the engine running, aiding and opposing results are similar. Figure 22 demonstrates this characteristic. With the engine running in the

"Aligning Moments Aiding" (i.e., steer) configuration the on-center steer compliance, before full boost occurs, is similar to that with "engine off, moments aiding"; while the off-center compliance is similar to the "engine off, moments opposing" configuration.

Figure 23 shows the torque reaction at the steering wheel resulting from total aligning moment (aiding) at the tire contact patch, with engine running and off.

STEERING GEOMETRY - Steering ratio is measured by the rotary encoder mounted between the ball slide follower mechanism and the wheelhub transducer plate. The encoder has a resolution of 0.05 degrees at 4X.

If ground-level steer axis offsets are required, instrumented turntables are used. These are standard units (Mac Tools, Inc.) to which rotary encoders and rectilinear potentiometers have been added. They were formerly fitted with ears for the hydraulic cylinders and used on top of the wheelscales in compliance testing, but they suffered from surface brinelling and were too high, and so they were replaced by the bearing pads. They directly replace the wheelscale/bearing pad units, and measure both steer angle and pad X and Y displacements. The encoder has a resolution of 0.05 degrees, and that of the cermet pots is infinite.

In testing, the instrumented handwheel is turned through its lock-to-lock range, while the outputs of steer angle, pad displacement, camber and caster inclinometers are recorded.

For computation of steering ratio both least-squares and trigonometric equations are fit to the data, to describe the basic linear or nonlinear ratio and the effects of column universal joints. Steer angles of "inside" and "outside" wheels are plotted against handwheel angle, along with Ackerman angles, and Ackerman errors are computed.

Static camber, caster, and steering axis inclinations are computed from the equations presented in Reference 4. Caster is computed by variation of the camber transducer output with steer, and steering axis inclination is computed from variation of the caster transducer output with steer. Steering axis offsets at ground level can be computed from the rotation and translation of the wheelpad or the turntable center, provided that the location of the tire is known with sufficient accuracy. Wheelcenter displacements and steering axis offsets at spin axis level can be computed from wheel follower measurements along with wheel rim offset measurement.

COMBINED KINEMATICS AND COMPLIANCE TESTING - Topsy's mechanization of kinematics and compliance tests are essentially independent. For that reason "partial derivative" testing - varying body motion or wheel loading at constant tire patch forces or moments or *vice-versa* - is simple. For simultaneous variation of kinematics and compliance input variables it is necessary to burn an EPROM with the control program for each servo and clock their addresses in

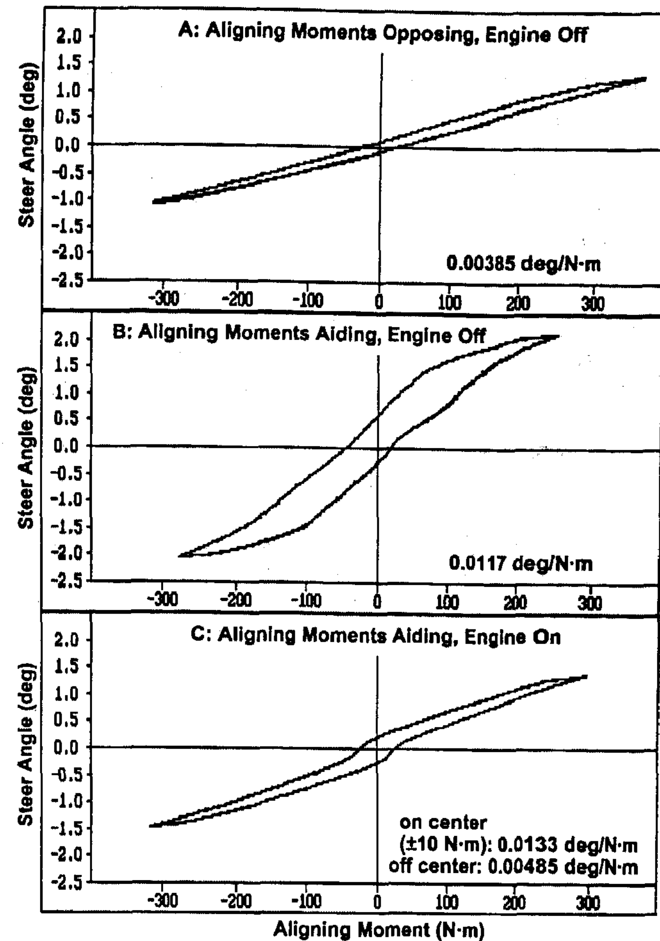


Figure 22: Steer Compliance

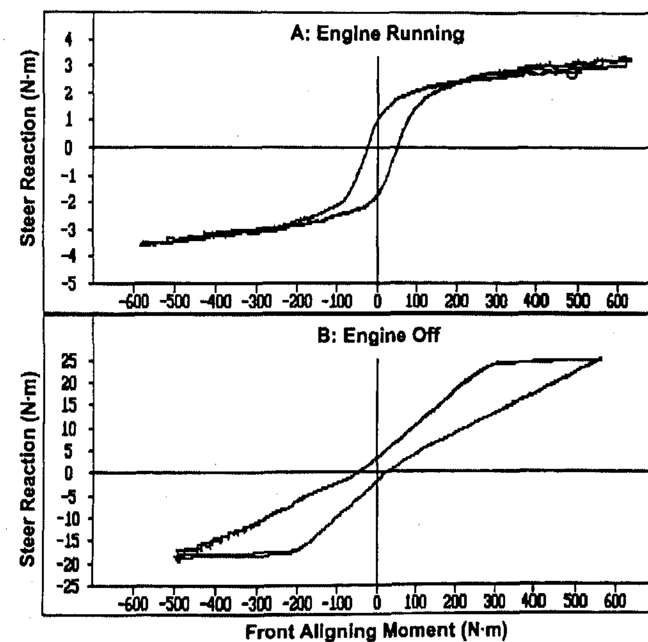


Figure 23: Torque reaction at steering wheel

unison - the electronics hardware required has been developed in another project [5].

SHOCKS AND STRUTS - Shocks, springs, and struts are measured on the vertical baseplate by simple restraining fixtures containing a load cell and a servo cylinder. The cylinder is driven at selected constant rates by a triangle-wave generator. A typical spring-rate result is shown in Figure 24 for a gas-pressurized strut. Determination of one point of a damping curve is shown in Figure 25. At each constant velocity the force is a function of both piston velocity and compression/expansion of the pressurized gas. Damping values are taken at zero deflection.

FREQUENCY RESPONSE METHODS - Frequency response methods (in present software development) utilize the servo cylinders of 1.3 or 2.6 cm^2 area, fitted with load cells, in the ride/roll crossmembers to generate sinusoidal ride forces and roll moments. The cylinders produce continuous off-resonance 2.5 - 5 cm peak-to-peak motion to 3 times resonant frequency. The software computes sprung mass inertial parameters from measured forces, accelerations, and displacements.

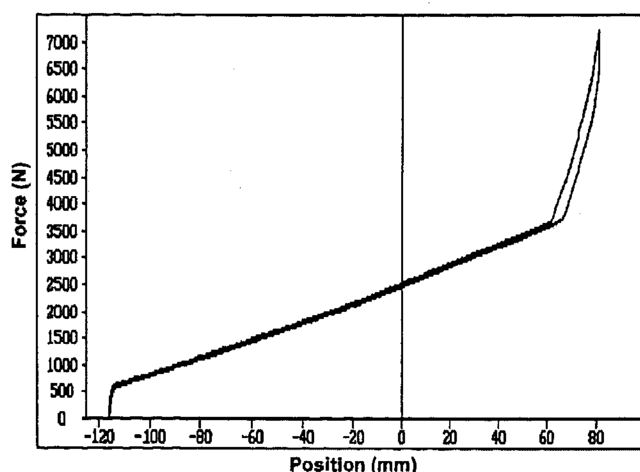


Figure 24: Strut spring rate

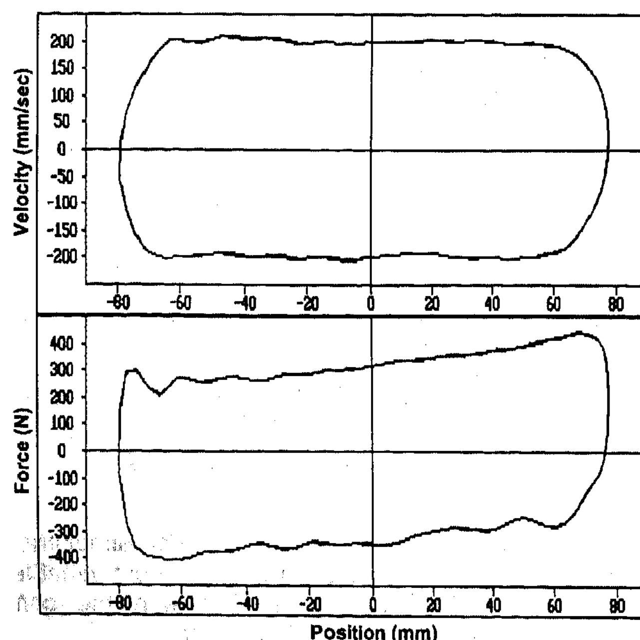


Figure 25: Strut damping

Measurements on the unsprung masses are more difficult, because of tire variations. A tire-supported wheel hub can be oscillated vertically to 30 Hz, and the static spring rate of the combination of tire and suspension spring is measured in kinematics testing. However, non-rolling and rolling tires are significantly different [6], and even the non-rolling tire spring rate is known to double between 0 and 30 Hz [7], so the "easy way" appears to have little value. If the tire is eliminated, the vehicle weight must be carried by bias pressure in the hydraulic cylinders. One system under consideration uses two servo cylinders supporting a beam attached to the wheelhub, replacing the wheel and tire.

Steering system dynamics can be studied by applying sinusoidal steer moments to the wheelpads, up to at least 30 Hz. The handwheel can be fixed with its torque measured by the "second steering wheel"; or it can be free with steering wheel motion measured by a rate gyro or angular accelerometer. The relationship between wheel angles and handwheel can be measured, but the nonrolling tire precludes moment-input relationships.

Some lateral tire dynamics can be estimated simply from lateral tire spring rates as measured in compliance tests, as described in References 8 and 9.

SPECIFYING AND REVIEWING A TEST PROGRAM - Specification of a test program consists of selecting the data required and the form in which it is desired. Identification codes are given in compressed form in Appendix 2. A test specification consists circling the appropriate codes in the complete version contained in Reference 10.

The primary record for each test program is a series of two-hour videotapes. From this record all other data can be generated.

For all test runs raw data (vs time) is supplied to the customer on floppy disks, in his choice of formats. Some or all plotting and curve fitting can be done by the customer to save costs, or it can be done by Heitz Chassis Lab. All data is referenced to hours, minutes, and seconds on the test videotapes, so an engineer can "watch" the test run from which each data plot came.

A typical test program will generate much more data than is actually used. The excess consists of deliberate redundancies in which a parameter is measured in different ways for comparison, and of "checking data" - hydraulic cylinder pressures as well as forces; various vehicle or component deflections; etc.

This information is generated for observation during a test, and is recorded under the adage "It is better to have it and not need it than to need it and not have it." This "excess data" remains available for review if questions arise, and is supplied on floppy disk at customer request.

The voice commentary, miscellaneous sounds of the test, and video camera view, all in synchronism with the instrumentation data, are additional forms of data which remain available for reviewing the testing procedures.

SUMMARY

This paper has described a compact laboratory for measurement of the inertial, kinematic, and compliance parameters required to model automobile control responses. The laboratory is oriented toward research rather than production. The approach is modular, based on an "infrastructure" of multipurpose components which are combined temporarily with a few special-purpose fixtures as required for each type of test. The modular approach permits addition of new capabilities without extensive redesign. The authors believe that in some instances the flexibility afforded by the modular approach may allow test capabilities not possible in "production" facilities. When not in use the facility is stored, so that its space is available for other projects.

REFERENCES

[1] Nedley, A.L., and Wilson, W.J., "A New Laboratory Facility for Measuring Vehicle Parameters Affecting Understeer and Brake Steer," SAE Paper 720473, National Meeting, May 1972.

[2] SAE J1574 - Measurement of Vehicle and Suspension Parameters For Directional Control Studies: Test Equipment and Procedures.

[3] Bixel, R.A., Heydinger, G.J., Guenther, D.A., Nevak, S.J., "Sprung/Unsprung Mass Properties Determination without Vehicle Disassembly", SAE Paper 960183, SAE SP-1141, International Congress, February 1996.

[4] January, Daniel B, "Steering Geometry and Caster Measurement", SAE Paper 850219, International Congress, February 1985.

[5] Heitzman, E.J., and Heitzman, E.F., "A Programmable Steering Machine for Vehicle Handling Tests", SAE Paper 971007, SAE Congress, March 1997.

[6] Richard E. Rasmussen and Anthony D. Cortese, "Dynamic Spring Rate Performance of Rolling Tires," SAE Paper 680408, Midyear meeting, May, 1968.

[7] Tsymberov, Anatoly, "An Improved Non-Intrusive Automotive Suspension Testing Apparatus with Means to Determine the Condition of the Dampers", SAE PAPER 960735, International Congress, February 1996.

[8] Chiesa, A., and Rinonapoli, L., "Vehicle Stability Studied with a Non-Linear Seven Degree Model", SAE Paper 670476, Mid-Year Meeting, May 1967.

[9] Winsor, F.J., "Effect of Tire Lateral Elasticity on Vehicle Dynamics", Automobile Dynamics Research Working Paper No.19, Princeton University Department of Aerospace and Mechanical Sciences, June 30, 1967.

[10] Internet http://home.navisoft.com/atiheitz

APPENDIX 1: DERIVATION OF EQUATIONS FOR CG CALCULATION

Figure A1 shows the table/vehicle assembly in untilted and tilted configurations. For the tilted case, taking moments about the knife edge, we obtain:

$$\begin{aligned} FR\cos\theta = {} & [W_vH_v \tan\theta \cos\theta] \\ & + [W_v(Y+Y_s)\cos\theta] \\ & + [W_tH_t \tan\theta \cos\theta] + [W_tY_t \cos\theta] \end{aligned} \quad (A1)$$

Dividing both sides by $\cos\theta$ and solving for $(W_vH_v \tan\theta)$,

$$\begin{aligned} W_vH_v \tan\theta = {} & FR - W_vY - W_vY_s \\ & - W_tH_t \tan\theta - W_tY_t \end{aligned} \quad (A2)$$

Differentiating, noting that Wv, Wt, Yt, Yv, R, Hv, Ht are constant, while θ, F, Ysh vary;

$$W_vH_v\, d\tan\theta = R\,dF - W_v dY_s - W_tH_t\, d\tan\theta \quad (A3)$$

Rearranging to solve for Hv,

$$H_v = \frac{R}{W_v}\frac{dF}{d\tan\theta} - \frac{W_v}{W_v}\frac{dY_s}{d\tan\theta} - \frac{W_tH_t}{W_v} \quad (A4)$$

For the table alone, denoting the force by F';

$$F'R\cos\theta = W_tH_t \sin\theta + W_tY_t \cos\theta \quad (A1a)$$
$$F'R = W_tH_t \tan\theta + W_tY_t \quad (A2a)$$
$$W_tH_t \tan\theta = F'R - W_tY_t \quad (A3a)$$

Differentiating as above and rearranging;

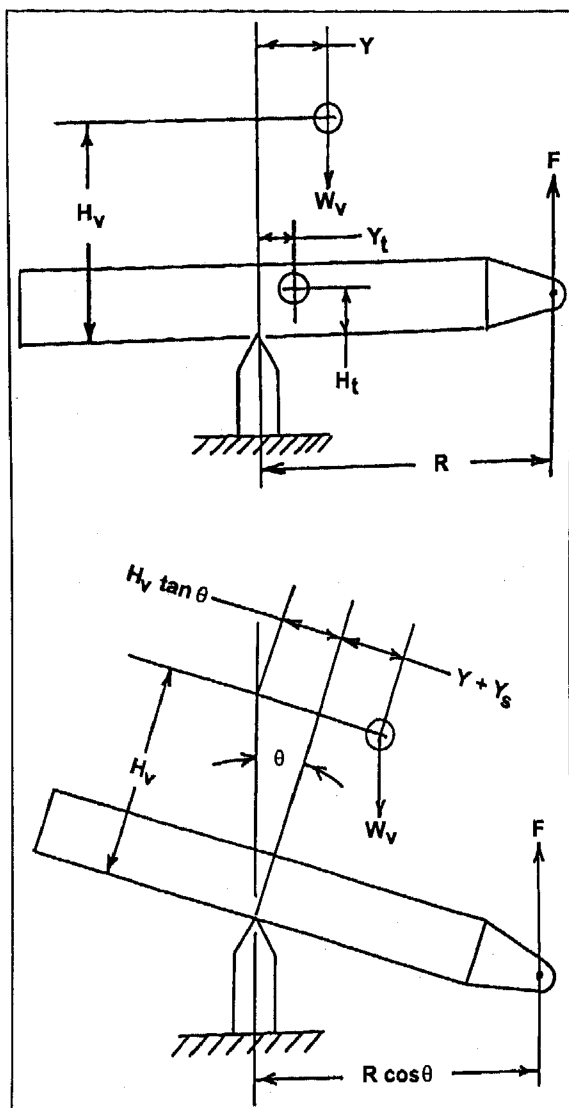


Figure A1: CG Geometry

$$H_t = \frac{R}{W_t}\frac{dF'}{d\tan\theta} \qquad \text{(A4a)}$$

Substituting this value for Ht into equation (A4);

$$H_v = \frac{R}{W_v}\left[\frac{dF}{d\tan\theta} - \frac{W_t}{W_t}\frac{dF'}{d\tan\theta}\right] - \frac{dY_s}{d\tan\theta} \qquad \text{(A5)}$$

H_v, the height of the CG above the knife edge is thus found from the slopes of the total assembly and table alone force/tilt angle relationship, plus that of the lateral vehicle displacement with respect to the table. The weight of the table itself drops out of the equation: however, the force/tilt angle relationship for the table alone must be measured with all ancillary equipment such as locating screwjacks and potentiometers in the same locations as for the vehicle test.

CORRECTION TO PRIMARY DATA - In the derivation of equation (A5) the effect of angularity in the

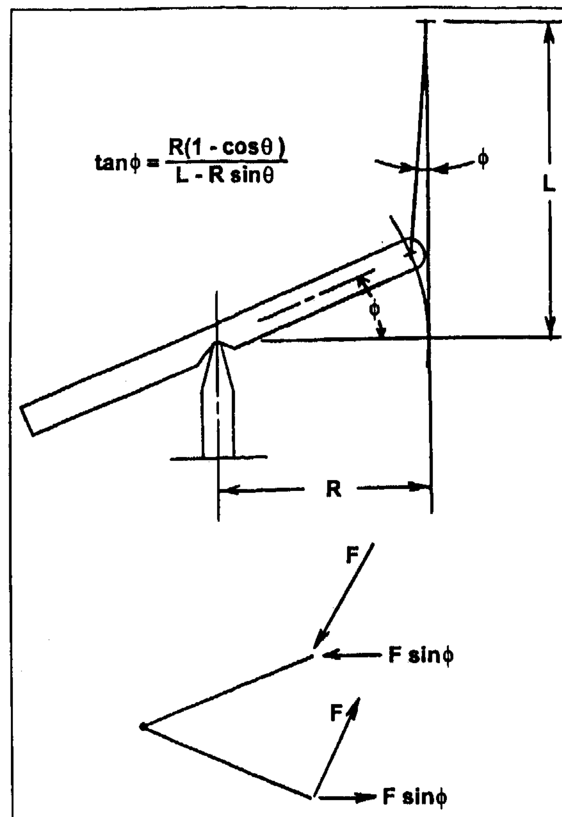


Figure A2: Sine Error

vertical actuator is ignored. This angularity does cause a small error in the slope dF/d(tan θ) for positive and negative tilt angles, with the slope being lower in the "lifting" or "raising" direction. The resulting cosine error is 0.005 percent, but sine error is in the order of 0.1 percent. They are computed and corrected as follows.

Figure A2 shows a shortening of the torque arm, R(1-cos θ), which together with the actuator length (L - R sin θ) forms the angle ϕ. This angle amounts to 0.56 degrees in the raising direction and 0.45 degrees in the lowering direction.

The resulting cosine error, which is always in the same direction, is corrected by computing ϕ and multiplying the load cell output signal by (cosine ϕ). However, correction of the resulting sine error is a little more complex.

In Figure A2 it can be seen that in the raised position the actuator is in compression, producing a clockwise moment equal to (FR cos θ) about the knife edges; while the sine component produces a moment (F sin ϕ) x (R sin θ) which opposes that moment.

In the lowered position the actuator is in tension, producing a counterclockwise moment around the knife edges. However, in this case the sine component produces an added, counterclockwise moment.

The correction for the moment (FR $\sin\phi$ $\sin\theta$) is made as follows. For small enough angles sine and tangent can be assumed equal, so that:

$$F \sin\phi = F \tan\phi = \frac{FR(1-\cos\theta)}{L - R\sin\theta} \quad \text{(A6)}$$

and the moment,

$$FR \sin\phi \sin\theta = FR^2 \sin\theta \frac{1-\cos\theta}{L - R\sin\theta} \quad \text{(A7)}$$

In equation (A1) the moment FR cos θ must be corrected by (A7). Dividing through by (R cos θ), we get a corrected value for force F.

$$F_{corr} = F\left[1+\frac{R(1-\cos\theta)}{L - R\sin\theta}\tan\theta\right] \quad \text{(A8)}$$

With this correction, and with the gas tank filled to overflow to limit fuel slosh, the difference in slopes between positive and negative tilt angles is eliminated.

CG ERROR ANALYSIS - The measurement of CG height is seen from equation (A5) to consist of four separate measurements: the force/angle slopes with and without the vehicle, the lateral deflection/angle slope, and the "road" height above the knife edges. The maximum error in the resultant CG height is therefore found as the weighted sum of the four individual measurement errors. The weighted error contribution from each source is found as the percentage error in that measurement, multiplied by the contribution of the measurement to the net CG height. The percentage errors in the individual measurements is found by differentiation, as follows.

For the contribution to total error of table plus vehicle, from equation A5,

$$\frac{R}{W_v}\frac{dF}{d\tan\theta} \quad \text{(E1)}$$

At the end points, for "small angles", this equation becomes:

$$\frac{R}{W_v}\frac{F}{\theta} \quad \text{(E2)}$$

Differentiating,

$$dH_v = \frac{R}{W_v}\frac{dF}{d\theta} + \frac{R}{W_v}\frac{Fd\theta}{\theta^2} + \frac{dR}{W_v}\frac{F}{\theta} + \frac{RdW_v}{W_v^2}\frac{F}{\theta} \quad \text{(E3)}$$

Multiplying and dividing by $\frac{R}{W_v}\frac{F}{\theta} = \frac{R}{W_v}\frac{dF}{d\theta}$

$$(\delta H_v)_{v+t} = \frac{R}{W_v}\frac{dF}{d\theta}\left[\frac{\delta F}{F} + \frac{\delta\theta}{\theta} + \frac{\delta R}{R} + \frac{\delta W_v}{W_v}\right] \quad \text{(E4)}$$

Similarly, for the table alone,

$$(\delta H_v)_t = \frac{R}{W_v}\frac{dF_t}{d\theta}\left[\frac{\delta F_t}{F_t} + \frac{\delta\theta}{\theta} + \frac{\delta R}{R} + \frac{\delta W_v}{W_v}\right] \quad \text{(E5)}$$

For lateral deflection,

$$(\delta H_v)_y = \frac{dY_s}{d\theta}\left[\frac{\delta Y_s}{Y_s} + \frac{\delta\theta}{\theta}\right] \quad \text{(E6)}$$

And for the road - knife edge distance

$$(\delta H_v)_h = \delta H_v \quad \text{(E7)}$$

The total error is

$$\sigma H_v = (\sigma H_v)_{v+t} + (\sigma H_v)_t + (\sigma H_v)_y + (\sigma H_v)_h \quad \text{(E8)}$$

The individual percentage errors are found from the full-scale linearity-plus-hysteresis specifications of all transducers, along with the measured end-point parameter values (to obtain percent-of-reading), and measurement resolution.

The total error is dominated by the $(\sigma Hv)v+t$ term and by the "roadway height" over the knife edges, because their "weight" in determination of Hv is greatest. These errors are minimized by calibrating the combination of load cell force and moment arm together, and by measuring the height of the "road surface" over the knife edges directly under the front and rear axles, before each test. In the latter measurement the deflection curve of the simply-supported reference beam under its own weight (1.0 mm at center span) is included in the measurement.

The maximum errors calculated from equation (E8), with all errors stacked up in the same direction, are in the order of 0.6 percent. Rms errors are in the order of 0.3 percent. We believe these estimates to be realistic, because the individual calibrations encompass the testing range; the quantity (Ft x R) is actually a single calibration; hysteresis error is minimized by running three completely reversed test cycles; and resolution error is minimized by taking the best straight line through the resulting 60 000 data points.

In practice we believe the greatest error is most likely to be the setting of ride height, because front and rear ride heights are seldom specified with precision.

APPENDIX 2: SPECIFICATION OF KINEMATICS AND COMPLIANCE TESTS

STANDARD CONDITIONS

- All tests are run at, or referenced to, curb trim height.
- All kinematics tests are run with engine stopped.
- Steering system tests are run with engine running and with engine stopped. Only handwheel torque is plotted for the engine stopped condition.
- Front lateral force and aligning torque compliances with forces or torques adding are run with engine running; and with forces and torques opposing the engine is stopped.
- Front longitudinal force compliance tests are run with engine running.
- All rear compliance tests are run with engine stopped.

EQUATIONS FITTED TO PLOTTED DATA - Third-order least-squares polynomial equations are fitted to most data plots. Where dictated by the character of the data, three-element linear plots (through-center, and positive and negative extremes) are used instead. Examples of the latter are steer tests, where roadwheel steer angle vs aligning torque usually shows a distinct difference between on-center and off-center slopes; and wheel loads in ride and roll, where rates change as bump or rebound stops are engaged.

CHOICE OF DATA TO BE PLOTTED - Raw data (vs time) required for all plots will be supplied on floppy disks for all tests, and will be included in the test charge. All of the plots in the following listing can be generated from the floppy disk format. Since many of the data plots in the listing are useful only in some simulation models, and the quotation for Data Processing includes a separate charge for each plot, costs will be minimized by circling only those plots actually intended for use, or those for which fitted equations are desired.

ORGANIZATION OF DATA PLOTS - Data plots are organized into four groups:

Test type and configuration
    Non-standard test conditions if present
        Test
            Parameter plotted

Examples of non-standard test conditions are engine running or stopped, non-curb ride heights, etc. For example, a front compliance test at 30 mm jounce instead of curb height would be coded as FC-30J-SFA, etc.

IDENTIFICATION CODES FOR PLOTTED DATA

STEERING TESTS

FS: FRONT STEER, with engine running, except HTS
  RH: Roadwheel angles vs handwheel angle
    SAL Left roadwheel angle vs handwheel angle
    SAR Right roadwheel angle vs handwheel angle
    SAC Left and right side compared (overlaid)
    AVG Average of left and right sides
    OSR Overall steering ratio vs handwheel angle
    DIO Difference between inside and outside wheel vs handwheel angle, with theoretical Ackerman
    AED Ackerman error in degrees vs handwheel angle
    AEP Ackerman error in percent vs handwheel angle
    HTR Handwheel torque vs handwheel angle with engine running
    HTS Handwheel torque vs handwheel angle with engine stopped
  RW Parameters vs Roadwheel angles
    DIO Difference between inside and outside wheel angles vs inside wheel angle, with theoretical Ackerman
    AED Ackerman error in degrees vs inside wheel angle
    AEP Ackerman error in percent vs inside wheel angle
    CAL/CAR Camber angle vs wheel angle
    CCL/CCR Caster change vs wheel angle
    DXL/DXR, DYL/DYR, DZL/DZR Deflections vs wheel angle

COMPLIANCE TESTS

FC: FRONT COMPLIANCES
  SFA: SIDE FORCES ADDING (in parallel)
  SFO: SIDE FORCES OPPOSING
  LFS: LONGITUDINAL FORCES, SYMMETRICAL
  LFL: LONGITUDINAL FORCES, LEFT SIDE ONLY
  LFR: LONGITUDINAL FORCES, RIGHT SIDE ONLY
  ATA: ALIGNING TORQUES ADDING
  ATO: ALIGNING TORQUES OPPOSING
    FT/MT Applied force/moment vs time
    SAL/SAR Steer angle vs applied force
    CAL/CAR Camber vs applied force
    CCL/CCR Caster change vs applied force
    DXL/DXR, DYL/DYR, DZL/DZR Deflections vs applied force
    HRT Handwheel reaction torque vs applied force
  SFO: RCL, RCR Wheel load change vs applied force, for roll center height
  LFS: RCL, RCR Wheel load change vs applied force, for anti-dive angle

RC: REAR COMPLIANCES
  SFA: SIDE FORCES ADDING, (in parallel)
  SFO: SIDE FORCES OPPOSING
  ATA: ALIGNING TORQUES ADDING (parallel steer)
  ATO: ALIGNING TORQUES OPPOSING
  LFS: LONGITUDINAL FORCES, SYMMETRICAL
  LFL: LONGITUDINAL FORCES, LEFT SIDE ONLY
  LFR: LONGITUDINAL FORCES, RIGHT SIDE ONLY
    FT/MT Applied force/moment vs time
    SAL/SAR Steer angle vs applied force/moment
    CAL/CAR Camber vs applied force/moment
    CCL/CCR Caster change vs applied force/moment
    DXL, DXR, DYX, DYR, DZX, DXR Deflections vs applied force/moment
RC - SFO: RCL,RCR Wheel load change vs applied force, for roll center
RC - LFS: RCL,RCR Wheel load change vs applied force for anti-dive angle

RIDE AND ROLL SPRING RATES AND KINEMATICS

RID: RIDE spring rates and kinematics
  FSR: FRONT SPRING RATES IN RIDE
  FSR: REAR SPRING RATES IN RIDE
    DZT Wheel travel with respect to sprung mass vs time, left and right sides
    WLL/WLR Wheel load vs wheel travel
    TZT Tire travel (spin axis reference point with respect to ground) vs time, left and right sides
    TLL/TLR Wheel load vs tire travel

  FKN: FRONT KINEMATICS in ride
    SAL/SAR Steer angle vs wheel travel
    CAL/CAR Camber angle vs wheel travel
    CCL/CCR Caster change vs wheel travel
    DXL, DXR, DYX, DYR, DZL, DZR Deflections vs wheel travel

PXL/PXR Deflection of wheelpad with respect to Frame, vs wheel + tire travel
PYL/PYR Deflection of wheelpad with respect to Frame, vs wheel + tire travel
STL/STR Shock absorber travel vs wheel travel

ROL: ROLL SPRING RATES AND KINEMATICS
FRR: FRONT ROLL RATES
RRR: REAR ROLL RATES
DZT Wheel travel with respect to sprung mass, left and right sides, vs time
SRA Suspension roll angle vs time, left and right sides
WLL/WLR Wheel loads vs suspension roll angle
RM Roll moment vs suspension roll angle
CRR: COMPOSITE ROLL RATES
VRM Vehicle roll moment (total roll moment) vs suspension roll angle
RSD Percent of vehicle roll moment in front vs suspension roll angle

ROL - FK: FRONT KINEMATICS IN ROLL
ROL - RK: REAR KINEMATICS IN ROLL
SAL/SAR Steer angle vs suspension roll angle
CAL/CAR Camber angle vs suspension roll angle
CCL/CAR Caster change vs suspension roll angle
DXL/DXR, DYL/DYR Deflections vs suspension roll angle
TZT Tire travel vs time, left and right sides
DZT Total travel, suspension plus tire, vs time, left and right sides
VRA Vehicle roll angle with respect to ground vs time, left and right sides
PXL/PXR, PYL/PYR X and Y Deflections of wheelpad with respect to Frame, vs vehicle roll angle
STL/STR Shock absorber travel vs roll angle

971058

Handling Analysis with Vehicle Dynamics Simulator

Katsutoshi Horinouchi, Takashi Yonekawa, Tamio Kanou, Seisyu Utsumi, and Yoshihisa Nagahara
Toyota Motor Corp.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

We have developed a vehicle test system called the Vehicle Dynamics Simulator (VDS). The system measures the handling characteristics in a transient state in the laboratory. The automobile suspensions are moved as on a road with the machine providing relative motion by force transducer platform beneath each tire. The detailed measurements of transitive motions and forces given to the wheel clarify the kinematics and compliance characteristics contributed to the good handling performance and stability.

This paper presents the system introduction and the results of analyzing the suspensions characteristics by the new analytical technique for breaking down into a variety of compliance components in a transient state.

INTRODUCTION

We have developed a vehicle testing system called the vehicle dynamics simulator (VDS) which analyzes the handling characteristics of a vehicle in the laboratory. VDS is outlined in figures 1 and 2. This testing system is featured by its ability to measure the behavior of steering and suspension systems in detail in a transient state by simulating the actual running conditions provided by a platform beneath each tire in the lab.



Figure 1: Vehicle Dynamics Simulator

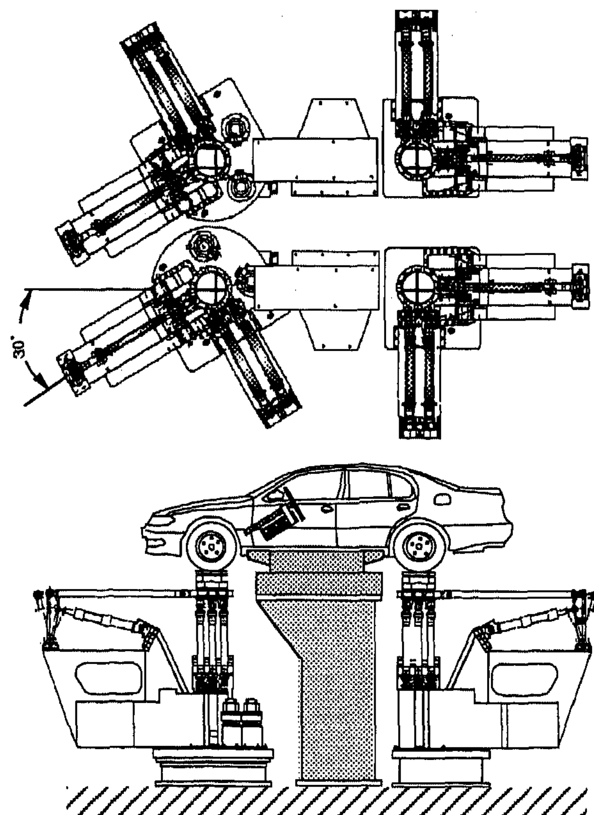


Figure 2: Elements of Vehicle Dynamics Simulator

Currently, the road simulator with four-wheel flat belt is well known as the typical on-road simulation system.[1][2] The four-wheel flat belt system is almost ideal for on-road simulation in the lab since it is capable of simulating even the rotation of tire on flat surface. However, this system has

difficulties analyzing an emergency accident avoidance behavior accompanied by a sudden change in steering since the flat belt unit itself is heavy and the system has a limited response in the steering direction. Moreover, this system involves difficulties controlling reverse input from the tire-contacting surface, therefore the system is unsuitable for in-depth measurement of compliance characteristics of the suspension by reverse input control.

Although there are a variety of systems that measure suspension characteristics by static reverse input [3][4], none of them are capable of simulating the dynamically-compounded inputs under the on-road conditions, and checking the characteristics actually attributable to the suspension for four wheels at the same time. We have prepared this paper since we have succeeded in applying the VDS to measure the transient suspension behavior under on-road conditions, and establishing the analytical technique for breaking down into a variety of compliance components at a transient state by making use of a reverse input controlling function.

SIMULATION TEST

SYSTEM CONFIGURATION - The VDS consists of the tire-contacting surface inputting four-wheel four degree-of-freedom (vertical, lateral, longitudinal and steer directions) plus two degree-of-freedom (steer direction, base steer) hydraulic exciter as the principal system and the steering actuator which has been set to allow synchronized control. Table 1 shows the performance of the main exciter. Unlike a durability tester, the exciter has been tuned with a special attention to the control performance for the low-frequency component because the exciter is intended for analysis of handling performance.

ROAD SIMULATION METHOD - For the control software to simulate the on-road conditions, we have used the RPC[5] control which has been successful with the durability tester. To fix the body, we have mainly employed two methods: one of which uses rigs to fix the rocker section of the body (figure 2), and the other uses cable to fix the hypothetical gravitational center of the body (leaving the body movable in jounce, pitch and roll directions, and fixed in yaw, lateral and longitudinal directions).

We have used a spindle coupled six-component axle meter[6] and a steering wheel meter as the detecting devices

Table 1: Specifications

	Direction :	Stroke,	Load, (Max.)	Frequency
Performance	Vertical :	±100mm,	20kN,	50Hz
	Lateral :	↑	↑	↑
	Longitudinal :	↑	↑	↑
	Steer :	±5deg,	2kN-m,	↑
	Base steer (front corner only) :	±30deg,	—	2.5Hz
Measurements on each wheel-pan	Six degree-of-freedom force and moment Three direction displacement and Steer angle			

in the simulation. We also have time-series data on the spindle force and the torque in four-wheel four directions (vertical,lateral,longi.,steer), and the steering angle during the on-road test as well as the data on the motions of the body. After the simulation, we can process the behavior of the suspension analyzed in the lab and the motions of the body during the on-road test in synchronization. Figure 3 is the flow chart of road simulation.

While the simulation is impossible without the on-road test at this moment, we will present a solution at the paragraph of Development of the Real-Time Simulator System (RTSS) in this paper.

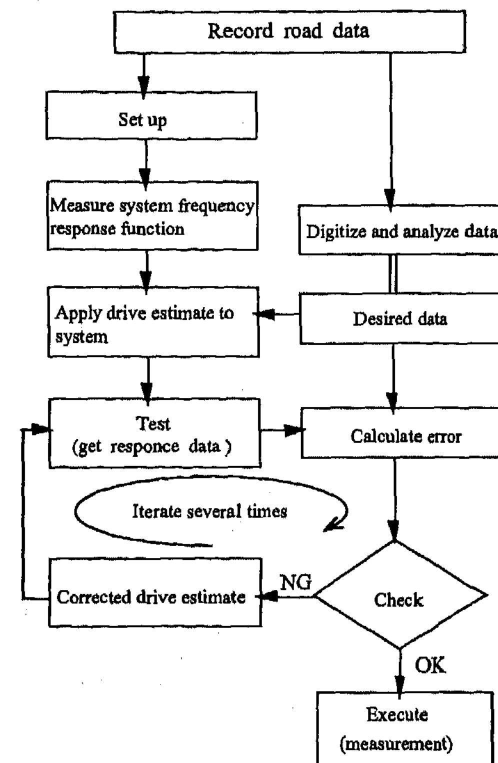


Figure 3: Flowchart of road simulation (RPC)

EXTENT OF ON-ROAD TEST SIMULATION - If the VDS is going to be used effectively as the handling performance analyzer, the system needs to simulate not only from the low turning G area to the critical turning high G area, but also from the low steering rate area to rapid steering area such as emergency avoidance driving.

Figure 4 compares six-component axle force data for the lane change in the low G area, the double lane change in the high G area and the braking in turn from the high turning G obtained in the road test and the simulation. When these data are collected, power steering is on, and the absorber temperature is controlled. In every case, the precision of the six-component axle force meter was good enough (so close that the VDS data overlaps) that, with the exception of braking force from critical turning, conditions could be simulated with an error no greater than 20N, relative to a force of 20kN with 0.1% non-straight-ahead movement.

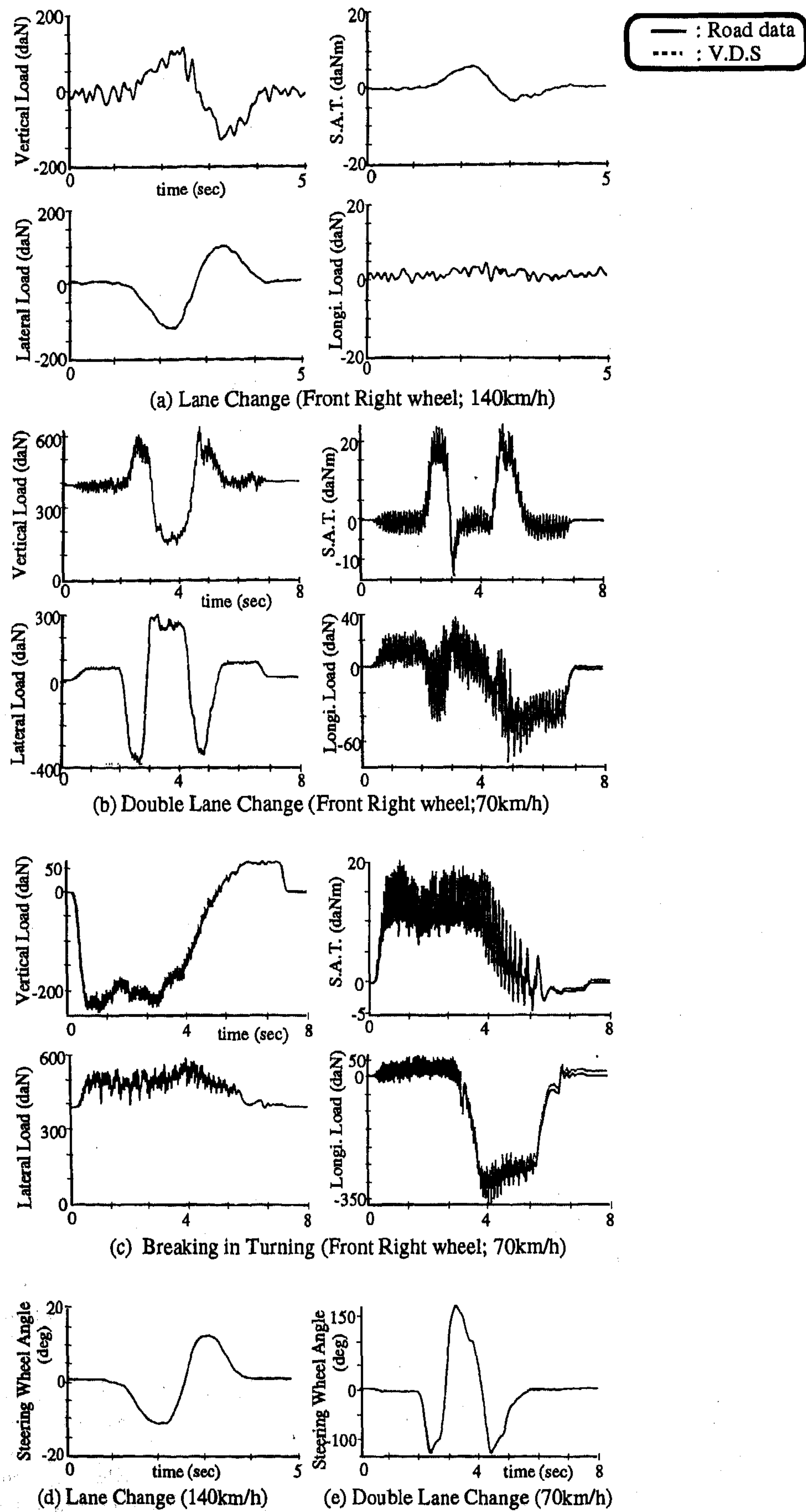


Figure 4: Spindle forces and steering angle repeatability

Table 2: RMS error on each channel

(Double Lane Change)

Channel			Ful scale ／10V	RMS error (%) [≦2%]
Front	left	vertical	2000daN	0.08%
		lateral	2000daN	0.12%
		longi.	2000daN	0.11%
		steer	100daNm	0.52%
	right	vertical	2000daN	0.13%
		lateral	2000daN	0.10%
		longi.	2000daN	0.10%
		steer	100daNm	0.67%
Rear	left	vertical	2000daN	0.05%
		lateral	2000daN	0.04%
		longi.	2000daN	0.03%
		steer	100daNm	0.17%
	right	vertical	2000daN	0.08%
		lateral	2000daN	0.06%
		longi.	2000daN	0.06%
		steer	100daNm	0.69%
Steeringwheel angle			540deg	0.19%

(DC~25Hz)

Table 2 shows the error in terms of RMS value which is the general indicator of the simulation performance of RPC. The fact that the error in the RMS value is less than 1% for every directional component during the double lane change shows that the simulation performance of the system is considered satisfactory. Similar results are obtained in other cases.

Thus, we have secured the simulation of the handling performance from the low turning G area to high turning G area and from the low steering rate area to the high steering rate area with the system, and are thus able to perform the analysis.

EXAMPLE OF APPLICATION

VALIDITY OF ON-ROAD TEST SIMULATION - It is extremely difficult to measure the behavior of the suspension and the force applied to the suspension for four wheels at the same time during the on-road test. A lab tester is required to have excellent simulation performance and repeatability. In our testing system, once we make the simulated excitation waveform, we can simulate the same behavior again and again with this waveform. We can simultaneously obtain a wide variety of data under a single driving condition by setting additional detection devices.

Since our testing system is equipped with the six-component meter on the tire-contacting surface (wheel-pan), the system is capable not only of measuring the displacement (four-wheel stroke, the displacement of the contact point, etc.), but also measuring the exchange of force on the contacting surface at the same time. Moreover, the system is capable of analyzing the pitch, roll, jounce and jack up during the on-road simulation with the body completely fixed by converting the vertical stroke of each wheel. The system also constantly measures the steer angle, the camber angle, and the steering wheel angle relative to the body, and also can easily measure the displacement of the suspension point and the localized body displacement whenever necessary. Figure 5 shows an example of measured behavior.

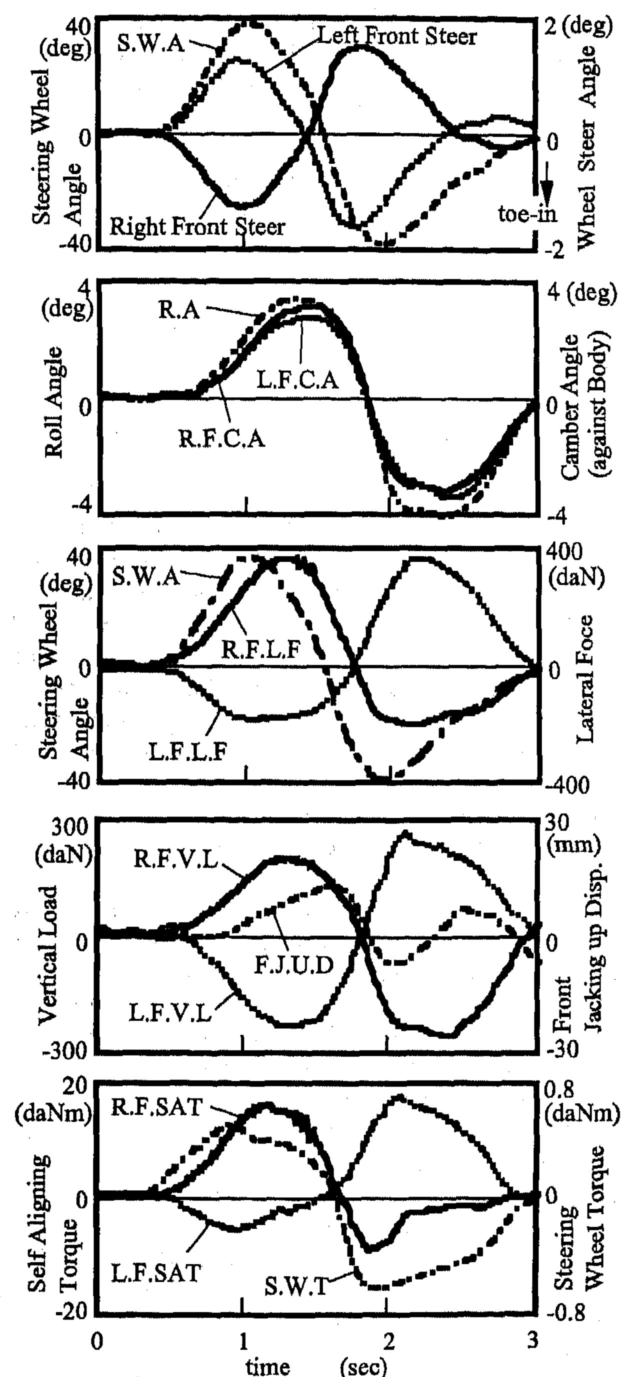


Figure 5: Example of measured results at lane change

Furthermore, the system is capable of synchronized processing of the above data and those data measured during the road test including the body slip angle and the yaw rate in order to determine the tire slip angle of each wheel.

IMPORTANCE OF ANALYSIS OF DYNAMIC COMPLIANCE COMPONENT - When designing a suspension, it is important to determine the geometry (roll steer, roll camber, roll center, etc.) and various compliance characteristics (lateral force steer, lateral force camber, etc.). However, there has been only a few published reports describing how these specifications actually exert an influence on the suspension and at what contribution rate and whether or not these specifications have an actual effect as designed at a specific moment of on-road test under a combination of circumstances (such as application of lateral or longitudinal force while rolling).

For analytical clarification of such contribution rate, we need not only measure the change in the steer angle, but also the motion of each suspension arm and bushing, vertical stroke and the input conditions in a synchronized manner. However, it is extremely difficult to break down the measured data on-road test into various compliance components.

With the ease of reverse input control provided by our testing system, we have developed a technique to clarify the contribution of a specific compliance component at a transient state on an actual vehicle.

DESCRIPTION OF ANALYSIS PROCEDURE - The body is totally fixed, and the on-road test simulation is conducted by inputting the actual running data and through RPC control. With this step, the force applied from the testing system to the contact surface of each tire is determined. On the basis of the waveform to be inputted to the contact surface, RPC control is carried out for example by preparing the simulated desire waveform at the Self Aligning Torque (SAT) of almost zero. Then, the same control is carried out by setting only the lateral force at zero. After conducting a series of simulation to break down the force applied to the contact surface, the control in which only the steering system is moved with the vertical stroke at zero is conducted. It should be noted that the vertical stroke is kept constant during the breakdown of surface force. Then, the data after each control is compared with the data obtained from the waveform simulating the original to check the ratio of each component. Figure 6 is the flow chart of the break down method.

Figure 7 is an example shown in terms of a change in the steer angle, and Figure 8 depicts the change in a schematic form. This diagram shows the results of a rather strong lane change (at 140km/h) of a vehicle with relatively good handling performance, indicating the steering angle, the change in the steer angle of the outer front wheel during the initial steering, and the change in the steer angle of the outer rear wheel.

Figure 7 (a) shows an example of breaking down the change in the steer angle of the outer front wheel in the

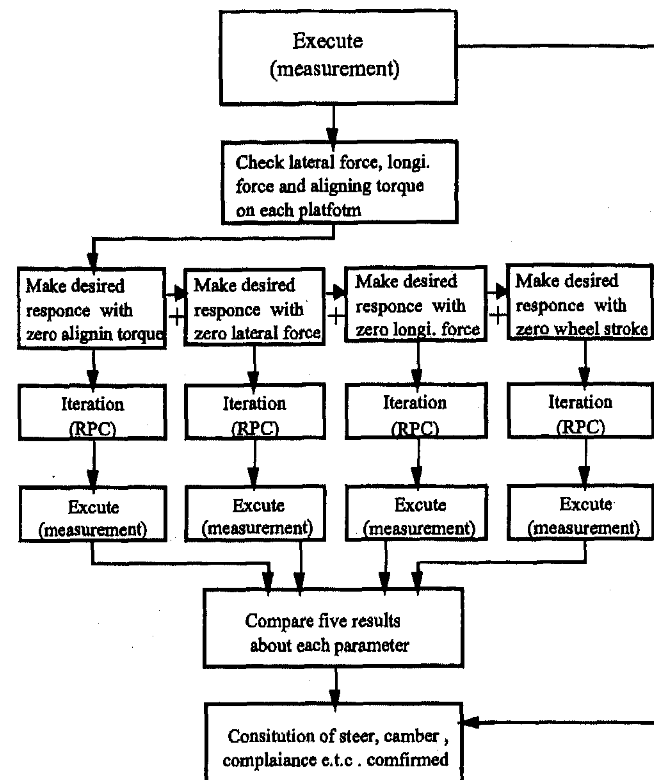


Figure 6: Flowchart of steer constitution analysis method

initial stage of turning into the roll steer component, the lateral force steer component and the SAT steer component. The α curve indicates the change in the steer during the on-road simulation based on the simulation of the original without inputting the torque in the SAT direction, and the increase in the steer angle (oversteer direction) is expressed as the SAT steer component. Likewise, the condition in which the lateral force is canceled is expressed as the lateral force steer component, and the change in the condition where the wheel stroke is canceled as the roll steer component. It goes without saying that the steering wheel angle is so controlled that the same input is made at a transient state in all cases.

Figure 7 (b) shows the case of the rear wheel turning. The same type of on-road simulation can achieve the breakdown analysis for the rear wheel. Such breakdown analysis of compliance components will help us understand the actual contribution of a specific feature of the suspension in a moving vehicle at a transient state, and examine the functioning of the feature in relation to the design target, and design the most suitable chassis system by understanding the characteristics of a vehicle with excellent performance.

EXAMPLE OF ANALYSIS RESULTS AND DISCUSSION - Figure 9 shows the breakdown data obtained during the same lane change for Vehicle B which uses a different type of suspension for both front and rear

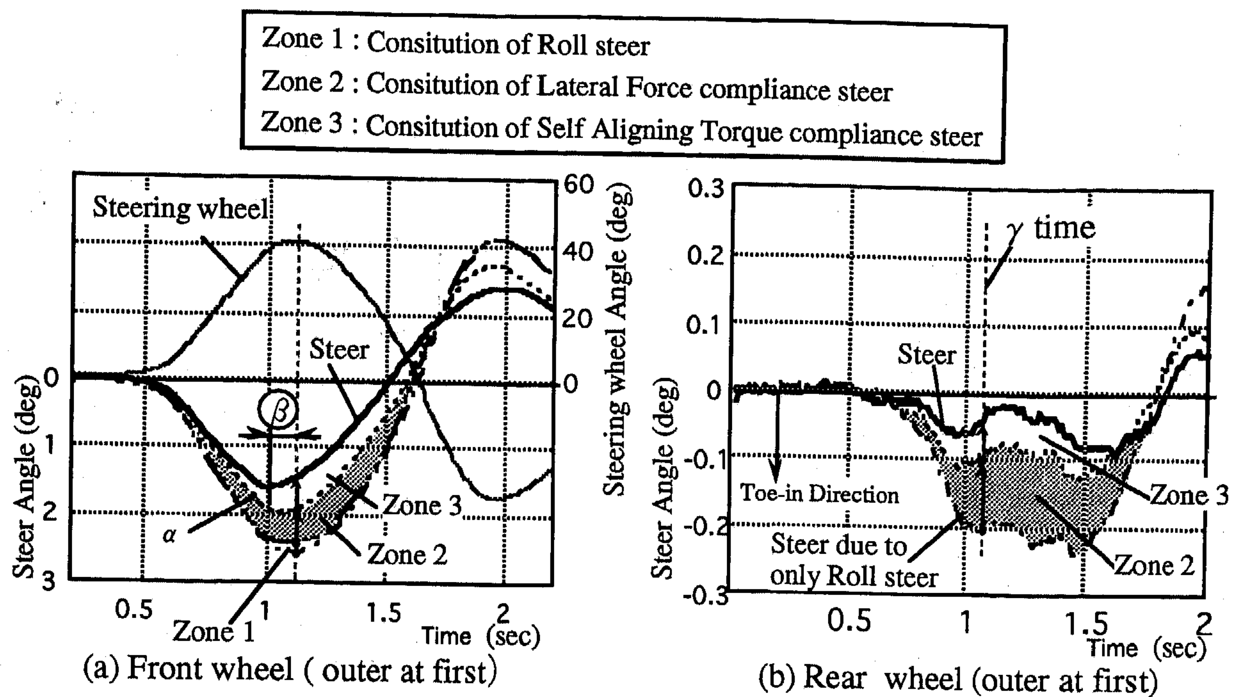


Figure 7: Analysis of steer angle constitution (vehicle A)

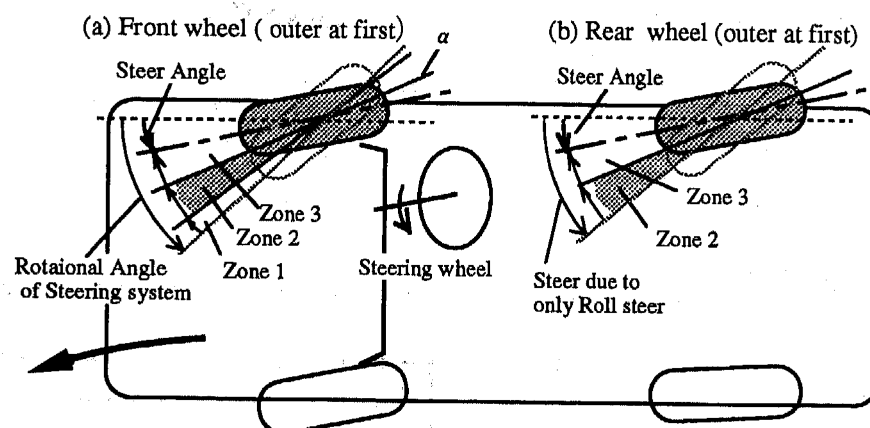


Figure 8: Steer angle constitution model

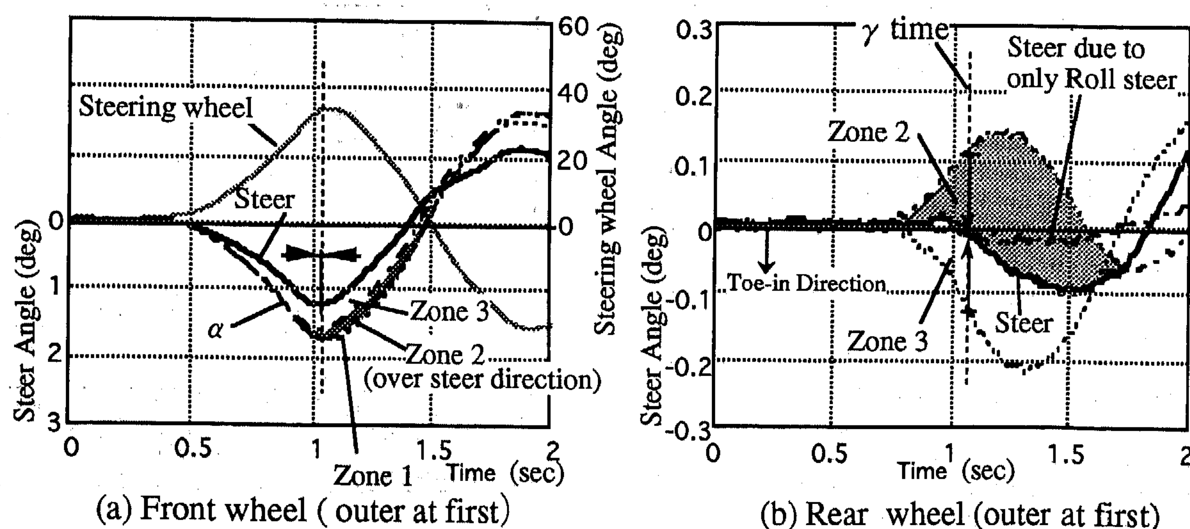


Figure 9: Analysis of steer angle constitution (vehicle B)

wheels. In front wheel, Vehicle B does not display the phase advance in the steer angle against the steering for Vehicle A shown in the section β in figure 7, indicating that the actual steer has lagged in appearance. Figure 10 shows the front wheel slip angle (as calculated from the body slip angle and the front wheel steer angle) of Vehicle A. A considerable lag in the body slip angle indicates that the front wheel slip angle has lagged behind the steering and that the phase advance in the steer angle is still insufficient. A greater lag is predicted for Vehicle B, showing its inferior controllability of the vehicle motion.

The phase lag is caused by the fact that Vehicle B has the SAT steer component as the main component while Vehicle A has the lateral force steer as the main component. Figure 11 shows generation of lateral force and SAT on the front wheel. Since SAT decreases when the tire slip angle is around 3 degrees, the peak of the SAT steer component appears to be advanced relative to the lateral force. Therefore, wherever the SAT steer component is the main component, the steer angle appears to increase (and the steering wheel return component decreases) when the SAT decreases, and the resultant steer angle appears to have less phase advance relative to steering.

Figures 7 and 9(b) show the results of breakdown for the rear suspension. To make the matter simpler, figure 12 shows the steer component at a given moment γ (equivalent to the turning lateral acceleration = 4 m/s²) in a bar chart. In Vehicle A, the roll steer component is significantly large, indicating lateral force oversteer, but the torsional rigidity of SAT direction is relatively high and the SAT steer component is relatively small. In total, the characteristics in the understeer direction have been secured. In Vehicle B, the roll steer is almost neutral, but the lateral force causes understeer. Since the torsional rigidity is low, the SAT steer is significantly changing to the oversteer side. Overall, Vehicle B has secured near-neutral steer.

The results show that in order to improve the steer characteristics of Vehicle B, the lateral force understeer component needs to be increased for the front wheel, and the SAT oversteer caused by an increase in the torsional rigidity needs to be solved for the rear wheel.

ANALYSIS OF OTHER SUSPENSION SPECIFICATION - In addition to breakdown of steer angle components shown above, the testing system is capable of component breakdown for the change in the camber angle, and the contact point displacement (displacement of the vertical load center during a given time period) by making use of the characteristics of the contact surface control. Moreover, this system can perform various analyses other than the suspension characteristics analysis at a transient state. In order to verify the breakdown characteristics of each component shown above, our testing system makes use of the merit of reverse input control to make in-depth measurement of normal suspension characteristics (roll steer, lateral force steer, SAT directional rigidity, lateral force camber, longitudinal force steer, etc.) under a

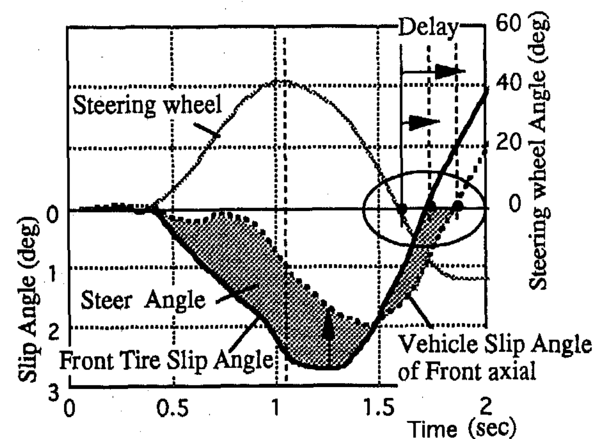


Figure 10: Front wheel slip angle

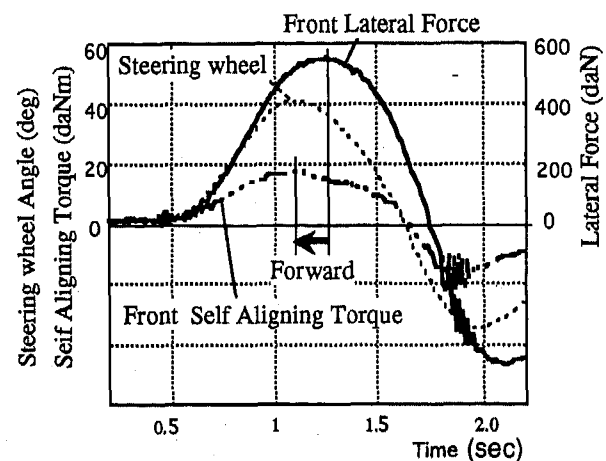


Figure 11: Front wheel lateral force and SAT

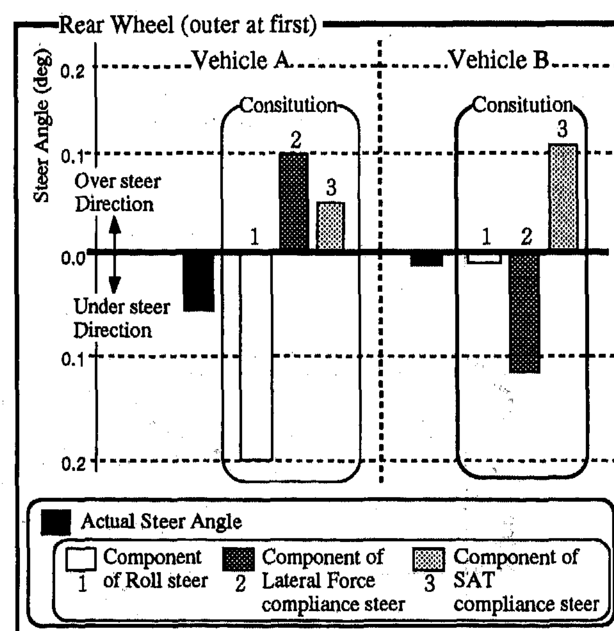


Figure 12: Component of rear wheel steer angle (γ time)

combination of inputs, and also measure above characteristics by making a dynamic input whenever necessary. Furthermore, the system dynamically determines the wheel rate to easily predict the damping force characteristics under the suspension assembly condition by the calculation from the wheel stroke rate, the vertical load and the wheel rate.

As described above, our testing system has made possible in-depth verification and feedback of the data on an actual vehicle at a transient state. By doing so, the system will contribute to improvement of the precision in the prediction of moment-by-moment performance during the design stage, and will help develop a new chassis with superior performance.

DEVELOPMENT OF THE REAL-TIME SIMULATION SYSTEM (RTSS)

We have developed the Real-Time Simulation System (RTSS) which evaluates and analyzes the motion of a vehicle within the framework of a laboratory. As described in the above paragraph of Road Simulation Method, both the on-road test and the lab testing are required in the on-road simulation using the VDS. This testing method achieves an extremely high level of precision in simulating the on-road conditions. Moreover, the system can easily measure the motion of each chassis component relative to the body, and analyze the relationship between the motion of the component and the handling performance in greater detail because the body is completely fixed.

On the other hand, the RTSS is intended for testing in the laboratory alone as in the case of an ordinary bench test, and thus it is capable of evaluation and analysis without on-road testing. This system also has the merit of accurately measuring the exchange of force on the tire-contacting surface because the body is fixed at the virtual gravitational point.

Real-time simulation [7][8] using the actual vehicle has been in use mostly for establishment of ABS and other control systems. We have applied the RTSS to the suspension (and part of the body motion) of an actual vehicle. In our opinion, the system is an effective tool for dynamic behavioral analysis of the chassis (component) motion.

SYSTEM CONFIGURATION - This system is consisted of two major components: one is the VDS and the other is the Real-Time Simulator (RTS). The RTS is a computer application designed to calculate the motion of a vehicle on the two-dimensional surface. Figure 13 shows the overall configuration.

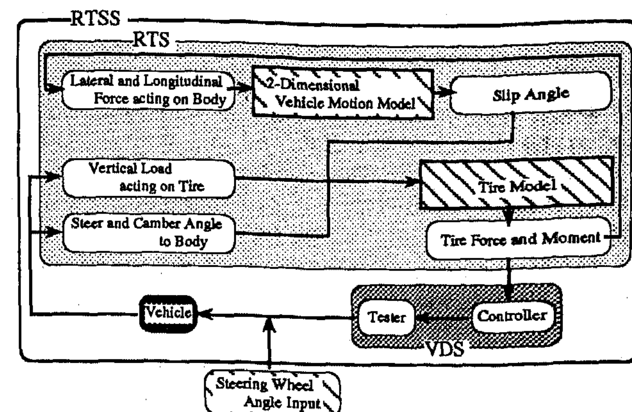


Figure 13: Structure of RTSS

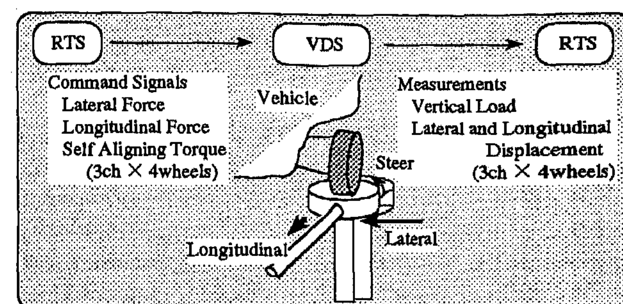


Figure 14: Function of VDS

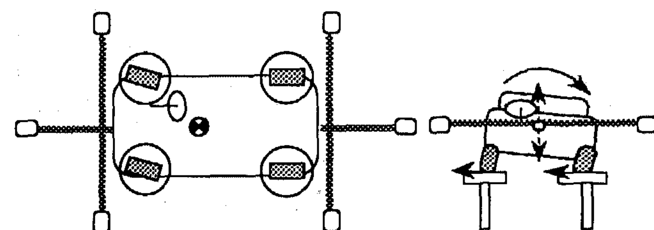


Figure 15: Method of body clamp

This system uses 13 out of 19 degrees of freedom (lateral, longitudinal and steer direction times four wheels plus steering angle) of VDS. The excitation machine is controlled by load control (and torque control). In other words, the excitation machine is so controlled that the load is applied to the vehicle as instructed by the RTS. (The excitation machine is controlled by detecting three components (lateral force, longitudinal force, steer directional torque) with the six-component meter built in each wheel-pan.) (figure 14) The body is fixed by fixing the virtual gravitational point of the vehicle with cables. In fixing the body, a longitudinal axis running through the gravitational point of the vehicle is postulated, and three pieces of cable are used respectively to fix the intersections of this axis and the front end of the body and the rear end of the body. In this method, the body is fixed in longitudinal, lateral and yaw directions, and is left movable in jounce, pitch and roll directions (figure 15). In other words, this system does not simulate the vehicle motion in the horizontal surface direction, but simulates other motions including vertical motion on the bench.

Three laser displacement gages are used for each wheel to measure the steer angle and the camber angle relative to the body. The data are sent to the RTS on a real-time basis, and used for calculation of tire force. A free gyroscope is used to measure the pitch angle and roll angle of the body relative to the ground. The data are sent to the RTS on a real-time basis and used for calculation of the camber angle relative to the ground.

The six-component meter built in the wheel-pan of each wheel is used to measure the vertical load. The measured data are sent to the RTS on a real-time basis, and used for calculation of the tire force (figures 14 and 15).

The stroke of the excitation machine outputted on the testing system side is inputted. The data are sent to the RTS on a real-time basis in terms of the tire-contacting point displacement (lateral and longitudinal) relative to the body (figures 14 and 15). With the testing system, the steering wheel can be turned according to the waveform (e.g. sine wave) intended by the operator. In our verification test, we have simulated the driver's steering waveform during the on-road test in the lab.

ALGORITHM AND MODEL - The RTS is a computer application designed for calculation of the motion of a vehicle on the horizontal surface from the motion of the suspension and the body on the testing system. The RTS has two main parts: one is the tire model and the other is the two-dimensional vehicle motion model. Figure 16 shows the functions of the RTS.

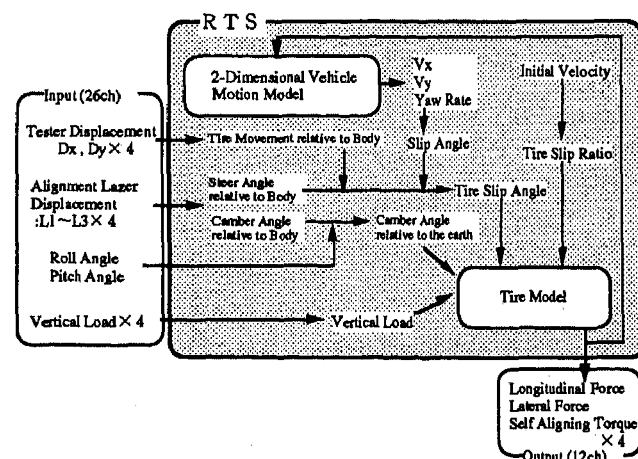


Figure 16: Function of RTS

The tire model used in the RTSS is based on the experimental data already obtained on the flat belt tire tester. The tire model is expressed in mapping of experimental data, and the gap between available data is interpolated. Figures 17 and 18 show the outline of the tire model. The motion of a vehicle on the horizontal surface is calculated from the longitudinal and lateral force generated by the tire. For this model, a number of basic vehicle specifications (empty vehicle weight, yaw moment of inertia, wheel base, etc.) need to be pre-measured. This model is outlined in figure 19.

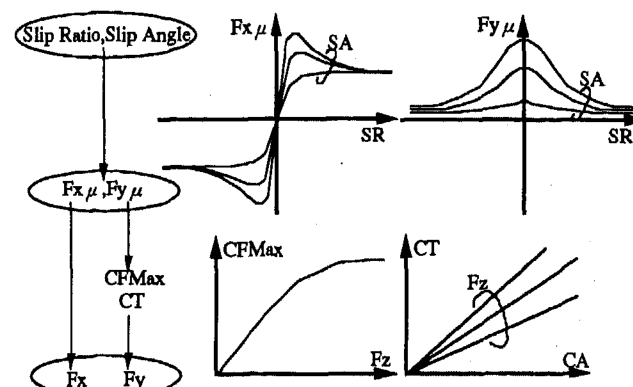


Figure 17: Tire model (lateral and longitudinal force)

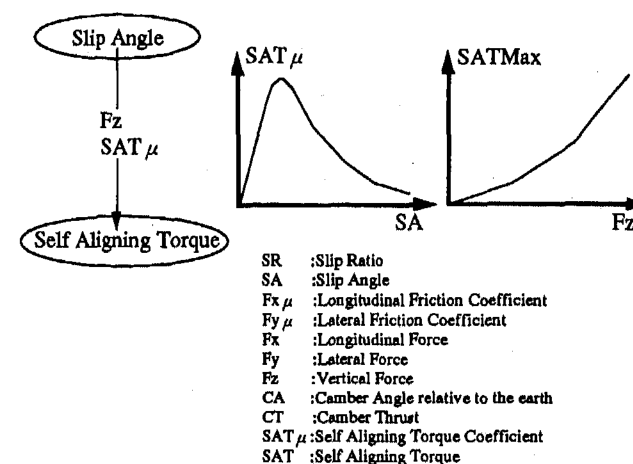
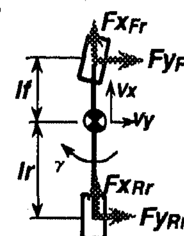


Figure 18: Tire model (SAT)

$$M\left(\frac{dVx}{dt} - Vy*\gamma\right) = Fx_{Fr} + Fx_{Rr} - Cd*Vx^2$$

$$M\left(\frac{dVy}{dt} + Vx*\gamma\right) = Fy_{Fr} + Fy_{Rr}$$

$$I\frac{d\gamma}{dt} = Fy_{Fr}*lf - Fy_{Rr}*lr$$



M :Weight
I :Yaw Inertia
lf / lr :Distance between C.G. and front/rear axle
Vx/Vy :Longitudinal/Lateral Velocity
γ :Yaw Rate
Fx / Fy :Longitudinal/Lateral Force

Figure 19: 2-Dimensional vehicle motion model

VERIFICATION TEST AND RESULTS - We have conducted a test to verify the on-road simulation performance of the RTSS. The running condition is the lane change at 140km/h. We have conducted the on-road test, and then the test using the RTSS in the lab under the same vehicle condition and the same steering input. During the on-road test, we have recorded the measurement at a proving ground. In the RTSS test, the calculated value from

the RTS is used for the motion on the horizontal surface, and all other items including vertical motion are measured in the lab. A picture of the RTSS testing a vehicle is shown in figure 20.

Figures 21 and 22 show a comparison of two-dimensional flat surface motion (Results of on-road test vs. Results of RTSS test calculation). In terms of flat-surface motion (yaw rate and lateral acceleration), the RTSS has simulated the on-road condition almost perfectly. However, the RTSS has not simulated the minor details since disturbance from the ground and any other matters is not taken into consideration in this test. The RTSS has perfectly simulated the steering angle during the on-road test since the steering wheel actuator has been set to simulate the steering during the on-road test.

The RTSS has closely simulated the on-road condition in terms of the motion of the vehicle in the lab (tire force in figure 23, roll angle in figure 24). However, a certain phase delay is found especially in the roll angle, which is caused by the delay in the load control system of the hydraulic excitation machine and the method of fixing the vehicle. Probably because of these factors, the actual vehicle part in the lab has shown the tendency to lag in its motion when we have attempted to simulate a quicker behavior.

However, such tendency can be sufficiently minimized by changing the setting of the load control gain of the excitation machine. There will not be a problem in evaluation and analysis which are the purposes of this system (although the system will have limited range of application).

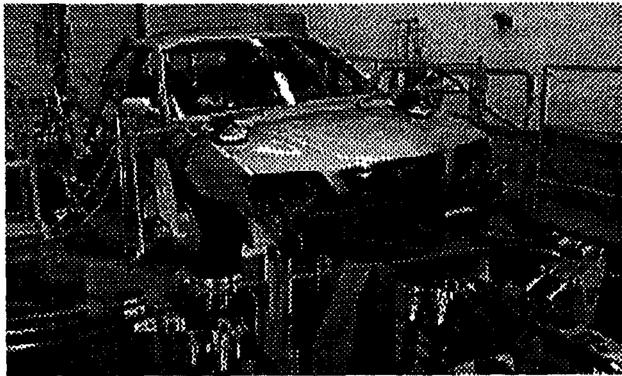


Figure 20: RTSS test in the laboratory

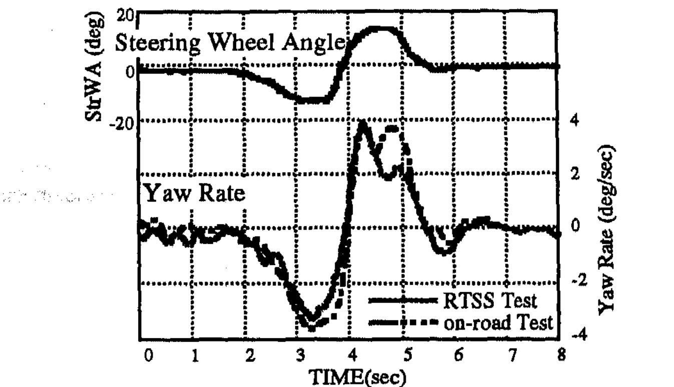


Figure 21: Comparison of yaw rate and steering wheel angle

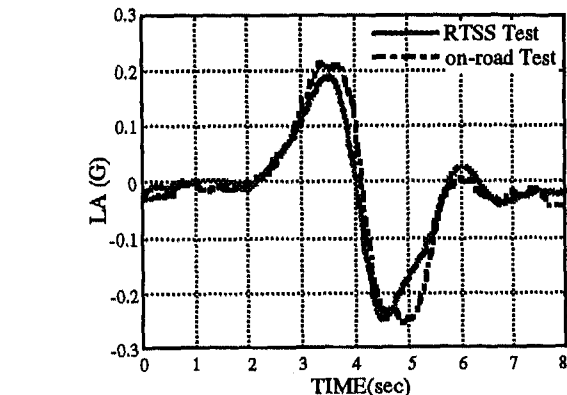


Figure 22: Comparison of lateral acceleration

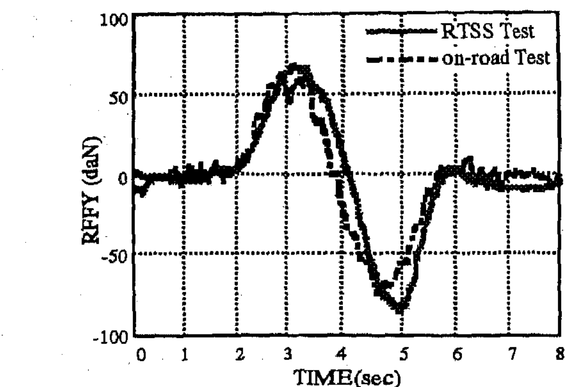


Figure 23: Comparison of right-front tire lateral force

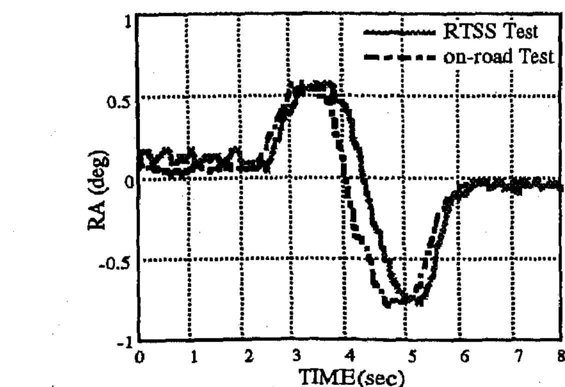


Figure 24: Comparison of roll angle

CONCLUSION

(1) In order to analyze the handling performance of an automobile, we have developed the VDS which analyzes the characteristics of a suspension in a transient state in the lab. We have verified that this testing system is capable of simulating the behavior in the range from low turning G to critical turning G and the range from slow steering to rapid steering in the emergency accident avoidance.
(2) We have established a technique that makes possible breakdown of factors of suspension characteristics of an actual vehicle in a transient state which has been considered difficult. For example, we have shown that the system is capable of clearly indicating the contribution of various steer angle components in a transient state.
(3) Since above steer components (lateral force steer component, SAT steer component,e.t.c.) have different phases in a transient state, we have made use of the difference in characteristics and verified that we can use the system to achieve optimum design of a suspension with excellent stability and controllability.
(4) We have shown that this system is a general-purpose system capable of measuring both static and dynamic specifications at once.
(5) We have developed the RTSS which uses the actual suspension and body of a vehicle and verified that RTSS is capable of simulating the dynamic vehicle motion by the lab test alone. The test with RTSS has the following merits:

- Ease of dynamic measurement of behavior and input load of each chassis component
- The tire model may be modified so that the system can be used for simulation of behavior under various tire and road conditions.

Finally, we would like to thank the people of MTS Systems Corp., MTS Japan Ltd., Toyota Macs Corp., Toyooki Kogyo Co., Ltd., Yokogawa System Engineering Corp., and Fujitsu Ltd. for their generous cooperation in our efforts to introduce this new testing system.

REFERENCES

1. Langer, W., "Durability Testing with Flat Surface Roadway Technology", IPC-8, 211, October 1995.
2. Mimuro et al., "Study of Interior Turning Test", JSAE Pre-publication Papers (912, 1991-10), 912153.
3. Jenniges, R.J., "The Application of Optical Methods in Vehicle Suspension Kinematics and Compliance Measurement", IPC-8, 212, October 1995.
4. Automax Co., Ltd.: Product Information on Basic Vehicle Characteristics Measurement Devices and Products
5. Linden, N., et al., "Techniques of Remote Parameter Control", 2nd International Symposium on Automobile Experiment Methodology and Invention, p. 789-801.
6. Higashijima et al., "Development of Multiaxial Wheel Hab Dynamometer", JSAE Pre-publication Papers (936 1993-10), 9306840.
7. Mori, Y et al., "A Simulation System for Vehicle Dynamics Control", SAEpaper, 910240.
8. Makita,M et al., "Development of Real Time Simulator", JSAE Pre-publication Papers (912 1991-10), 912153.

971059

Simulation in the Development of ASMS

B. Lammen, F. Kemmler, U. Judaschke, and S. Müller
ITT Automotive Europe GmbH

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper describes the contribution of simulation to the development of the new Automotive Stability Management System (ASMS) of ITT. The benefits and limitations of simulation especially with respect to experimental testing with prototype vehicles are discussed. The paper will outline how cost and time to market have been reduced by Off-Line Simulation (OLS) and Real-Time Simulation (RTS). During the development of ASMS, new control algorithms were designed and first validated in the laboratory. Simulation has offered an insight into the vehicle dynamics that is difficult to obtain with prototype vehicles. It has been possible to study the interactions of the vehicle control system and vehicle dynamics under all circumstances. Some simulation examples of typical maneuvers are discussed.

INTRODUCTION

At the end of 1997 the new Automotive Stability Management System of ITT will be available on the market. ASMS combines the well known Anti-Lock Brake System (ABS) and Traction Control System (TCS) with a new Active Yaw Control (AYC) to a comprehensive vehicle safety system. ASMS assists the driver and enhances the vehicle stability under all kinds of driving conditions. In addition to conventional ABS/TCS, the steering wheel angle, yaw rate, lateral acceleration, and master cylinder pressure of the vehicle are measured. AYC monitors the yaw and side slip reaction of the vehicle to driver steering inputs. If the vehicle tends to instability, AYC automatically initiates wheel individual brake intervention. Thus, the driver is able to handle the vehicle safely if the vehicle approaches the limits of lateral stability. The principle effect of ASMS is shown in figure 1. In case of undesired oversteering or understeering, the AYC module of ASMS calculates a compensating yaw torque. ASMS generates this yaw torque by applying brake forces (F_B in figure 1) at the appropriate wheels.

The development of the new vehicle control systems ASMS demands detailed information about vehicle dynamics. This information can be obtained either from testing with a prototype vehicle or from theoretical calculations based on computer simulations. Both methods have particular advantages and disadvantages.

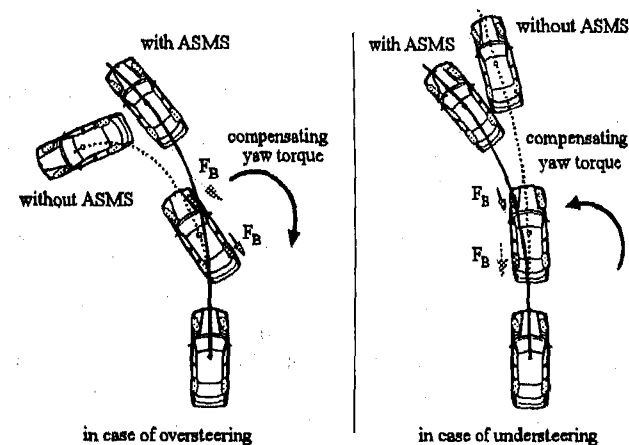


Figure 1: ASMS in case of oversteering and understeering

Testing is time consuming and requires high technical expenditure of vehicles, measurement equipment, testing areas and staff for support and maintenance. Extreme or hazardous maneuvers are difficult to realize and experimental results are not exactly reproducible. However, only experimenting can give final evidence of the functionality of a vehicle control system.

Simulation, on the other hand, can relieve vehicle prototype testing from time consuming work. It enables continuous development in the laboratory and shortens the development process of a new vehicle control system. During the design of algorithms, non-effective strategies are eliminated at an early stage. Basic test procedures are executed automatically. Maneuvers, external influences and failures can be reproduced exactly to compare and evaluate algorithms. The development of the series code can be supported by Real-Time Simulators [1], which combine the controller hardware and software with simulation models of the vehicle and its components.

However, simulation always needs a close relation to vehicle prototype testing. To achieve reliable simulation results, the models have to be verified based on measured data from vehicle testing. The more sophisticated the modeling, the more

parameters have to be provided and adapted to different vehicles. Some parameters can be directly measured at the prototype vehicle or identified from experimental data, for the others construction data from the automobile manufacturers is necessary. This requires close cooperation with customers.

The paper shows how simulation techniques have been used to reduce development time and costs of ASMS. In the following chapter an overview of the simulation tools is given and the modeling techniques for the different components of the control loop are briefly discussed. Afterwards a variety of simulation examples will be presented in detail to demonstrate the functionality of ASMS and to show the contribution of simulation in the development process. Special attention will be paid to maneuvers and external influences on the vehicle dynamics, which are difficult to generate in vehicle tests. Finally, an outlook on future use of simulation techniques in the development of vehicle control systems at ITT will be given.

SIMULATION TOOLS

According to the aim of a simulation study the complexity of modeling and an appropriate simulation tool have to be chosen. Principally, Off-Line Simulation (OLS) and Real-Time Simulation (RTS) are distinguished.

OFF-LINE SIMULATION - In OLS all components are modeled mathematically and represented as software modules. The simulations are run on workstations. The complexity of the models is not limited by real time constraints. With the help of Mechanical-System-Simulation (MSS) tools like ADAMS* the 3D-multi-body-dynamics of a vehicle are represented in detail with a high number of degrees of freedom. The interactions of the vehicle control system and the vehicle dynamics in extreme situations are evaluated. Thus, OLS offers the possibility to analyze 3D-maneuvers which are difficult to realize and reproduce in experiments. Furthermore, internal values are available in OLS, which can not be measured in experiments. In addition to MSS other tools like e.g. Xmath/SystemBuild** are used which support controller design and offer mathematical methods for system analysis and graphical modeling techniques. The postprocessing of simulation results is partially automated. After modifications in the controller software a variety of standard maneuvers are automatically simulated and evaluated to verify basic functionalities of the code.

REAL-TIME SIMULATION - In RTS the ABS, ABS/TCS or ASMS controller is integrated as hardware-in-the-loop. The code runs on the series controller and interacts with the Real-Time Simulator. The hardware of the RTS is a compact and portable device based on a 19" rack, which only needs a standard PC for downloading of the simulation models and parameters as well as for user interactions and disk control. With about $25.000 it is less cost-intensive than workstations running a real-time-OS (e.g. VxWorks). Until today 40 RTS-systems were built up. So, almost each software development engineer has access to such a system and can simulate maneuvers. This ensures short development time and relieves the test department.

MODEL DESIGN - The decisive task of each simulation model is an accurate representation of system dynamics. Furthermore high flexibility and low maintenance costs are advantages of efficient simulation tools. Therefore OLS and RTS environment are structured into the following basic modules:

- vehicle dynamics,
- powertrain,
- tire / road,
- track,
- driver,
- hydraulic unit
- and electronic controller.

Each of the simulation modules have a well defined interface and can be adapted to the specific demands of different examination goals. The combination of these basic modules to an overall simulation environment is shown in figure 2.

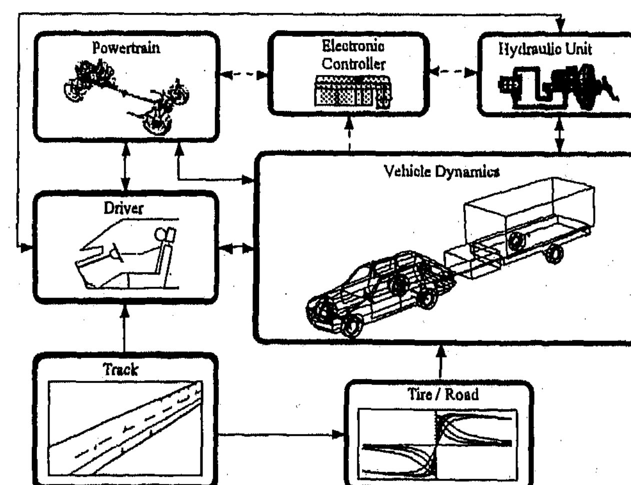


Figure 2: Structure of the simulation environment

The functionality of the modules is briefly described below.

Vehicle Dynamics - Depending on the basic effects which have to be regarded by means of simulation, an appropriate set of differential equations has to be chosen to describe the vehicle dynamics. This equation set determines the number of degrees of freedom of the vehicle model and the complexity of motion representation.

* ADAMS is a trademark of Mechanical Dynamics Inc.

** Xmath and SystemBuild are trademarks of Integrated System Inc.

Simple driving maneuvers on a plane road require at least two translational degrees of freedom in x and y direction and three rotary degrees of freedom in *yaw*, *pitch* and *roll* direction for the vehicle, while the movement in z-direction can be neglected. At the moment, RTS calculates e.g. movements in z-direction only for the sprung rigid car body, and not for the unsprung wheels. This model simplification is necessary due to the limited calculation performance of the processing unit (six T805 transputers) and with respect to real-time constrains.

More complex driving maneuvers which explain the dynamic behavior of a vehicle with respect to downward and sideward inclinations of a road require a full 3D-vehicle model with six degrees of freedom for the vehicle body and additional degrees of freedom for the dynamics of the suspension. ADAMS vehicle models e.g. passenger cars, off-road vehicles or light trucks include among 60 to 70 degrees of freedom.

Furthermore the examination of particular influences like coupling between a vehicle and a trailer demands more sophisticated sets of differential equations. Car / trailer combinations with one or two trailer axles and an overrunning brake contain 40 to 50 parts and have approximately 80 degrees of freedom.

The use of C-coded vehicle dynamics modules or ADAMS descriptions depend on scope of examination.

Powertrain - Especially for examinations in the context of traction control the description of the powertrain becomes an important part of the simulation environment. This module takes into account interdependencies between engine torque, rotational speed, throttle flap angle and interventions of the engine management like fuel cut off or spark retard as well as stiffness and damping of cardan and joint shafts.

Furthermore different drive concepts like front, rear and four wheel drive can be selected. All drive concepts allow an adequate consideration of differential gear locking torques.

Depending on the examination target the dynamic behavior of the powertrain is described with the help of C-coded modules or by means of ADAMS.

Tire / Road - The tire/road module describes the forces and torques which are generated by the interaction between road surface and tire. For simulation of ABS, TCS and ASMS maneuvers

- a modification of the HSRI-tire model which was developed by Dugoff, Fancher and Segel [2] and
- the MAGIC-FORMULA-tire model which was developed by Pacejka [3]

is used. Both tire/road models are described in C-coded modules.

Track - One of the essential advantages of simulation is the simple and repeatable definition of sophisticated environmental constrains. Therefore not only straight line and steady state cornering are offered but also any combination of bit by bit defined straight line parts and bends [4]. Furthermore special maneuvers (e.g. SAE lane change) can be chosen. The up-/downward inclination as well as the sideward inclination of a road can be defined freely. Examples for the possibilities of track definition are illustrated in figure 3 and 4.

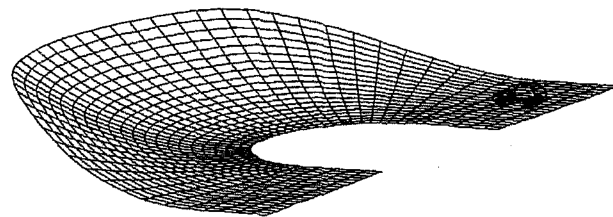


Figure 3: Definition of a banked curve

Figure 3 shows the graphical description of a banked curve which represents a combination of two straight line parts and a semicircle. The sideward inclination of the bank curve can be varied in the open interval $]-\pi/2, \pi/2[$.

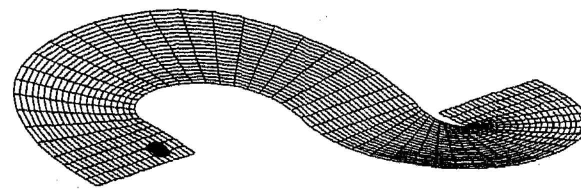


Figure 4: Definition of a serpentine

To simulate the influences of up-/downhill driving, straight lines and bends can be joined together as in figure 4. Here three straight line parts and two semicircles are combined with a downhill inclination. Furthermore a small sideward inclination in both bends is defined.

For the examination of effects of unevenness on the road different kinds of periodic waviness functions can be superimposed to the basic geometry of the road. An example of a sine function with a positive amplitude on the right half of the lane and a negative amplitude on the rest of the road is shown in figure 5.

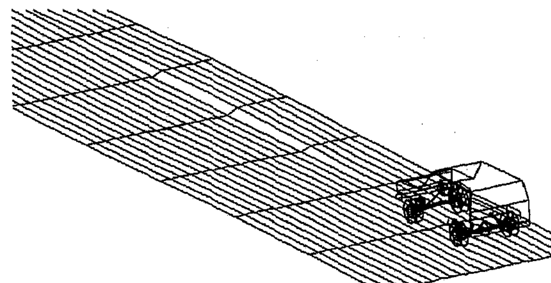


Figure 5: Definition of unevenness on the road

In addition to the geometrical description of the road in 3D, friction coefficients can be assigned to the road surface or parts of it. Therefore the road surface can be split into different parts while each part represents a special tire/road combination. The μ-split hill road in figure 6 is defined with the help of two parts, one part for the low-μ side and one part for the high-μ side.

All track models are described in C-coded modules.

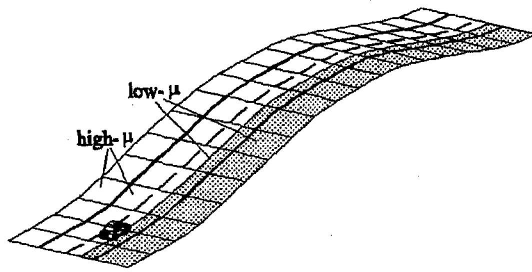


Figure 6: Definitions of road properties

Driver - In the driver module all inputs for vehicle handling are generated. The driver module is able to carry out

- open loop maneuvers and
- closed loop maneuvers.

In case of open loop maneuvers a sequence of time or distance depending actions are predefined and passed to the vehicle actuators. The throttle flap, the clutch, the gear box and the brakes are operated in a repeatable way.

On the other hand closed loop maneuvers like track following, driving with constant velocity or acceleration can be obtained by means of C-coded driver modules.

Hydraulic Unit - In the development of ABS, ABS/TCS and ASMS controllers the hydraulic unit takes an important part. Therefore the differential equations for representing brake fluid behavior are described.

Different configurations like e.g. front/rear or diagonal split hydraulic units with special kinds of pumps and valve arrangements are provided.

The hydraulic unit is normally described by means of C-coded modules but in the case of RTS applications a hydraulic hardware can also be used.

Electronic Controller - All kinds of ABS, ABS/TCS and ASMS controllers can be integrated in the simulation environment. RTS includes hardware controllers while OLS uses the software codes of the control algorithms.

ASMS-SIMULATION

The simulation tools described above have been used to support the development of ASMS. To achieve reliable results the simulation models were verified based on vehicle test data.

VERIFICATION - Measured input signals of test maneuvers can be used as input signals for the simulation. Thus, a direct verification of the models by comparison of simulation and test results is possible. Since ASMS especially affects the yaw response of the vehicle to steering inputs, special attention has to be paid to the reliability of the model of the lateral vehicle dynamics. In figure 7 the measured yaw rate and the lateral acceleration of a testing maneuver and a simulation are compared to verify the vehicle model. During the maneuver the velocity was kept constant. The maneuver was carried out without ASMS.

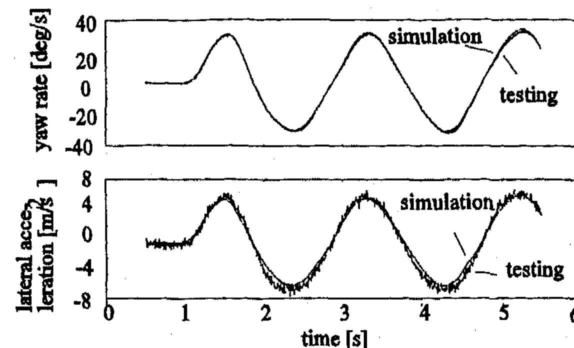


Figure 7: Comparison of testing and simulation

DESIGN OF THE AYC ALGORITHM - The initial idea of an active yaw control was first validated by various Off-Line Simulations. After finding a promising algorithm it was implemented on a prototype controller in a test vehicle which was equipped with the required sensors and an experimental hydraulics system. The development of the AYC algorithm has then been carried out in a parallel procedure both in simulation and vehicle testing. Figure 8 shows the principle structure of the AYC algorithm. AYC includes an internal reference model, that permanently calculates the desired stable reaction of the vehicle to the steering input of the driver. If the deviation between the stable behavior of the reference model and the measured reaction of the vehicle exceeds certain thresholds, a tendency to instability is assumed and AYC stabilizes the lateral dynamics by wheel individual brake intervention which is superimposed to driver braking. The AYC intervention is coordinated with the ABS/TCS functionality.

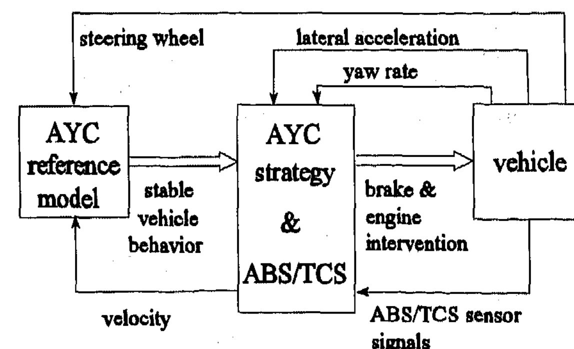


Figure 8: AYC algorithm

Parametrization of the AYC reference model - The parameters of the reference model have to be adapted to the particular type of vehicle to achieve optimal functionality and performance. The design of the reference model is a compromise between two objectives. On the one hand, an accurate representation of the vehicle is required to obtain smooth brake intervention and to prevent undesired reactions. On the other hand, the complexity of the model is limited by the real time constraints of the series micro controller. Some parameters as e.g. the mass of the vehicle can directly be measured. Others are determined in a partially automated procedure in OLS with the optimization function of Xmath as illustrated in figure 9.

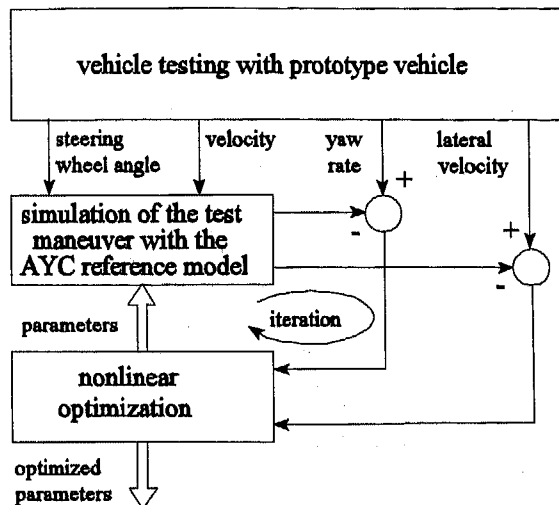


Figure 9: Parameter identification of AYC reference model

The procedure is based on data, that has to be gained from testing with prototype vehicles. A special set of maneuvers is carried out with the prototype vehicle. The maneuvers represent the desired stable behavior of the vehicle over the relevant range of velocity. During each maneuver, the steering wheel input and the velocity as well as the resulting yaw rate and lateral velocity are recorded. In the OLS the steering wheel angle and the velocity are then used to simulate the test maneuver with the AYC reference model and an initial set of parameters. The yaw rate, side slip angle, and lateral velocity are compared with the measured data from vehicle prototype. The result is combined to a quality value that shows how good the reference model fits the real maneuver with the given set of parameters. The nonlinear optimization function modifies the model parameters until an optimal fit is found.

Functionality of the AYC algorithm - The original ASMS control algorithm, as it runs on the series micro controller, was transferred to the workstation on which the OLS is carried out. The functionality of the AYC component of ASMS shall be demonstrated now with a SAE lane change on a snow surface that was executeded out in OLS. Figure 10 illustrates the course of the vehicle with and without ASMS.

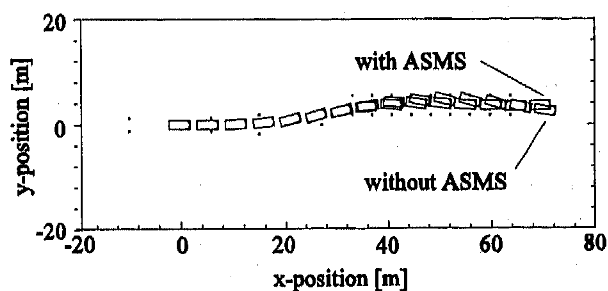


Figure 10: Vehicle course

The vehicle contour is marked in time steps of 0.333 s. The vehicle has an initial velocity of 16 m/s. The driver model of the OLS just tries to follow the course without braking the vehicle. The clutch is open. The brake pressures are plotted in figure 11. After about 1 s, AYC brakes the right front wheel. After about 2.3 s time when the vehicle has reached the left lane, there is a brake intervention at the left front wheel. There are also brake interventions at the rear axle but the level is much lower. With AYC the vehicle can execute the lane change without problems. Without AYC the driver model is not able to stabilize the vehicle in the left lane. The vehicle hits the pylons.

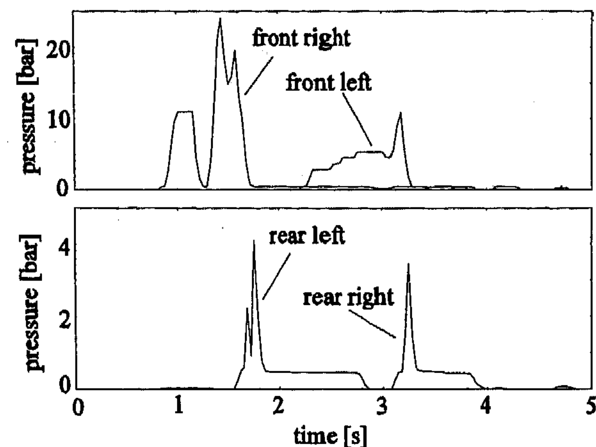


Figure 11: Wheel brake pressures

Figure 12 compares the steering wheel angle, velocity, yaw rate, side slip angle and lateral acceleration with and without ASMS. AYC reduces the steering work that is necessary for the lane change. When the vehicle with ASMS has reached the left lane, the yaw motion and the lateral acceleration are balanced much earlier. The side slip angle is significantly reduced.

Apart from the plots presented here, OLS provides all relevant internal signals of ASMS for an analysis of the functionality of ASMS. Furthermore all states of the vehicle dynamics are available and can be interpreted. To enable an efficient evaluation of the huge amount of data the post processing is automated up to a certain degree. 3D-animations of the maneuver and a standard set of plots can be generated. Different maneuvers can directly be compared with each other or with measurements from vehicle prototype testing. The data is stored in a format that allows an easy exchange between the simulation tools. Thus, a more sophisticated evaluation of results e.g. with numerical methods as they are offered by tools like Xmath is possible. For parameter studies maneuvers can be executed automatically in a loop in batch mode. Characterizing values can be extracted and displayed automatically to show the impact of parameter modifications on the functionality of ASMS.

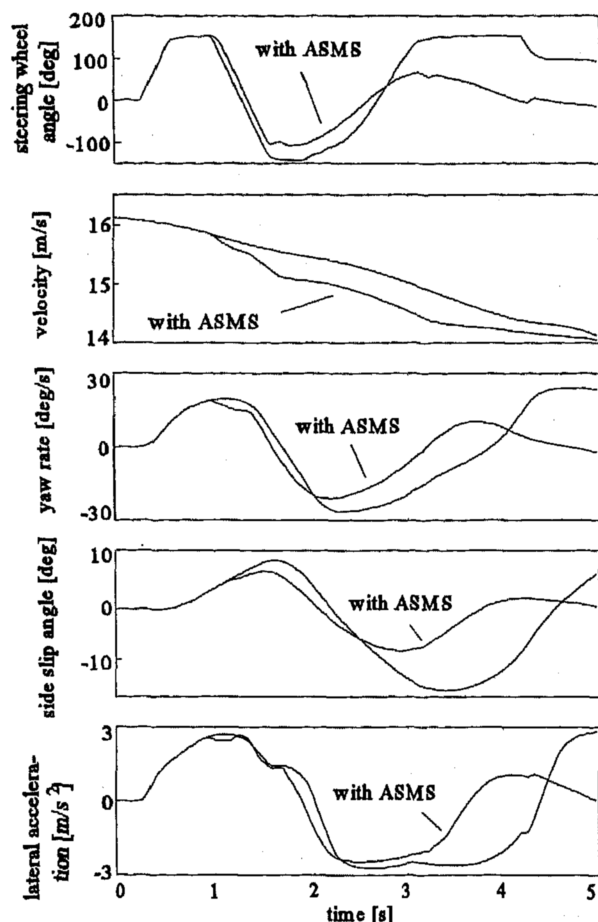


Figure 12: Vehicle states

DEVELOPMENT OF SERIES CODE - After the design of AYC and the decision for a definite algorithm it was necessary to code and integrate the algorithm in a prototype hardware environment. From this point the Real-Time Simulation has taken up an important part at code development because real hardware components (e.g. valve-, pump- and on/off-booster-dummies) are needed. Therefore the I/O-system of the RTS has to be flexible enough to grow with new hardware requirements of the ASMS controller from prototype up to series, that means an extension of the system and a configuration by software has to be possible. Figure 14 shows the basic structure of the RTS environment. We use a 16 bit SMP-bus (from Siemens) controlled by a T805-transputer as master and piggy-back-modules placed on a motherboard, which is equipped with a T225. This transputer works as Communication Interpreter for either a CAN- (Controller Area Network) or SCP- (Standard Corporate Protocol) Communication Controller and controls the transmission and the receipt of messages via the serial communication line. A typical application of field buses in simulation is the transmission of data for engine management and sensor signals e.g. the steering wheel angle. In the RTS different automotive manufacturer-specific communication protocols are implemented and can be selected by the user.

Via the protocol interface, which is a simple unidirectional parallel connection, internal values of the controller are read by the RTS. These data are displayed synchronously to the simulation data. So, incoming and prepared sensor signals or internal values of the controller like the estimated pressure in the wheel brakes can be compared with the corresponding simulation data. Another way to acquire data from both - Real-Time Simulation and internal controller values - is to use the ITT Data Acquisition System (DAS).

By means of a measuring system (DAS-Box) the simulation and controller data are collected with a predefined period of time and sent to a PC if requested. The transmission of simulation values from the RTS to the DAS follows via a serial communication line. Since the test department records the driving maneuvers with the same acquisition system, results of simulated and real maneuvers can be directly compared and verified. The integration of the DAS in the RTS is shown in figure 13.

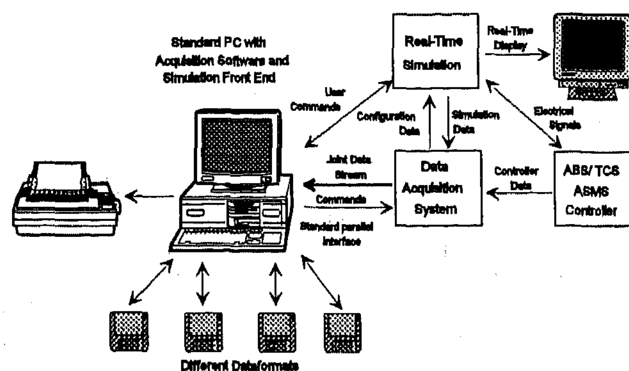


Figure 13: RTS with the ITT Data Acquisition System

Especially at the transition from prototype controller (floating point arithmetic) to series controller (integer arithmetic) and the integration in the existing ABS/TCS code, Real-Time Simulation was a very helpful tool to verify the fail-safe concept, that means electrical and physical range checking of sensor inputs, actuator outputs and the field bus communication. Before testing in the car the software engineers have the opportunity to check, if the modifications are correct and do not unintentionally influence other features of the software.

To ensure a simple use of the RTS, all parameters of the models and the hardware configuration settings for each controller are combined in projects. Furthermore, the simulation provides an appropriate set of standard maneuvers for such a project. With these maneuvers it is very easy to show whether the desired functionality and performance of the ABS, TCS or AYC is reached or not.

3D-COURSES - After describing the design of the AYC algorithm and the development of the series code with the help of RTS, the robustness of the resulting ASMS code under different external disturbances on the vehicle dynamics shall be discussed now. Such influences can e.g. be caused by non-plane driving courses like banked curves and serpentines. OLS offers the vehicle models and the 3D-courses to examine these effects. Due to real-time constraints, such simulations can not yet be carried out in RTS. In relation to experimental tests, OLS has the advantage that a large variety of banked

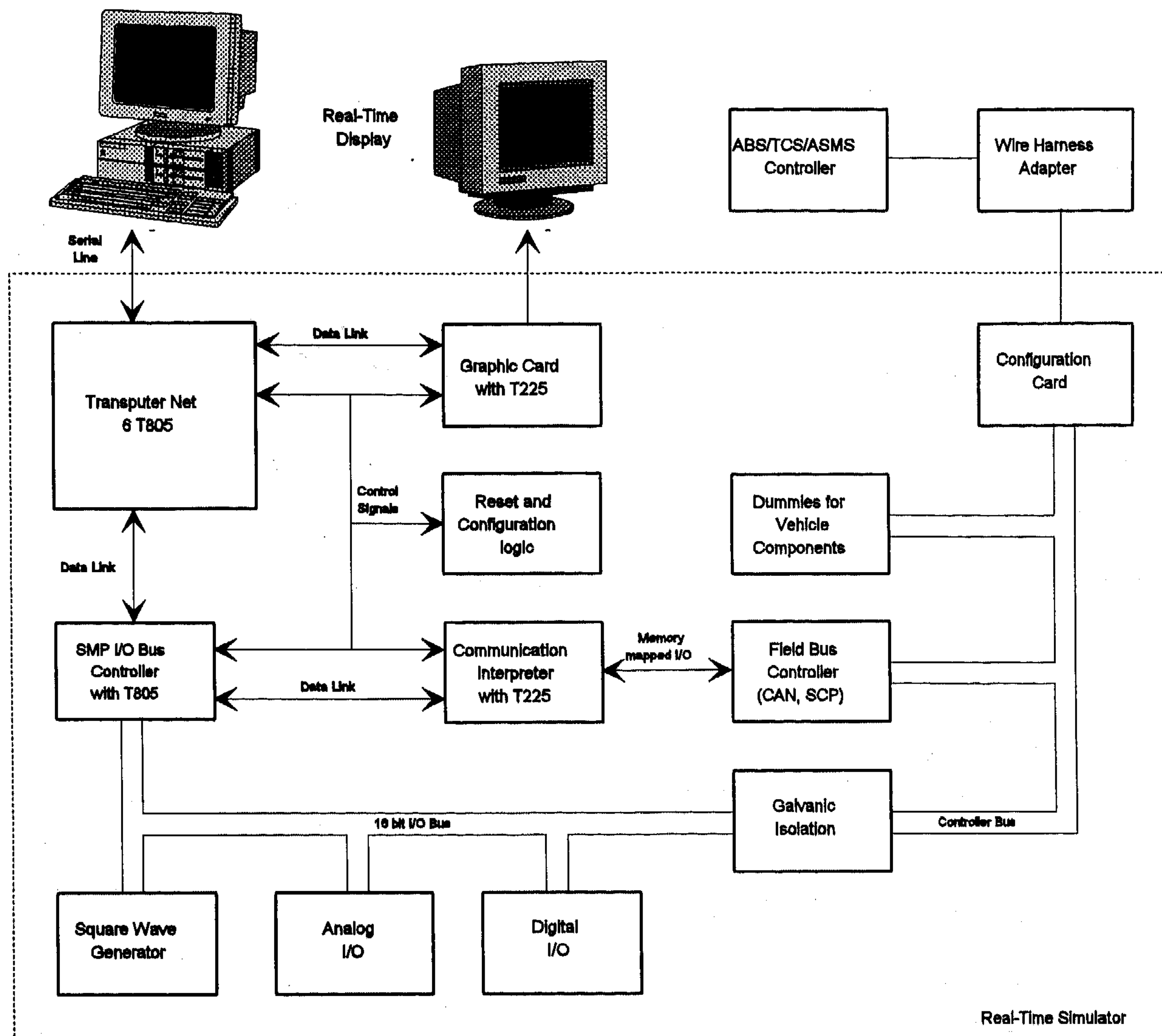


Figure 14: Block diagram of the Real-Time Simulation environment

curves or serpentine characteristics can be examined by simple parameter studies.

As a first example, some simulations shall be discussed that were carried out to study the impact of banked curves on the yaw behavior of a vehicle. Therefore a complex 3D-vehicle model described with the help of ADAMS is used to take 3D-motion effects into consideration.

Figure 15 illustrates the maneuvers. While the vehicle is driving in circle in a banked curve of constant inclination angle it is slowly accelerating to increase the yaw rate. The acceleration is low so that transient effects can be neglected. The driver module keeps the vehicle on a constant radius.

The ratio of the resulting yaw rate to the steering wheel angle as a function of the vehicle velocity is described at the top and the resulting vehicle acceleration at the bottom of figure 16. For a better comprehension of the maneuver dynamics the nominal ratio of yaw rate to the steering wheel angle of an appropriate parameterized linear one-track model is also plotted in the upper diagram.

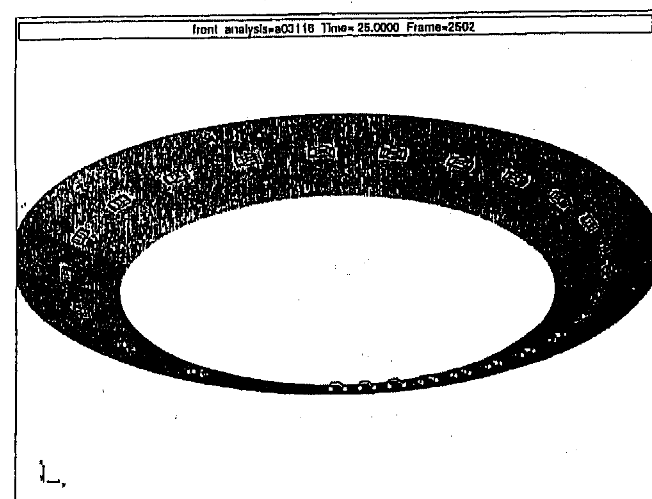


Figure 15: Closed loop driving maneuver in a banked curve

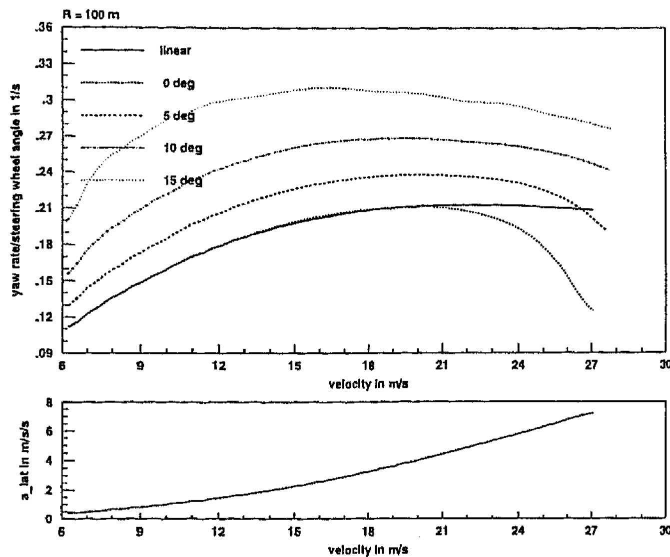


Figure 16: Oversteer/understeer behavior of a vehicle in a banked curve

Like it is expected, the yaw rate to steering wheel angle ratio of the linear one-track model and the simulated vehicle behavior on a banked curve with a 0 degree inclination are almost identical up to a lateral acceleration of 4 m/s². When lateral acceleration increases, a clear difference between the simulated vehicle dynamics and the linear single track model occurs which is caused by nonlinear behavior of the simulation model.

If there is a positive inclination of the road, the ratio of the resulting yaw rate to the steering wheel angle increases. That means that the same steady state yaw rate can be driven with less steering wheel angle. But on the other hand, this advantage of road inclination changes the dynamic behavior of a vehicle and may possibly cause undesired interventions of ASMS controllers if no adequate countermeasures would be integrated into the control algorithms. This is a typical example where OLS provides a deeper evaluation of control algorithms than driving tests can.

Another example to illustrate 3D-effects in the development of control algorithms is downhill driving on serpentines. Figure 17 shows a simple track-following maneuver on a snow-covered road with a rear driven vehicle. During this maneuver the driver runs into trouble caused by the drag torque of the engine. Here, the rear wheels increase their slip and the vehicle looses its stability. Normally an ordinary driver is not able to stabilize the vehicle by counter-steering.

The same maneuver with a rear driven vehicle which is equipped with ASMS is described in figure 18. In this case ASMS recognizes the tendency of instability and supports the driver with adequate countermeasures. This gives an ordinary driver the ability to follow the track easily.

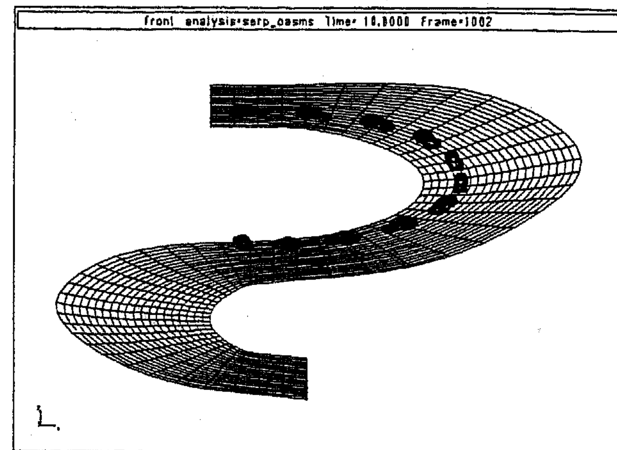


Figure 17: Closed loop driving maneuver on a serpentine without ASMS

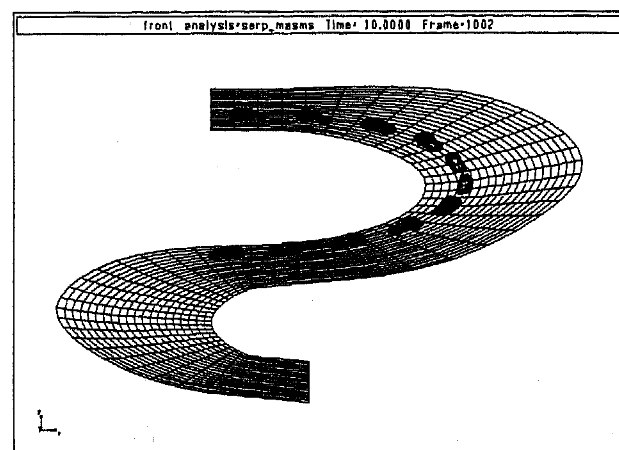


Figure 18: Closed loop driving maneuver on a serpentine with ASMS

CAR/TRAILER COMBINATION - In a similar way as it was described for 3D-courses, OLS offers the possibility to evaluate the robustness of ASMS for car/trailer combinations. Trailers cause additional forces on the vehicles so that the yaw torque around the vertical axis leads to undesired understeer/oversteer behavior. Tests with car/trailer combinations are hazardous for vehicle testing when driving with velocities that are on the limit of stability.

Influence of trailers on the lateral vehicle dynamics - In a first step some standard simulations without ASMS interaction were carried out to show the influence of trailers on the lateral vehicle dynamics. A passenger car/caravan combination and a passenger car/truck trailer combination were created in ADAMS, representing a heavy and light class of trailers. The models were evaluated in comparison to measurement data from literature.

The following example deals with the steady state behavior of the vehicle The velocity is kept constant. As steering wheel input, a ramp over the interesting range of steering wheel angles is given. The gradient of the ramp is small so that transient effects can be neglected.

Figure 19 illustrates the required steering wheel angles to achieve a certain lateral acceleration. Obviously the heavy caravan as the trailer causes the largest deviations to the normal steady state behavior of the vehicle. Consequently the following investigations will focus on the car/caravan combination.

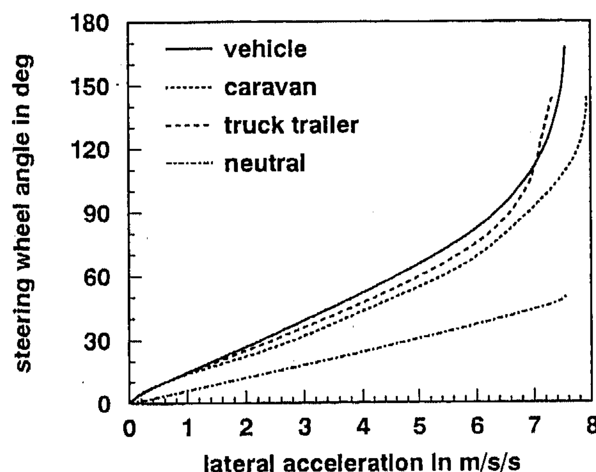


Figure 19: Steady state understeer/oversteer behavior of a passenger car with trailers: lateral acceleration

Lateral stability test - For the modelled passenger car/caravan combination without ASMS a lateral stability test was carried out according to the ISO 9815 standardized criteria. The damping characteristics of the car/trailer combination as the result of the lateral stability test is shown in figure 20. It is characterized by the reference speed damping of 0.1 at 80 km/h, the reference damping speed of 88 km/h (the damping equals 0.05) and the zero damping speed of approximately 93 km/h which means that velocities higher than 93 km/h can lead to trailer oscillations with growing amplitudes of the articulation angle between the towing car and the trailer.

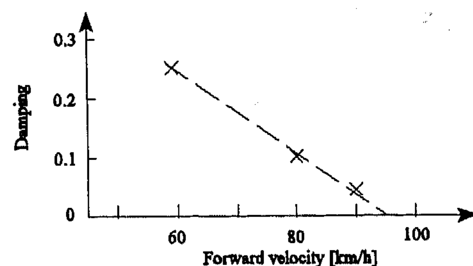


Figure 20: Damping characteristics of the articulation angle of the passenger car/caravan combination

Two selected maneuvers are presented below to demonstrate the effect of ASMS to the passenger car/trailer combination.

Impulse response with ASMS - As a first maneuver the impulse response according to ISO 9815 was simulated on a dry pavement. Figure 21a shows the steering wheel input. The other charts of figure 21 compare the maneuver with and without ASMS. Figure 21b shows the forward velocity. It is chosen as the zero damping speed at the limit of stable trailer behavior.

Figures 21c-e compare the articulation angle between the towing passenger car and the trailer and the yaw rates of car and trailer. The oscillation of the signals decreases with ASMS. Without ASMS the maneuver is instable.

Figure 21f shows the wheel individual brake pressures that are generated by ASMS to reduce the undesired effects of the trailer. Up to a certain degree the ASMS pressure modulation compensates the additional yaw torque around the vertical axis of the car which is caused by external forces of the trailer. Moreover, ASMS has a positive influence on the oscillation of the whole passenger car/trailer combination. The superimposed animation of the maneuver is shown in figure 21g.

Double step response with ASMS - The second maneuver is again carried out on dry pavement. The trailer oscillation is caused by steering to the left, and after a short period of time, by counter steering to the right. The driver input and the velocity are plotted in figures 22a and 22b. The superimposed animation of the passenger car/trailer behavior is illustrated in figure 22g. It shows that the maneuver with ASMS is stable whereas without ASMS the whole combination spins around and slides. The trailer imposes additional forces on the towing passenger car that cannot be compensated without ASMS. The individual wheel pressure modulation of ASMS as it is shown in figure 22f reduces the undesired effects so that the stability is improved. These effects are also illustrated in figure 22c-e. The articulation angle and the yaw rates both of the car and the caravan are much smoother with ASMS.

SENSORS - The simulations above have mainly focused on the robustness of ASMS. However, simulation was also contributed to hardware specifications. The influence of the sensor dynamics, filtering, A/D-conversion and sensor failures on the performance of ASMS can easily be studied in OLS. The following example demonstrates, how the tolerable misalignment of the yaw rate sensor can be specified based on a parameter study in OLS.

Apart from a small sensitivity error the main effect of misalignment is that portions of the pitch and roll motion of the body are superimposed to the yaw rate signal. To study these effects an extreme maneuver at a velocity of 70 km/h was chosen, that is considered to be the worst case. A sine signal of 1.5 Hz with an amplitude of 180° as steer input causes a strong roll motion of the vehicle body. Furthermore a driver brake intervention of 140 bar between 1.0 and 2.0 s of simulation time is simulated to stimulate a pitch motion.

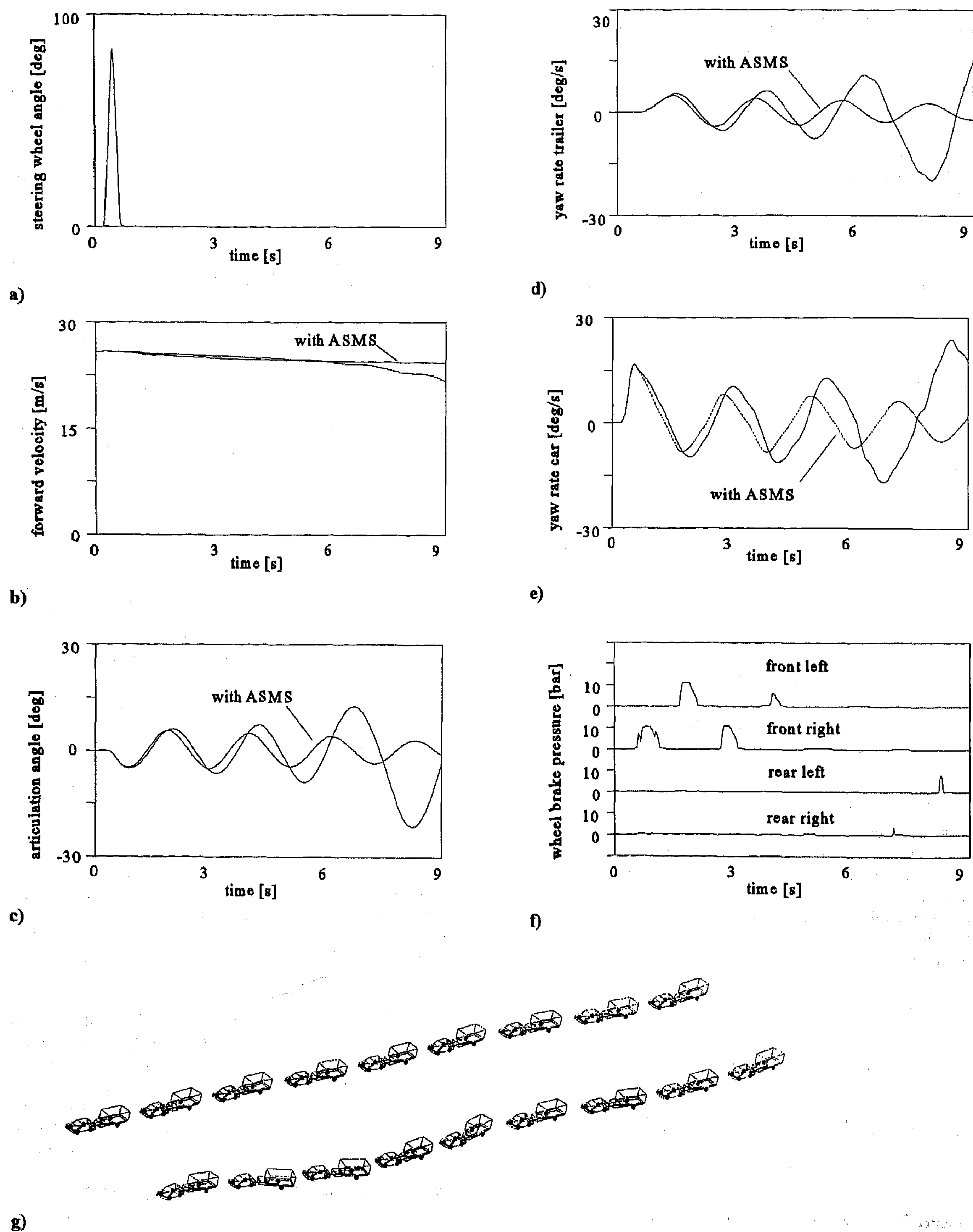


Figure 21: Impulse response

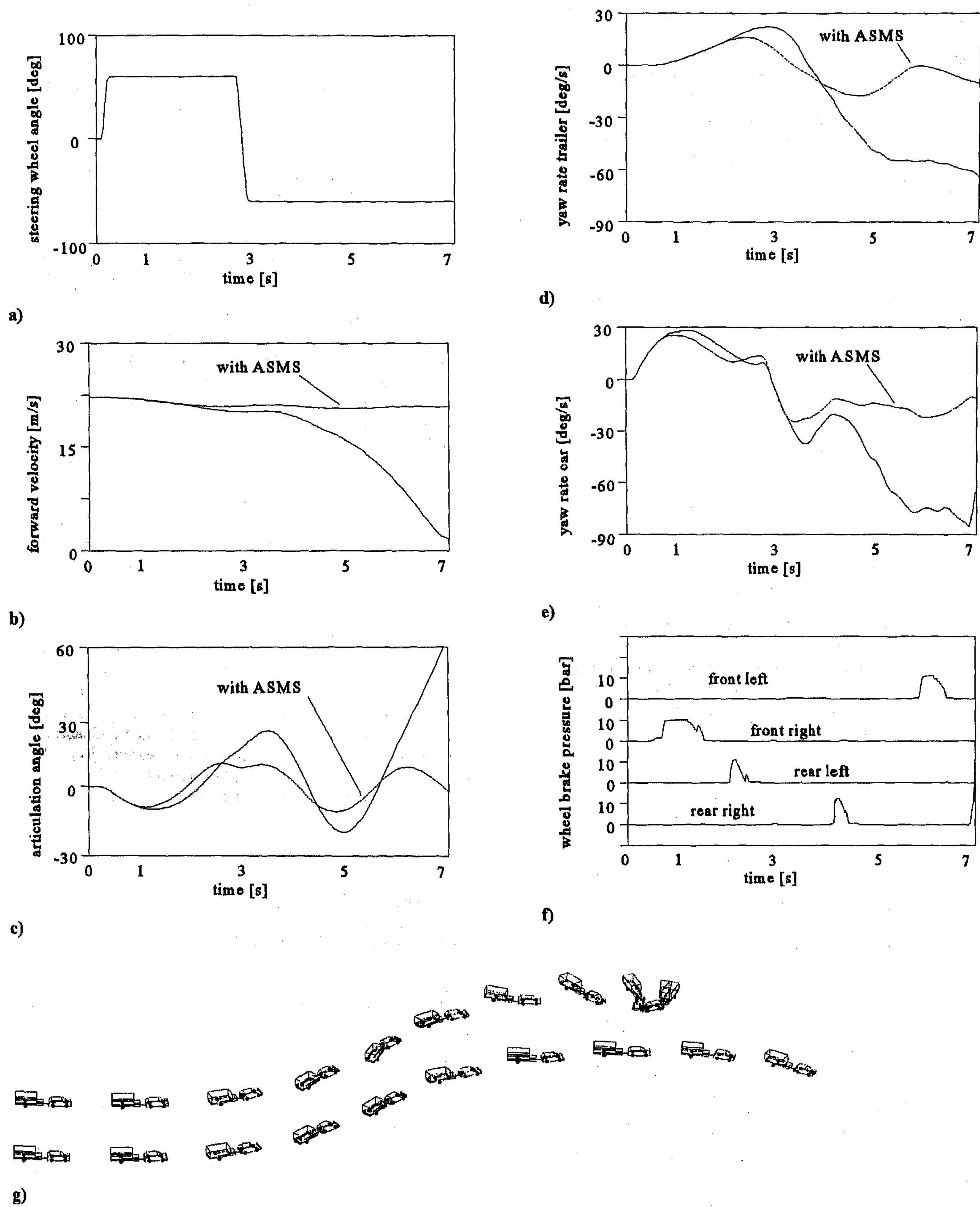


Figure 22: Double step response

Figure 23 shows the yaw, pitch and roll rate of the body calculated in the body frame. In an automated loop the yaw rate signals of the maneuver with all combinations of misalignments around the longitudinal and lateral axis of the body in steps of 1.0 between -10 and +10 have been calculated. The maximum of the absolute error with respect to the real yaw rate for each simulation was evaluated. Finally, the error was plotted as a function of the misalignment along the longitudinal and lateral axis.

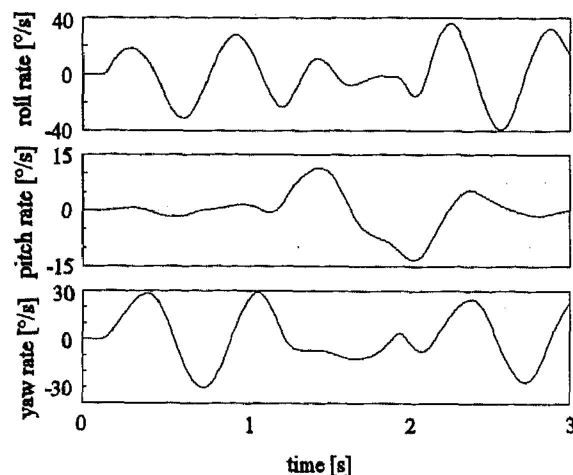


Figure 23: Angular velocities of the body

Figure 24 shows lines of constant error that have been interpolated from the simulation results. The dashed circle shows that with a tolerance of 5 misalignment a yaw rate error of 3.5 /s in the worst case has to be accounted for in the AYC-control algorithm. Furthermore, the parameter study shows, that the sensitivity against misalignment around the lateral axis is higher than around the longitudinal axis.

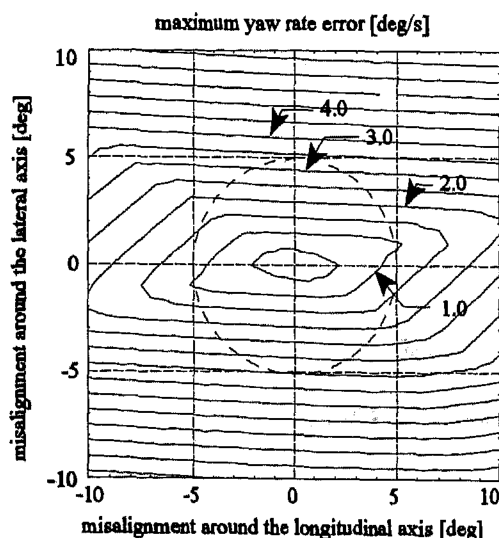


Figure 24: Maximum yaw rate error by misalignment

SUMMARY AND CONCLUSIONS

Simulation has rendered an important contribution to the development of the new Automotive Stability Management System (ASMS) from ITT. In addition to the ABS and TCS function, ASMS includes an Active Yaw Control (AYC) that assists the driver when the vehicle tends to instability. Simulation has offered the possibility of continuous development in the laboratory. Up to a certain level the design of basic control strategies has been carried out in simulation. Simulation has relieved vehicle prototype testing from time consuming work. The development of the code for the series micro controller in integer arithmetics is supported by Real-Time Simulators which combine the series controller hardware with simulation models of the vehicle dynamics and all other components of the control loop. Modifications in the code could be verified by simulating standard maneuvers before testing in the vehicle. In Off-Line Simulation special attention has been paid to those issues which are difficult to study in testing. As it has been pointed out in the examples of the paper, external influences on the vehicle dynamics due to disturbances e.g. by trailers or 3D-courses can be examined in simulation. Simulation offers an insight to dynamic effects that can hardly be obtained in testing. It is possible to configure maneuvers and disturbances arbitrarily. The results are reproducible. Parameter studies with a huge number of simulations can be executed and evaluated automatically. Nevertheless, the functionality and performance of the code always have to be proven by vehicle prototype testing.

The growing functionality and complexity of the vehicle control systems demand more and more testing and simulation efforts for the design and verification of algorithms. The progress in simulation software tools, computer performance and modeling of the vehicle dynamics will offer new possibilities to simulation. The step from the initial idea of an algorithm that is developed in simulation to the verification by vehicle prototype testing will become easier. Regarding shorter product life cycles "Rapid Prototyping" will become a key issue in the development of new vehicle control systems.

REFERENCES

[1] Fennel, H.; Mahr, S.; Schleysing, R.: Transputer-Based Real-Time Simulator- A High Performance Tool for ABS and TCS Development, SAE Paper 920643, 1992

[2] Dugoff, H. et al.: An Analysis of Tire Traction Properties and Their Influence on Vehicle Dynamic Performance, SAE Paper 700377, 1970

[3] Pacejka, H. et al.: Tyre Modeling for Use in Vehicle Dynamics Studies, SAE Paper 870421, 1987

[4] Judaschke, U.: Verfahren zur kollisionsfreien Führung und Koordination mobiler Transportsysteme, VDI Fortschritts-bericht, Reihe 12, Nr. 225, pp. 16 - 30, VDI-Verlag GmbH, Düsseldorf, Germany, 1994, ISBN 3-18-322512-3

971060

The Effects on Motorcycle Behavior of the Moment of Inertia of the Crankshaft

Takeshi Kimishima, Toyokazu Nakamura, and Takahisa Suzuki
Honda R&D Co., Ltd.

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

The moment of inertia of the crankshaft cannot be ignored when analyzing the dynamics of a motorcycle. In this research, the tire friction force (calculated by drag and tire side force) was used as an index of the drive performance. The ratio of roll rate and steering torque (here after referred to as a roll rate gain) was used as an index of the cornering performance, and it was analyzed as the influence of the moment of inertia of a crankshaft on the drive performance as well as cornering performance.

As a result, the influence on drive performance and cornering performance by the moment of inertia has been found.

INTRODUCTION

An example of the theoretical research of the relationship to the motion of motorcycles and the gyroscopic effects is as follows:(1). When comparing motorcycles with automobiles, the contribution rate to the motion of motorcycle by the rider's body position is much greater. Moreover, in turning while banking, the analysis of motorcycle motion is difficult.

In recent years, however, technological advances have allowed higher engine speed, larger engine displacement, and a lighter engine. The effect of the moment of inertia of the crankshaft on a motorcycle has become an important resarch subject.

In this research, the changing moment of inertia of the crankshaft, and the analysis of the drive performance and cornering performance of a motorcycle was carried out from the viewpoint of engine gyroscopic and flywheel effect.

THEORETICAL ENGINE GYROSCOPIC EFFECT

GYROSCOPIC MOMENTS

Straight Running - The inertial effects by which a rotating object tries to maintain the same direction of the rotation axis and the precession which generates the gyroscopic moments in the rotation axis of that object, when the object rotates on an orthogonal axis are characteristics of that object at high speed. A simple model is shown in Fig.1.

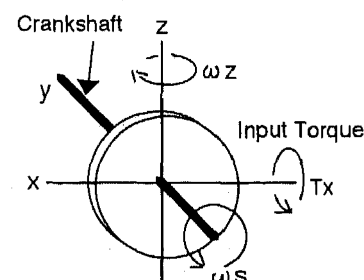


Fig.1 Direction of Generated Gyroscopic Moments

$Ic \times \omega s \times \omega z = Tx$ ----- (1)

Tx : Input Torque on Roll axis
Ic: Moment of inertia of crankshaft (kgm/s^2)
ωs: Engine speed (rad/sec)
ωz: Yaw rate (rad/sec)

In the case of a motorcycle that is running straight as the small moment Tx is input around the roll axis, a gyroscopic moment occurs around the Z axis as shown in formula (1). If Tx is assumed to be a disturbance, the angular moment of the crankshaft $Ic \times \omega s$ represents the stability against the disturbance.

Initiate Action - When attempting to lean the motorcycle, the moment that is input around the roll axis is given as a function of steering torque (Ts) and the moment due to the movement of the rider (Tr).

$$Tx = f1(Ts)+f2(Tr) \text{ ----- } (2)$$

f1(Ts):Moment generated by steering torque
f2(Tr):Moment generated by movement of the rider

It is assumed in the test mode of this study that the moment due to the movement of the rider does not vary depending on the moment of inertia of the crankshaft. This is becauseit is difficult to measure the moment due to the movement of the rider. Attention was paid to the steering torque.

When initiating action to turn from straight running, the moment that is input around the roll axis can be expressed by formula(2)

$$Ix \times \frac{d\omega x}{dt} + Ic \times \omega s \times \omega z = Tx \text{ ----- } (3)$$

Ix: Moment of inertia of Roll axis (kgm/s^2)
ωx: Roll rate (rad/sec)

This shows that (Tx) is divided into the gyroscopic moment around the Z axis, and that the moment lets the body turn in the rolling direction. The complete model is shown in Fig.2

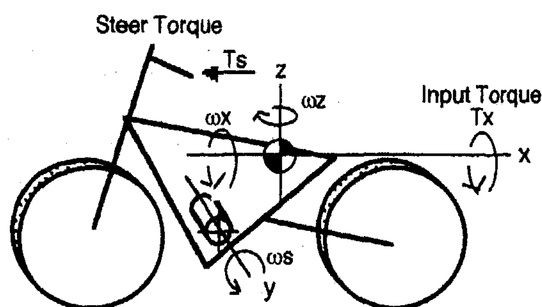


Fig.2 Complete Model

If a motorcycle is consided in a simplistic manner, it can be defined as a two axis free gyro that is supported by contact points of the front and rear tires and is free to yaw and roll.

Wheel and Engine Gyroscopic Moment - Fig.3 shows the engine/transmission system of a typical motorcycle. The gyroscopic moment on motorcycle of 750cc class by which the engine speed of 6000 rpm and the roll rate of 30deg/sec are shown in Fig. 4. The gyroscopic effect of the wheels are larger than the gyroscopic effect of the engine. However, about 9% of all gyroscopic moments that are generated take place within the crankshaft of the motorcycle. The mainshaft has the opposite gyroscopic effect, while the influence on cornering performance cannot be ignored.

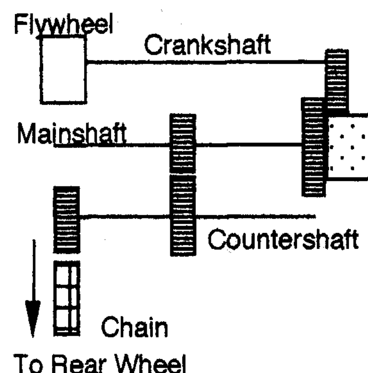


Fig.3 Engine Transmission System of a Typical Motorcycle

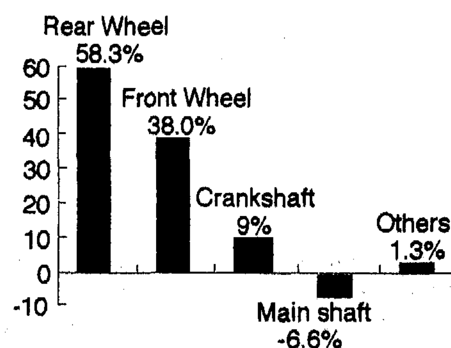


Fig.4 The Gyroscopic Moment Ratios in a Motorcycle with the Transmission System in Fig.3

THEORETICAL FLYWHEEL EFFECT

REAR TIRE FRICTION CIRCLE- The ability to controll engine speed due to throttle operation is enhanced by the adequate moment of inertia of the crankshaft. When this is achieved, the throttle can be operated with ease and the rider of the motorcyclle will be able to control it without exceeding the tire friction circle, as well as balancing drag and side force.

In the research, attention to drag and side force during actual riding and flywheel effect regarding drive performance while turning, were also analyzed.

Fig.5 shows the calculation model. During acceleration a moment is created when the front of the motorcycle is raised. Therefore, changes of axle load (Wb) and rear axle load (Wr') are expressed in formula (4) and (5).

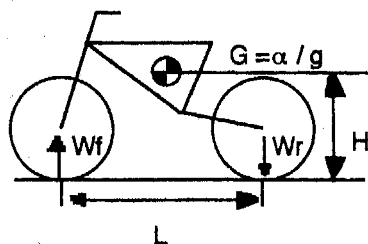


Fig.5 Calculation Model for Change of Axel Load

$$W = (\alpha / g) \times H / L \times W \quad \text{----- (4)}$$

$$Wr' = Wr + Wb \quad \text{-----------------(5)}$$

W: Curb weight
Wr: Rear axle load
Wf: Front axle load
Wb: Change of axle load distribution
a : Acceleration
H: Center of gravity height
L : Wheel base
G : Acceleration G

The side force of the front and rear tires act as the force that balances the centrifugal force on the center of gravity of the motorcycle while it is turning. The tires then generate a force proportional to the contact load. The side force of the front and rear tire can be calculated by simply considering the amount of change in the axle load of the front and rear wheel during turning acceleration and deceleration.

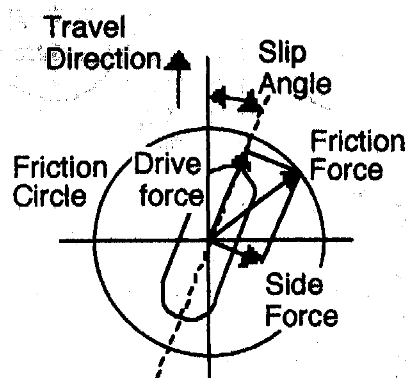


Fig.6 Concept of Friction Circle

$$\mu W \geq F = \sqrt{S^2 + T^2} \quad \text{----------- (6)}$$

F: Composite Force
S: Side force on tire
T: Drag
W: Contact load on tire
μ: Road surface friction coefficient

Fig.6 shows the concept of a friction circle. Formula (6) pertains to the action of drive force while turning. Formula (6) shows that the tire friction force cannot exceed the tire friction circle radius (limit value). The tire friction circle radius is the force resulting from drive force and side force calculated by the contact load and the friction coefficient of road surface.

NUMERICAL SIMULATION

In order to examine the influence of the gyroscopic effect of the crankshaft from a theoretical point of view before the actual running test has been performed, the simulation of a simplified model and motorcycle model using mechanism analysis language was administered through the gyroscopic effects of the engine (2).

SIMULATION OF ROLLING MOTION THROUGH A SIMPLISTIC MODEL - Fig. 7 shows a simplified model. The masses that are assumed as A, B, C, D are shown in Table 1. They are arranged on a rectangular frame which is representative of the body of a motorcycle. The rolling behavior was generated by roll torque being input at the center of gravity of this model, which is assumed to be floating weightless in space. The simulation was executed by four specifications shown in Table 2.

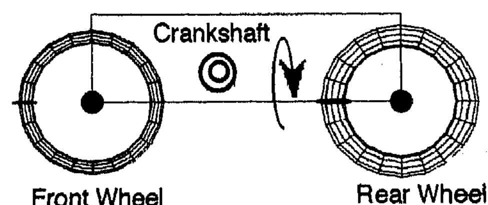


Fig.7 Simplified Model

Table 1 Moments of Inertia
$Kg \cdot m \cdot s^2 \cdot 10^{-4}$

Front Wheel	446.0
Rear Wheel	701.0
Crankshaft	7.7

Table 2 Calculation Specifications

	Front Wheel	Rear Wheel	CrankShaft
A	Not Rotated	Not Rotated	Not Rotated
B	Rotated	Not Rotated	Not Rotated
C	Rotated	Rotated	Not Rotated
D	Rotated	Rotated	Rotated

(The rotation speed of each part corresponds to motorcycle velocity 60km/h.)

Fig. 8 shows the result of the simulation. The roll rate is increased by the input torque and eventually reaches a constant value if the gyroscopic effect is not useful, as in case A Table 2. A yaw rate was not generated.

When the gyroscopic effect is sufficient, as in case B, C, and D in Table 2, the gyroscopic moment is generated by the roll rate. The body of the model is rotated to turn with the yaw axis and the roll rate has been decreased. As the gyroscopic effect increases, the roll rate continues to decrease and the maximum value of the roll rate becomes smaller.

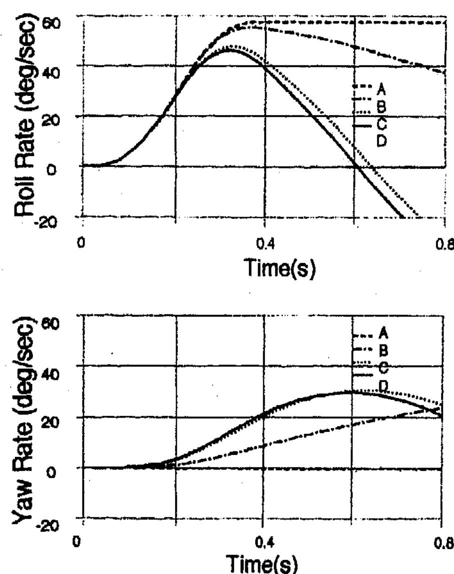


Fig.8 Result of Simplified Model Simulation

SIMULATION FOR INITIATING A TURN BY MOTORCYCLE MODEL - Fig. 9 shows the motorcycle model, and the front and rear suspension and the degree of freedom for the rotation of steering system are given.

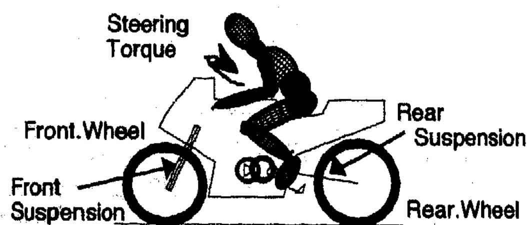


Fig.9 Motorcycle and Rider Model

The simulations of transient characteristic for initiating a turn by the input steering torque shown in Fig. 10 at 60 km/h were compared with the moment of inertia of the crankshaft shown in Table 3. (There was no input by the body movement of rider. The rider model was fixed on the motorcycle model.)

Table 3 Moment of Inertia of Crankshaft

$Kg \cdot m \cdot s^2 \times 10^{-4}$

Case.1 (Small Mass)	7.7
Case.2 (Large Mass)	17.7

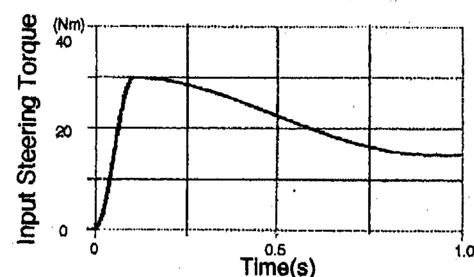


Fig.10 Input Steering Torque

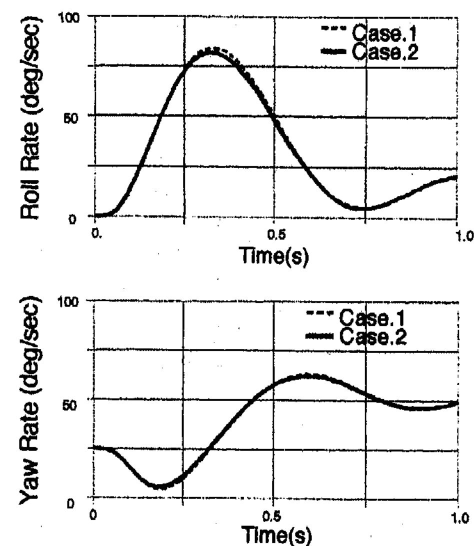


Fig.11 Result of Motorcycle Model Simulation

The maximum value of roll rate was found to be smaller with a larger moment of inertia of the crankshaft, as shown in Fig. 11. From simulation results mentioned above, there is an influence on the rolling motion during the initiation of a turn when the moment of inertia of the crankshaft increases, while the roll gain decreases.

OUTLINE OF TEST

TEST MOTORCYCLE AND TEST RIDER - The motorcycle used in this test is a sports type. The engine displacement of 750cc shown in (Fig.12), as well as the parameters are shown in Table 4. The test rider was a skilled rider who had an understanding of motorcycle dynamics.

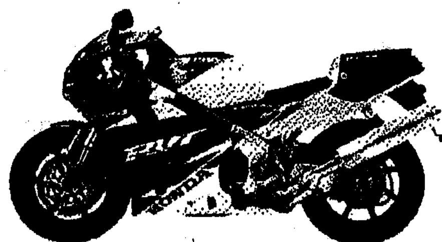


Fig.12 Test Motorcycle

Table 4 Specifications of Test Motorcycle

Curb Mass (kg)	211
Wheel Base (mm)	1410
Caster Angle (deg)	24.5
Trail (mm)	92
Front Tire Size	130/70ZR16
Rear Tire Size	190/50ZR17
Displacement (cm^3)	749

SPECIFICATIONS OF MOMENT OF INERTIA - The specification of the moment of inertia were changed by replacing a flywheel mass on the left end of the crankshaft (shown in Fig. 13). Specifications that concern the moment of inertia are shown in Table 5. The testing of cornering performance was executed by using specifications 1, 2, 3, and 4. Specifications 1, 2, and 3 were also used in order to test drive performance.

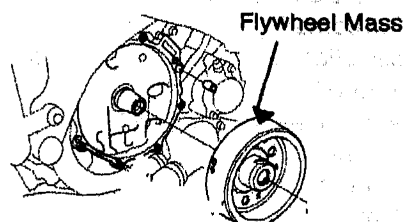


Fig.13 Install of Flywheel Mass

Table 5 Moment of Inertia

	Crankshaft $Kg{\cdot}m{\cdot}s^2 \times 10^{-4}$	Mainshaft $Kg{\cdot}m{\cdot}s^2 \times 10^{-4}$
Specification1	7.72	17.57
Specification2	9.90	17.57
Specification3	11.73	17.57
Specification4	9.90	29.08

TEST MODE - The test was conduced using three separate procedures shown in Table 6. The items for subjective evaluation are shown in Table 7, Table 8 displays the subjective rating.

Table 6 Test Method

Cornering Performance : Loop Turn

Test Pattern	Steady turn; 45R; Terminate action; Initiate action; Velocity : 60 km/h
Test Method	· turning of radius 45m at 60km/h · use 2nd gear

DrivePerformance : Acceleration in a turn

Test Pattern	Width 10m; Velocity 60 k/h; 20 R; Passage point; Velocity 100 k/h; Velocity 100 k/h
Test Method	· accelerating in testing lane of radius 20m and width 10m · running through the passage point

Table 7 Feeling Evaluation Item

Test Mode	Feeling Evaluation Item
Cornering Performance	· Cornering Performance · Perceived Steer Effort
Drive Performance . Acceleration in turning	· Acceleration · Traction Feeling · Throttle Controllability

Table 8 Subjective Rating Scale

Subjective Rating Scale (Cornering Performance)

Subjective Rating	Light	Rather Light	Average	Rather Heavy	Heavy
Rating Scale	5	4	3	2	1
	sensitive ←				→ insenstive

Subjective Rating Scale (Others)

Subjective Rating	excelent	good	fair	poor	bad
Rating Scale	5	4	3	2	1

TEST RESULTS

CORNERING PERFORMANCE - Fig.14 illustrates an example of the measurement data used in the analysis of cornering performance based upon the results of the simulation. The results of the subjective evaluation test on cornering performance and steer effort based on a 5-point scale are shown in Fig.15. Attention was given to cornering ability (ease in cornering).

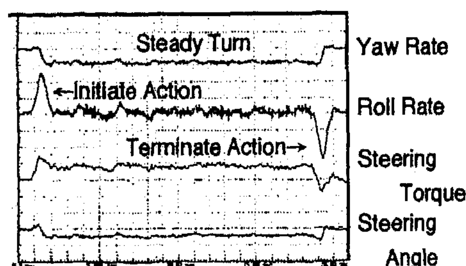


Fig.14 Example Data of Loop Turn

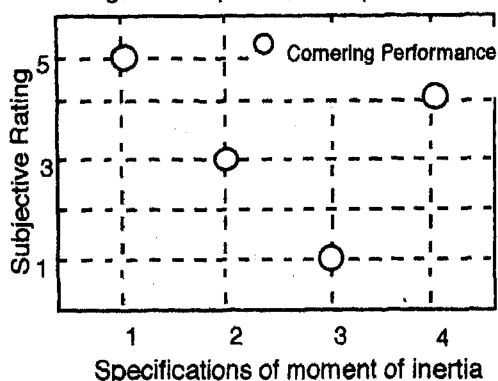


Fig.15 Subjective Rating of Cornerint Performance

As a constant to represent the cornering performance, the yaw rate convergence time as shown in Fig.16 was defined and its values in different specifications of the crankshaft were compared.

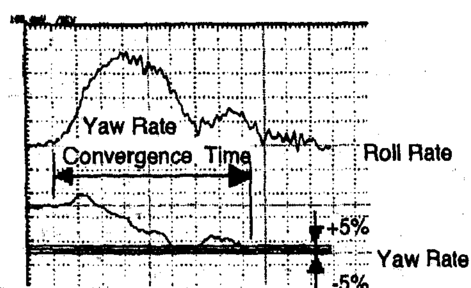


Fig.16 Yaw Rate Convergence Time

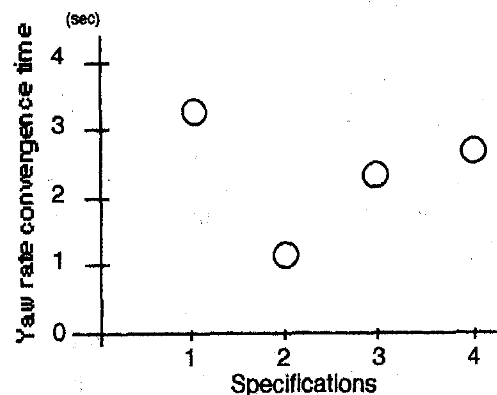


Fig.17 Comparison of Yaw Rate convergence Time

Fig.17 shows a comparison of yaw rate convergence time. Specification 2 has a short yaw rate convergence time and can obtain cornering stablity quickly. Specification 1 has a long yaw rate convergence time and while turning, is increased in time to obtain cornering stability. From this result, one can say that a motorcycle which has an adequate moment of inertia of the crankshaft can quickly obtain cornering stablity.

We considered the difference betweeen yaw rate convergence time by the ratio of the roll rate and steering torque. Comparisons of roll gains using specifications 1, 2, 3, and 4 are demonstrated in Fig. 18.

As for specification 1, the roll rate to the steering torque oscillates and the maximum value of the roll rate is smaller than other specifications. Apparently, the the rider is unable to steer well because of the sensitivity of cornering performance.

The steering torque and roll rate of specification 2 are larger than that of specification 1 and 3. It is possible that the rider can actively steer the motorcycle due to the fact that the cornering performance is only moderately sensitive.

The roll rate gain of specification 3 is smaller than 2 because the cornering performance of 3 is less sensitive than specification 2. Furthermore, in a steady turn the steering torque becomes agitated and unstable.

In specification 4 both the steering torque and roll rate were increased by an enlargement of the moment of inertia in the main shaft. This enlargement in the mainshaft incrased more than specification 2. However, the steering torque during a steady turn remained unstable.

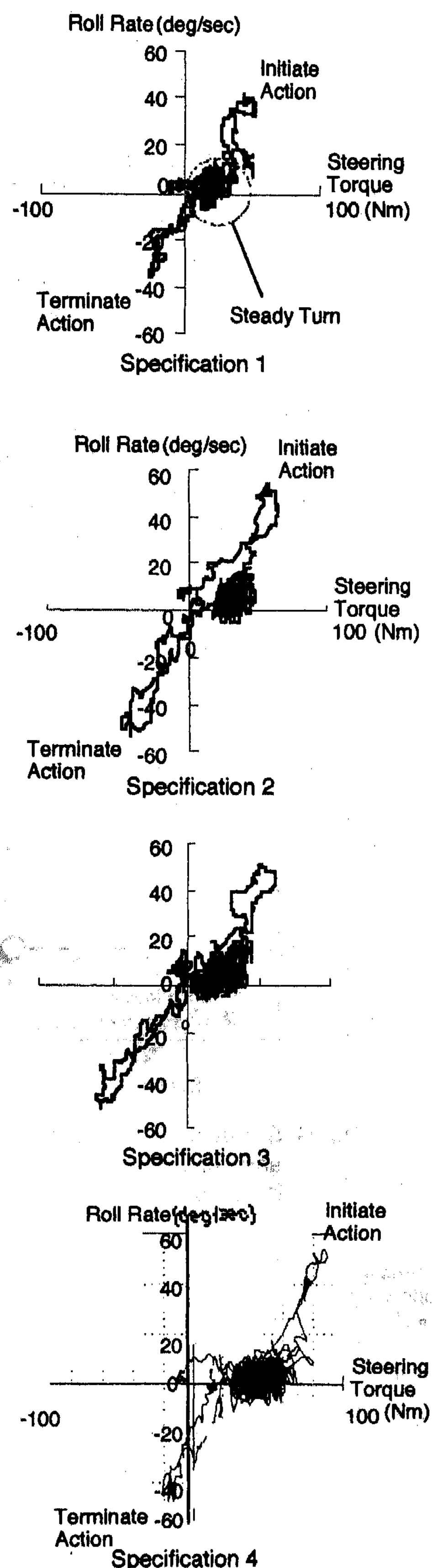


Fig.18 Comparison of Roll Gain

In Fig.19, the linear regression calculation excludes the steady turn portion of the roll rate gain data. The linear regression coefficient is defined as the roll gain coefficient. The results are demonstrated in Fig. 20. Without specification 1, the roll gain coefficient coincides with the inclination to subjective evaluation in cornering performance. The sensitivity and insensitivity of cornering performance will be expressed by the roll rate gain coefficient.

Fig.20 illustrates that the roll rate gain coefficient of specification 3 was smaller than specification 2, and the specification 3 exhibited insensitive cornering performance in the subjective evaluation. Therefore, the cornering performance will be insensitive due to the increase in the moment of inertia of the crankshaft.

Moreover, becauseof the fact that specification 4 is seen as having a more sensitive cornering performance than specification 2, one can say that an increase in cornering performance sensitivity correlates with an inclease in the moment of inertia of the main shaft. This is because the gyroscopic effect of the main shaft acts contrary to the crankshaft.

As mentioned above, the influence of the moment of inertia of the crankshaft on cornering performance is assesed by the roll rate gain.

Fig. 21 shows the comparison of the gyroscopic moment of the crankshaft and main shaft for each test specification. The gyroscopic moment of the crankshaft and mainshaft are almost the same as specification 2. Specification 2 receives a suitable subjective evaluation result in cornering performance. Thus, it balabces both gyroscopic moments and results in good cornering performance.

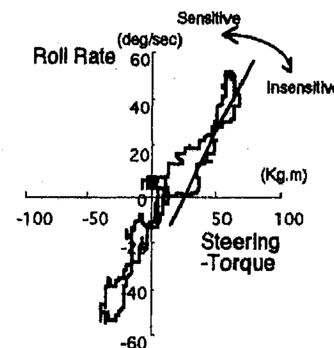


Fig.19 Example for Linear Regression

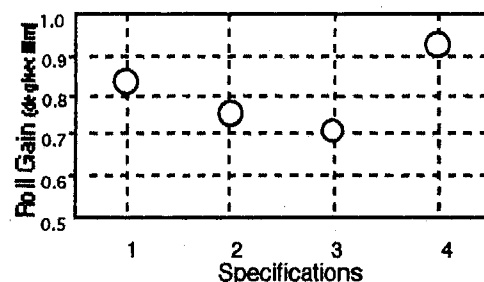


Fig.20 Comparison of Roll Gain

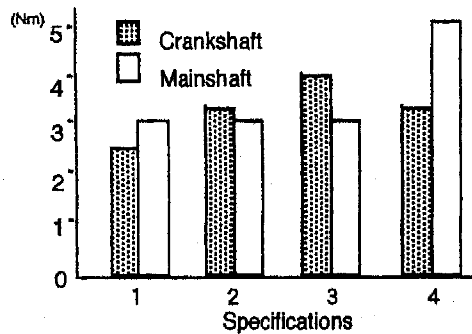


Fig.21 Comparison of Gyro Moment of Crankshaft and Mainshaft

Due to the gyroscopic effect, the moment of inertia of the crankshaft becomes larger, and the stability increases. This is demonstrated in the theoretical formula. Therefore, roll rate generated by the steering torque becomes smaller, and the roll rate gain decreases. From the results illustrated above, one can say that an optimum moment of inertia of the crankshaft provides good cornering performance.

DRIVE PERFORMANCE - Although it is known that reduced moment of inertia of the crankshaft enhance responses from the engine and provide greater linear accelerating performance, the response becomes too sensitive during cornering acceleration, which causes driving to be difficult. In this paper, the flywheel effect and the acceleration characteristics of cornering were studied in order to determine an adequate moment of inertia of the crankshaft.

Chassis Dynamometers Measurement - Fig.22 shows the waveforms of torque on the crankshaft and the countershaft of the transmission as measured using a chassis dynamo. Each torque fluctuation of specification 1 is larger than in specificatio 2. This is considered to be the cause of difficulty in throttle controlability.

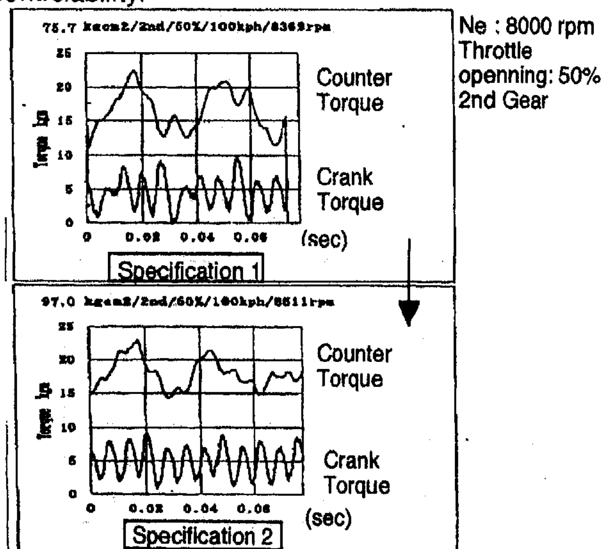


Fig.22 Crankshaft and Countershaft Torque Waveform

Acceleration mode in turn - Fig.23 displays an example of measurement data. The results of the evaluation of acceleration, throttle control, and drive feelilng in a turn utilizing a 5-point scale can be seen in Fig.24.

Specification 2 obtained a favorable subjective evaluation result in cornering performance, and overall test results that were quite good.

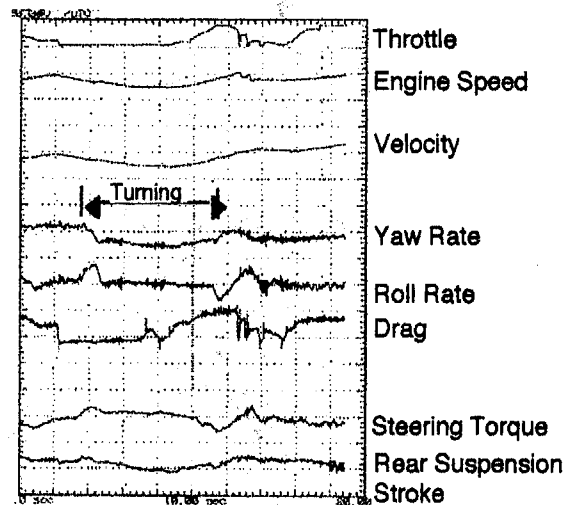


Fig.23 Exsample Data Acceleration in Turns

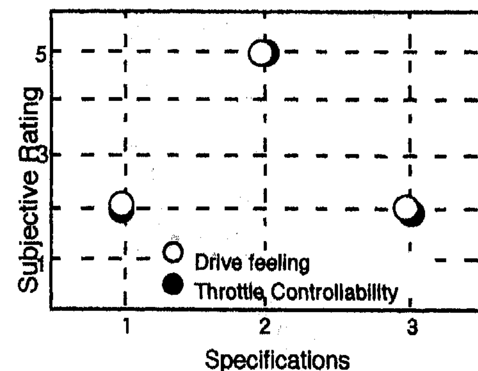


Fig.24 Subjective Evaluation of Acceleration and Throttle Control and Traction in Turns

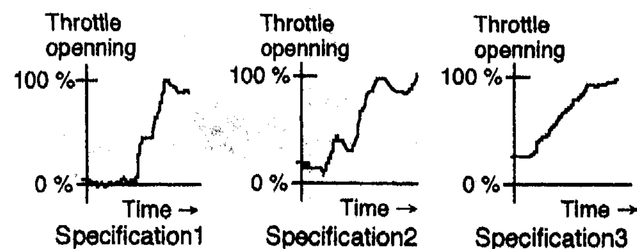


Fig.25 Comparison of Throttle Time Wave form

Changes in throttle operation over time are shown in Fig.25. It demonstrates that the rider cannot open the throttle due to an overly sensitive response in specification 1. The throttle of the motorcycle in specification 3 can open earlier, however, the engine response is insensitive and the motorcycle is unable to run at a highspeed.

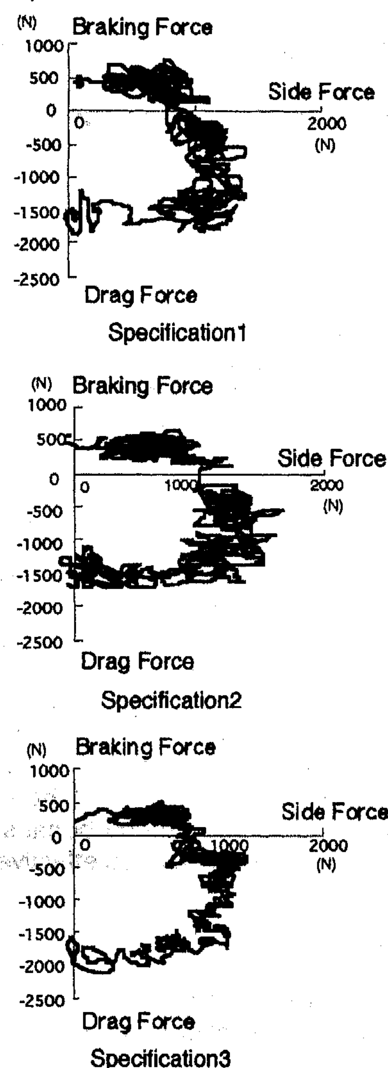


Fig.26 Comparison of Side Force and Drag in ActualRunning

The diagram for driving force and side force of the rear tire is demonstrated in Fig. 26. Side force increases rapidly because the moment of inertia of the crankshaft in specification 1 is small. In specification 2, the driving force and side force are balanced and a large tire friction circle is obtained. The tire friction circle of specification 3 was smaller than specification 2.

Fig. 27 shows an average value of the tire friction force calculated by formura (6),(7) from the initiation to the termination of the turn in each specification.

$$\bar{F} = \frac{\sum \sqrt{S^2+T^2}}{N} \quad \text{------------ (7)}$$

$\bar{F}$: Average Composite Force

S: Side force on tire

T: Drag

N : Number of Datas

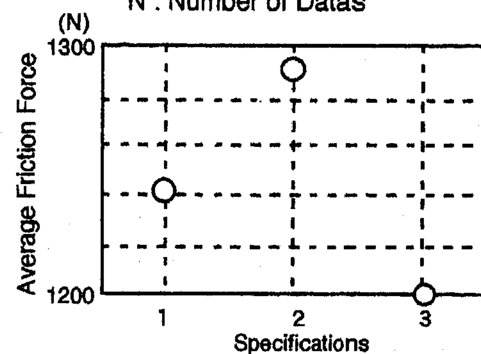


Fig.27 Comparison of Average Friction Force

The driving force and side force are effectively used in specification 2. Therefore, it can be said that the tire friction force will influence the traction.

Investigation of Tire Friction Circle Limit - If the moment of inertia of the crankshaft is reduced, the engine response becomes sensitive. This results in driving difficulty, so sliping will occur easily. Therefore, we studied the threshold for slippage from the test results on a tire (3).

Fig.28 is an illustratin of the tire friction circle. The relation to the tire friction circle limit was investigated by using the tire friction force during actual running and the tire characteristic data used in this research. The camber angle was fixed at 20 degrees, and the measurement of the tire characteristic was executed.

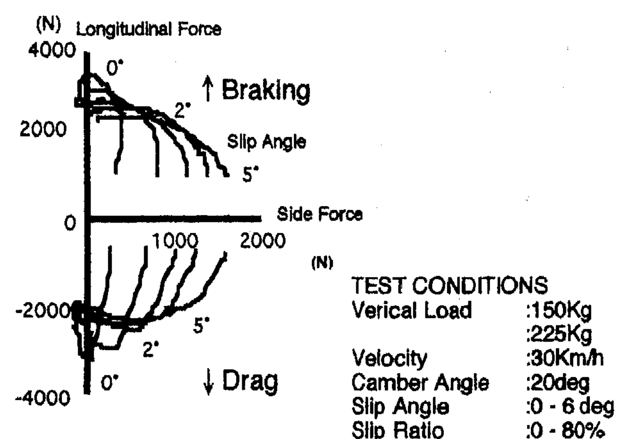


Fig.28 Measured Tire Friction Circle

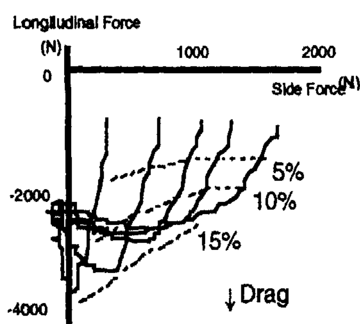


Fig.29 friction circles for each slip ratio

Fig.29 shows the tire friction circle for each slip ratio. The actual running data of specification 1 has been combined with the data of tire characteristics and is displayed in Fig.30. The rear axle load is ridden by the motorcyclist in a static (150kg) and has a maximum value wile turning (225kg). The data of tire characteristics in each rear axle load was used.

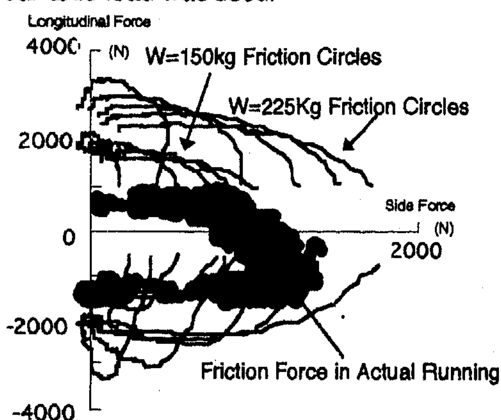


Fig.30 Friction Circle and Drag and Side Force of Specification 1

Slip ratio waveform is demonstrated in Fig.31. A worn tire was used in order to test cornering acceleration mode. The test indicated that the slip ratio was about 20%.

From Fig.29,30 and 31, one can conclude that the maximum slip ratio expected in specification1 is about 10%. The effect of slip angle at 10% slip ratio on the longitudinal friction coefficient shown in Fig.32 is small but it can be seen that the lateral friction coefficient largely varies due to the slip ratio shown in Fig.33.

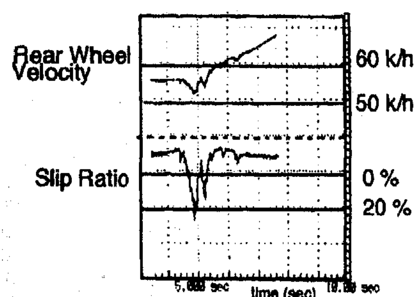


Fig.31 Slip Ratio Waveform

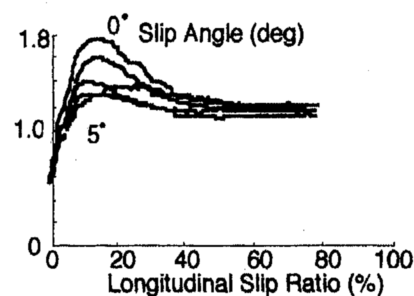


Fig.32Coeffecient of Longitudinal Friction

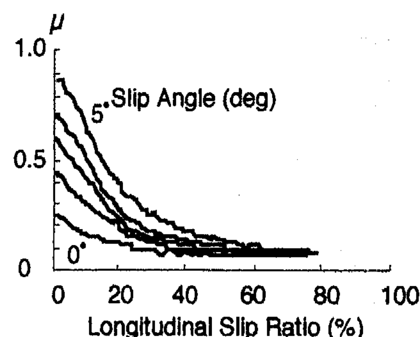


Fig.33 Coeffecient of Lateral Friction

The torque fluctuation of the crankshaft and countershaft of the transmission in specification 1 is fairly large. The motorcycle in specification 1 may be prone to slipping due to the maximum side force generated near the limit of the friction circle.

In the results mentioned above concerning the flywheel effect, the change of the moment of inertia of the crankshaft increases the ability to control the engine speed with the throttle. This verified the fact that the feeling of traction was enhanced because the rider utilized driving force and side force effectively.

CONCLUSION

The effects of the moment of inertia of the crankshaft in a motorcycle can be concluded as follows:

1. Both the gyroscopic effect of moment of inertia of the crankshaft and the main shaft influences the cornering performance of motorcycles.
2. In acceleration during a turn, a small moment of inertia of the crankshaft enhances the acceleration performance, but the engine speed control by throttle operations becomes sensitive. Riders can effectively use driving force and side force when the moment of inertia of the crankshaft is at the optimum value and the traction feeling is enhanced.
3. Good cornering performance and good driving performance (traction feeling) are produced by the optimum moment of inertia of the crankshaft on the motorcycle.

This research is a part of fundamental investigation of the gyroscopic effects on motorcycle handring dynamics. The necessity of further analysis on cornering and traction performance, as well as the need for additional research concerning the dynamics of motorcycles are expected in the future.

REFERENCES

(1)Tsuyoshi Katayama and Tomoo Nishimi, "Gyroscopic Effects on Motorcycle Dynamics", Journal of Japan Automobile Research Institute.

(2)Hirohide Imaizumi and Takehiko Fujioka, "Motorcycle Motion Analysis using Mechanical System Simulation Language", JSAESpring Convention Proceedings 942, 9433137, 1994

(3)L.SEGEL, "Tire Traction on Dry, Uncontaminated Surfaces",
University of Michigan, Ann Arbor, Michigan

971062

Comparison of Ideal Vehicle Lane-Change Trajectories

Nathaniel H. Sledge, Jr. and Kurt M. Marshek
University of Texas

Copyright 1997 Society of Automotive Engineers, Inc.

ABSTRACT

This paper seeks to compare various models of desired, or *ideal*, vehicle lane-change trajectories or paths and to determine which is best based on selected criteria. Background information is presented covering the utility of lane-change maneuvers, lane-change terminology, and known desired open-loop trajectories. The performance of a vehicle may be assessed by measuring its deviation from an ideal path during simulated or actual limit lane-change maneuvers. Therefore, several techniques for assessing vehicle performance against an ideal trajectory are demonstrated. Similarly, in the present effort, performance indices such as integral penalty (cost) functions are used for assessing candidate lane-change trajectories. Therefore, ideal trajectory comparison is essentially treated as an optimization problem with prescribed continuity and boundary conditions, where a path's length, curvature, and rate of change of curvature are taken as its costs. The maximum constant velocity or *critical speed* is employed as an additional discriminator between the candidate paths. The results of side-by-side comparison of candidate ideal trajectories are presented and discussed.

1. INTRODUCTION

Several decades ago, straight-line running ability was considered the most important vehicle handling characteristic. Today, the lane-change maneuver has become a common method of measuring vehicle performance [1]. Many automotive researchers have used lane-change maneuvers for vehicle dynamic studies. The literature indicates overwhelmingly that lane-change tests are useful for evaluating vehicle directional control and handling performance, especially during limit maneuvers – that is, maneuvers that involve high-g lateral motion. There are several applications under this heading for lane-change analysis such as: transient response [2], open-loop performance [3][4], closed-loop performance [5][6][7], vehicle control engineering [8][9][10], intelligent vehicle studies [11][12], driver assessment and subjective evaluations [1][7][13], accident reconstruction [14][15], semi-trailer truck handling [17], motorcycle handling [18], and modeling and simulation studies [8][9][10][13][19]. Additionally, Fancher et al. [19] presented a comprehensive review of attempts to standardize lane-change handling performance test procedures.

The foregoing notwithstanding, a comparison of various ideal trajectories of the lane-change maneuver apparently has not been attempted, and therefore, this paper, presents a comparison of *ideal* vehicle trajectories for high-speed lane-change maneuvers. Trajectory functions are presented in closed form when possible under the assumption of constant velocity and in terms of path attributes: (1) curvature, (2) lateral displacement, and (3) longitudinal displacement. The primary means of comparison of desired vehicle trajectories in lane-changes are (1) length, (2) curvature, (3) the derivative of curvature, and (4) critical speed.

This paper is organized as follows. Section 1 presents an introduction to the subject of lane-change maneuvers. Section 2, Background, clarifies basic lane-change terminology, lists methods for assessing vehicle performance using ideal trajectories, and describes known desired lane-change trajectories. The problem statement is given in Section 3. In Section 4, Analysis, the comparison approach is explained, which includes criteria, assumptions, and the path measures of performance used to discriminate between the candidate paths. Results and discussion of results are presented in Section 5. Section 6 presents conclusions. Appendix A summarizes the concept of critical speed and explains its relation to lane-change kinematics.

2. BACKGROUND

2.1 TERMINOLOGY – Various vehicle maneuvers may be classified as lane-change maneuvers. Fricke [15] called one-half of a lane-change maneuver (at obstacle clearance) a *critical turn-away* maneuver. Figure 2.1 illustrates what Fancher et al. [19] defined as a *single lane-change* maneuver, which is the action of applying steering in one direction to displace a vehicle laterally followed by counter steering to recover the original direction of travel, while maintaining directional control and minimizing displacement to approximately one lane width. The *double lane-change* maneuver is similar, but

includes additional steering input to recover the original direction of travel in the original lane of travel. The double lane-change maneuver is sometimes referred to as a *sinusoidal steer* maneuver. The foregoing maneuvers may be termed collectively as lane-change maneuvers. However, this paper is concerned primarily with the features of the ideal single lane-change trajectory.

A single lane-change trajectory describes the motion of a vehicle (center of gravity) along a path from one lane to another as shown in Figure 2.2 In this paper the terms curve, path, pathline, and trajectory are used interchangeably to describe vehicle position throughout a lane-change maneuver.

In Figure 2.2, $f(x)$ is a single-valued function describing the trajectory (path) of the vehicle center of gravity, x is the longitudinal distance, y is the lateral displacement, x_e is the total longitudinal distance, y_e is the total lateral displacement, and $\theta(x)$ is the tangent direction angle.

2.2 EVALUATING VEHICLE PERFORMANCE USING IDEAL TRAJECTORIES – According to Fancher et al. [19], the three basic reasons for conducting lane-change maneuver tests are: (1) the maneuver's geometry emphasizes important facets of vehicle control or performance--vehicle yaw and sideslip response and vehicle ability to "recover tight directional control", (2) the lane-change maneuver is representative of maneuvers performed frequently on highways--obstacle avoidance and high-speed passing maneuvers, and (3) it can be used to discriminate between vehicles--luxury passenger cars vs. performance and "sports" cars--because vehicle performance during lane-changes varies significantly across the spectrum of vehicle types and models.

Many authors [3][2][10][20][21] have used lane-change maneuvers to assess vehicle lateral response, mostly for vehicle design purposes. Common performance measures, gathered from various sources, for both open- and closed-loop studies are listed in Table 2.1. The comparison approach of this paper is best related to the pathline deviation and curvature performance measures–Nos. 11 and 14 of Table 2.1.

Pathline deviation may be measured by various error performance indexes that have been proposed in the literature. The integrals of EQ (2.1) give examples of generic penalty or cost functions [22], which have their origin in the *calculus of variations*.

$$J=\int_o^T e^2\,dt \qquad J=\int_o^T te^2\,dt$$
$$J=\int_o^T |e|\,dt \qquad J=\int_o^T t|e|\,dt$$
$$J=\int_o^T\left(e^2+\dot{e}^2\right)dt \qquad J=\int_o^T t\left(\left|e^2\right|+\left|\dot{e}^2\right|\right)dt \qquad (2.1)$$

where $e = e(t)$ is error or deviation as a function of time, t, with $e(t) = y_c(t) - y(t)$, $\dot{e}$ is the rate of change of the error with respect to time, $y_c(t)$ is the objective or desired path, $y(t)$is the actual or simulated path, and the upper time limit, T, is chosen sufficiently longer than the maneuver period.

Similar performance measures have also been used specifically for lane-change studies. Fancher et al. [19] report use of the following measure of performance termed *lane-change deviation*:

$$\delta=\frac{1}{3.4}\int_o^{3.4}\left|y_e-y\right|dt \qquad (2.2)$$

where the chosen period 3.4 seconds is chosen based on experimental results. Similarly, Zellner et al. [18] proposed a variation on the *integral- square-error criterion*, $\overline{y^2}$ – see EQ (2.3) – to determine the experimental "departure of a [actual] response from the desired or optimum response." To measure the time required to complete a lane-change maneuver, Zellner et al. used the *integral-time absolute-error criterion*, $\tilde{y}$ – see EQ (2.4). This criterion "weights the path error multiplicatively," such that "errors occurring late in the transient response are penalized heavily" [22].

$$\overline{y^2}=\frac{1}{T}\int_o^T\left(y_e-y\right)^2dt \qquad (2.3)$$

where $T = x_e/V_0$ and

$$\tilde{y}=\frac{1}{T}\int_o^T t\left|y_e-y\right|dt \qquad (2.4)$$

2.3 KNOWN DESIRED LANE-CHANGE TRAJECTORIES – Various functions have been used to describe ideal lane-change paths. Some notable trajectory description schemes are listed in Table 2.2.

Circular Arcs Fett [23], Fricke [15], and Daily [14] provide derivations and geometry for the circular arc representation of an ideal lane-change path. Figure 2.3 depicts the critical turn-away distance for obstacle avoidance.

Fricke [15] presents a formula for one-half of the single lane-change maneuver for the purpose of determining the critical turn-away distance, d_x. He does so using the geometry of Figure 2.4, from which EQs (2.4) and (2.5) are derived for the purpose of determining the center of gravity path radius of curvature, ρ [24].

$$\rho=\frac{\left(2d_x\right)^2}{8d_y}+\frac{d_y}{2} \qquad (2.5)$$

$$d_x=\sqrt{2d_y\rho-d_y^2} \qquad (2.6)$$

The radius can also be described as shown in EQ (2.7), in terms of the critical speed formula – EQ (A.2).

$$\rho=\frac{V^2}{\mu_m g} \qquad (2.7)$$

and the equation

$$d_x = \rho \sin\varphi \qquad (2.8)$$

where

$$\sin\varphi = \sqrt{1-\cos^2\varphi} = \sqrt{1-\left(1-\frac{d_y}{\rho}\right)^2} \qquad (2.9)$$

and φ is the angle swept by the critical turn-away circular arc. After substitutions are made, EQ (2.10) provides an estimate of the required critical turn-away distance. EQ (2.10) can be recognized as Fricke's evasive maneuver distance formula [15]. EQ (2.11) echoes Limpert's [16] expression for lane-change distance under limit conditions because it is a critical turn-away equation where only 67 percent of the maximum lateral acceleration is used.

$$d_x = V\sqrt{\frac{2d_y}{\mu_m g}} \qquad (2.10)$$

$$d_x = V\sqrt{\frac{2d_y}{0.67\mu_m g}} \qquad (2.11)$$

where μ_m is the maximum value for the tire-road coefficient of friction.

Daily [14] takes this approach one step further by connecting the critical turn-away path segment to its mirror image, doubling the distance, d_x, to $x_e = 2d_x$, thereby completing the geometry of a single lane-change maneuver as shown in Figure 2.5. In contrast to Fricke's definition of d_y as the lateral displacement to avoid an obstacle, Daily considers the lateral distance, $y_e = 2d_y$, as the total lateral distance for completion of a single lane-change. A modification of EQ (2.10) to account for the total lateral displacement yields EQ (2.12). EQ (2.12) is a common form of the total lane-change distance equation.

$$x_e = 2.82V\sqrt{\frac{y_e}{\mu_m g}} \qquad (2.12)$$

where $x_e \equiv$ the total longitudinal distance
$y_e \equiv$ the total lateral displacement

Fett [23] described the geometry for the circular arc path of a single lane-change maneuver. He noted that the radius of curvature is a function of speed and friction, but he also contended that it is a function of the prescribed lateral displacement and the lane-change duration, t_e. Assuming constant velocity, the longitudinal position at time t is $x=V_0 t$ and the lateral position is given by

$$y(x) = \rho - \sqrt{\rho^2 - x^2} \quad 0 \le x < \frac{x_e}{2} \qquad (2.13)$$

$$y(x) = y_e - \rho + \sqrt{\rho^2 - x^2 + 2xx_e - x_e^2} \quad \frac{x_e}{2} \le x \le x_e \qquad (2.14)$$

where, in accordance with the geometry of Figure 2.4, the constant radius of curvature is described by

$$\rho = \frac{x_e^2 + y_e^2}{4y_e} \qquad (2.15)$$

For the circular path, Fett relied on experimental data and assumes the duration of a "forced" lane-change to be about 3.0 seconds. A forced lane-change is effectively one in which an oncoming obstacle approaches rapidly or where headway decreases quickly. Therefore, at constant velocity, he considers the total longitudinal distance for the lane-change maneuver to be $x_e = V_0 t_e$.

There are two other ways to set up the circular trajectory. *First*, the start and end points can be prescribed, then based on the curve geometry, the maximum velocities and lateral acceleration can be determined for stable traversal of the given path. *Second*, the maximum lateral acceleration, velocity, and lateral displacement can be prescribed, leading to the estimate of the total longitudinal lane-change distance, x_e.

Fett [23] noted the following disadvantages and shortcomings for simulating lane-changes using connected circular arc paths: *First*, special measures must be undertaken to determine which circle the vehicle is on. This problem is presented by the discontinuity at the inflection point between the two circular arcs. *Second*, neither circular arcs nor circular arc tangents will accommodate changing velocity (i.e., The path will over- or under-shoot the desired lateral position). And *third*, changing velocity produces a non-smooth curve and requires a more complex vehicle and/or trajectory model.

Polynomial – Eshelman and Desai [17] and Nelson [25] both explored the use of a single Cartesian polynomial (SCP) to describe an ideal lane-change maneuver path. The polynomial can be of any power, but high order polynomials increase complexity and computation time. EQ (2.16) is the general form of a polynomial of degree-n.

$$y(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \qquad (2.16)$$

Eshelman and Desai [17] determined an ideal lane-change path for semi-trailer trucks using a seven-degree polynomial and the following conditions: $y(0) = y'(0) = y''(0) = y'''(0) = 0$, $y(x_e) = y_e$, and $y'(x_e) = y''(x_e) = y'''(x_e) = 0$. Solving for the constant coefficients, they arrived at the following degree-7 polynomial:

$$y(x) = \frac{y_e}{x_e^7}\left[35x_e^3x^4 - 84x_e^2x^5 + 70x_ex^6 - 20x^7\right] \quad (2.17)$$

Another single Cartesian polynomial for lane-changes that satisfies the constraints given in Table 2.3 is provided by Nelson [25]:

$$y(x) = y_e\left[10\left(\frac{x}{x_e}\right)^3 - 15\left(\frac{x}{x_e}\right)^4 + 6\left(\frac{x}{x_e}\right)^5\right] \quad (2.18)$$

A comparison of curvatures and trajectories for both the 5th and 7th degree SCPs is shown in Figures 2.6 and 2.7.

Ramp Sinusoid – Several authors [16][18][23] have recommended the use of a ramp sinusoidal function to describe the ideal lane-change path. Fett [23] held that a sinusoidal trajectory was preferable to connected circular arcs because it provides a continuous function of continuous curvature and it accommodates changing velocity. The geometry for ramp sinusoidal analysis is shown in Figure 2.8.

Zellner et al. [18] proposed the following function for a sine-shaped lane-change path.

$$y(t) = \frac{y_e}{x_e}V_0t - \frac{a_yx_e^2}{4\pi^2V_0^2}\sin\left(\frac{2\pi V_0t}{x_e}\right) \quad (2.19)$$

where $x = x(t) = V_0t$ is the linear displacement in the direction of travel, $y(x)$ is the lateral displacement (usually equivalent to the lane width), and lateral acceleration, a_y, takes the form:

$$a_y = \frac{y_e}{x_e^2}2\pi V_0^2 \quad (2.20)$$

Substituting x for V_0t and EQ (2.20) for a_y reduces EQ (2.19) to a more familiar form [16].

$$y(x) = y_e\left[\frac{x}{x_e} - \frac{1}{2\pi}\sin\left(\frac{2\pi x}{x_e}\right)\right] \quad (2.21)$$

For the path function, $y(x)$, the minimum radius is given by:

$$\rho_{min} \cong \frac{x_e^2}{2\pi y_e} \quad (2.22)$$

Limpert identified four levels of lateral acceleration for lane-change distance and time estimation formulas: (1) *normal* describes relatively low lateral acceleration, (2) *severe* describes moderately high lateral acceleration, (3) *limit* indicates high acceleration, and (4) *maximum* is reserved for very high lateral acceleration. Limpert considered the assignment of the maximum lateral acceleration to single direction steering situations only, like critical turn-aways, which can be likened to the first half of a single lane-change maneuver [14][15]. Limpert provided a menu of choices for modification of the lateral acceleration, a_y:

(1a) Normal (single lane-change): $a_y = (0.1 - 0.0013V)g$
(1b) Normal ("feels" safe): $a_y = [0.025V - (V/44)^{1.85}]g$
(2) Severe: $a_y = (0.22 - 0.002V)g$
(3) Limit: $a_y = 0.67a_{y,\max}$
(4) Maximum (critical turn-away): $a_y = 0.85a_{y,\max}$

EQ (2.23) is in the form typically used to represent the total longitudinal distance for a lane-change.

$$x_e = C_xV\sqrt{\frac{y_e}{a_y}} \quad (2.23)$$

In EQ (2.23), the lateral acceleration, a_y, is equivalent to $\mu_m g$, where μ_m is the maximum coefficient of friction between the tire and road. Limpert [16] cites experimental results that yield a distribution of values for the constant coefficient, C_x, with modes at 2.6 and 2.67. Based solely on algebraic manipulation, C_x is approximately equal to 2.51.

Spirals – Kanayama and Miyake [26] proposed a novel lane-change trajectory while researching smooth local path planning for autonomous vehicles. They maintain that Eulerian, or Cornu, spirals are preferable to straight lines and circular arcs because of continuous and smooth paths and curvature. As shown in Figure 2.9, we construct a lane-change path by specifying two *endpostures* where the curvature is null. The endpostures are a combination of position and orientation information; for example: $(p_0, p_e) \rightarrow p_0=(x_0, y_0, \theta_0)$ and $p_e=(x_e, y_e, \theta_e)$, where position coordinates $(x_0, y_0) \neq (x_e, y_e)$ and θ_0 is the tangent direction of the path at each posture point. The piecewise continuous *curvature* for the Cornu spiral path varies linearly. Taking this feature into account, lane-change path curvature has the following piecewise structure:

$$\begin{aligned} \kappa(s) &= c(s,w)\, s + \kappa_0 && 0 \le s < l/4 \\ &= -c(s,w)\, s - 2\,\kappa_1(l/4) + \kappa_0 && l/4 \le s < 3\,l/4 \\ &= c(s,w)\, s - 4\,\kappa_2(3\,l/4) + \kappa_0 && 3\,l/4 \le s \le l \end{aligned} \quad (2.24)$$

Where l is the total *positive length* of the curve, κ_i are curvature constants, and $c(s,w) = 2/w(\beta)^2$ is the *sharpness* of the curve, and $w = w(\beta)$ is the *roundness* factor. β is the orientation angle between the first and final points and is given by

$$\beta = \tan^{-1}\left[\frac{y_e - y_0}{x_e - x_0}\right] \quad (2.25)$$

The key parametric variables describing the lane-change path are defined as follows:

$$\theta(s)=\int_0^s \kappa(\zeta)d\zeta \quad 0\le s\le l \tag{2.26}$$

$$x(s)=\int_0^s \cos\theta(\zeta)d\zeta \quad 0\le s\le l \tag{2.27}$$

$$y(s)=\int_0^s \sin\theta(\zeta)d\zeta \quad 0\le s\le l \tag{2.28}$$

where ζ is a "dummy" integration variable.

The endpostures are said to be symmetric if $\theta_0 - \beta = -(\theta_e - \beta)$; and this is the condition for parallel endpostures such as those for a single lane-change. The size d and deflection α of a symmetric curve (velocity is constant) are defined respectively as follows:

$$d=\sqrt{(x_e-x_0)^2+(y_e-y_0)^2} \tag{2.29}$$

$$\alpha=\theta_0-\theta_e \tag{2.30}$$

where θ is the curve tangent direction angle. For example, the deflection for a left turn is $\alpha = \pi/2$, and the deflection for a lane-change is $\alpha = 0$.

For a description of lane-change paths of constant velocity vehicles, the parallel case ($\theta_0 = \theta_e$, so $\alpha = 0$) of the Cornu spiral is used. The path inflection point is given by the intermediate posture that Kanayama and Hartman [27] refer to as the *least cost split point, q*.

$$q=\left(\frac{x_0+x_e}{2},\frac{y_0+y_e}{2},\beta-(\theta-\beta)\right) \tag{2.31}$$

The split point used in conjunction with each endposture yields two *clothoid* pairs that describe the ideal lane-change path. The minimum radius of curvature occurs at the turning points B and D as depicted in Figure 2.9 and 2.10.

Acceleration Profile – Prescribing a lateral acceleration profile enables analysis of the effect of lateral acceleration on a lane-change trajectory. Figure 2.11 compares three examples of acceleration profiles where $a_{y,\max} = \mu_m g$. The first is *sinusoidal* with a maximum amplitude equal to $a_{y,max}$ [12], the second is *trapezoidal* with maximum amplitude equal to 0.85 $a_{y,\max}$ [11], and the third, the crudest, is represented by a *square wave* with a maximum amplitude of 0.67 $a_{y,\max}$ [16].

Chovan et al. [12] prescribed a sinusoidal lateral acceleration (a_y) profile as a function of time:

$$a_y=a_{y,m}\sin(\omega t)=\left(\frac{2\pi y_e}{t_e^2}\right)\sin\left(\frac{2\pi t}{t_e}\right) \tag{2.32}$$

where $a_{y,m} \equiv$ maximum lateral acceleration = $0.7g$
$\omega \equiv$ lane-change angular frequency = $2\pi/t_e$
$y_e \equiv$ lane-change total lateral displacement
$t \equiv$ elapsed time into the lane-change
$t_e \equiv$ total time to complete the lane-change

The expression for lateral acceleration, a_y, can be integrated to determine lateral velocity and position:

$$V_y=v=\left(\frac{y_e}{t_e}\right)\left[1-\cos\left(\frac{2\pi t}{t_e}\right)\right]+v_0 \tag{2.33}$$

$$y(t)=\left(\frac{y_e t}{t_e}\right)-\left[\frac{y_e}{2\pi}\sin\left(\frac{2\pi t}{t_e}\right)\right]+v_0 t+y_0 \tag{2.34}$$

Chee et al. [11], in their attempt to control vehicle motion in an automated highway system (AHS), introduced the term virtual desired trajectory (VDT) for ideal lane-change curves. They acknowledged the use of lane-change trajectories based on reverse circular arcs and polynomial functions. They also introduced a trapezoidal lateral acceleration profile to prescribe a VDT. Their analysis focused on the maximum lateral acceleration capability of AHS vehicles during lane-change maneuvers.

Taking into account issues and constraints such as passenger comfort and vehicle capabilities, Chee et al. [11] first prescribe a trapezoidal acceleration profile and then observed its effect on the trajectory. EQ (2.35) represents the acceleration profile by a piecewise function of time, where $J_{\max}$ is the maximum jerk and the lateral acceleration is limited to 85 percent of the nominal maximum corresponding to Limpert's *maximum* case. The acceleration profile must be integrated to yield the velocity and position graphs depicted in Figure 2.13.

$$\begin{aligned} a_y &= J_{\max} t && 0 \le t < t_1 \\ a_y &= 0.85\, a_{y,m} && t_1 \le t < t_2 \\ a_y &= -J_{\max} t + 2\,(.85)\, a_{y,m} && t_2 < t \le t_3 \\ a_y &= -0.85\, a_{y,m} && t_3 < t \le t_4 \\ a_y &= J_{\max} t - 4\,(.85)\, a_{y,m} && t_4 < t \le t_e \end{aligned} \tag{2.35}$$

Limpert [16] used a square wave acceleration profile with an amplitude equal to 67 percent of the nominal maximum as an average lateral acceleration. This acceleration profile is represented by one period of a square wave as shown in Figure 2.14.

Bezier-Spline – In the Bezier spline approach, the spline algorithm interpolates only the prescribed endpoints; the other points are used to control the shape of the spline curve [28]. According to Apetaur and Opicka [8], a simple form of the Bezier Spline equation is:

$$B(p)=\sum_{k=0}^{n} p^k(1-p)^{n-k} S_{k+1} \tag{2.36}$$

In Figure 2.15, the spline trajectory is constructed using lane-change dimensions y_e=3.66 m and x_e = 62.14 m. These dimensions are inputs to the following knot algorithm:

$$x = x_e \frac{n}{n-1} \quad (2.37)$$

$$y = 0,\ 0 <= x <= \frac{x_e}{3};\ y = x\left(\frac{3y_e}{x_e}\right) - y_e,$$
$$\frac{x_e}{3} < x < \frac{2x_e}{3};\ y = y_e,\ \frac{2x_e}{3} <= x <= x_e \quad (2.38)$$

where n = 7 is the number of knots (points) and n-1 is number of equal segments along the x-axis. Again, the vehicle speed is assumed to be constant; therefore, the knot locations are symmetrical about the path center ($x_e/2$, $y_e/2$). In Figure 2.15, the Bezier spline curve is displayed along with the *cubic spline* curve using the same knots. For the given number of knots, the Bezier spline path is more desirable.

Other Lane-Change Trajectories – Several other trajectories, not presented in detail here, are discussed in the literature. Nelson [25] proposed replacing the circular arc with polar polynomials with continuous curvature. Guldner et al. [29] used *curvature profiles* to design reference paths for assessing automatic steering controllers. Kanayama and Hartman [27] offered the *cubic spiral* as an improvement over the Cornu spiral. Martinez-Alfaro and Flugrad [30] synthesized *B-Spline* curves to obtain optimal collision-free paths for automated vehicles. Finally, Chan [31] formulated trajectory design as an optimization problem using weighted penalty functions for acceleration and jerk.

Table 2.3 summarizes and provides representative mathematical descriptions of the trajectories presented in this subsection.

3. PROBLEM STATEMENT

The primary objective of this paper is to compare several lane-change trajectory functions and determine which is superior, based on selected criteria and assumptions. The lane-change trajectories to be compared are: (a) reverse circular arcs, (b) 5th degree Cartesian polynomial, (c) 7th degree Cartesian polynomial, (d) ramp sinusoid, (e) Bezier Spline, and (f) Cornu spiral.

3.1 CRITERIA – Table 3.1 is compiled from various sources [8][23][25][27][32] and provides a list of commonly desired criteria for ideal lane-change paths. In general, the criteria prescribe minimal length paths for smooth lane traversal with minimal yaw transients and jerk.

4. ANALYSIS

4.1 COMPARISON APPROACH – There are various ways to establish a common reference for comparing the candidate ideal lane-change trajectories. *First*, the maximum lateral acceleration, velocity and lateral displacement may be prescribed, allowing an estimate of the total longitudinal lane-change distance, x_e, and the path length, l – see Appendix A. *Second*, for geometrical consistency, the overall dimensions (x_e and y_e) may be prescribed, then, based on the resulting path geometry, the maximum velocities and maximum lateral acceleration can be determined for stable traversal of the given path – see Appendix A.* The second method will be used in this paper to compare the candidate paths, where x_e and y_e are prescribed and curvature and critical speed are determined.

Two important features of lane-change trajectory functions are curvature, $\kappa(x)$ and radius of curvature, $\rho(x)$. EQ (4.1) defines curvature in *explicit* form using Cartesian coordinates.

$$\kappa(x) = \frac{y''(x)}{\left[1 + y'(x)^2\right]^{3/2}} \quad (4.1)$$

EQ (4.2) also defines curvature in Cartesian coordinates but in *parametric* form, where $x = x(t)$ and $y = y(t)$ [33].

$$\kappa(t) = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{\left(\dot{x}^2 + \dot{y}^2\right)^{3/2}} \quad (4.2)$$

EQ (4.3) shows the reciprocal constitutive relationship between curvature and the radius of curvature.

$$\rho(x) = \frac{1}{|\kappa(x)|} \quad (4.3)$$

We seek the path with minimal curvature and relatively large minimum radius to minimize vehicle yaw transients, lateral acceleration, and jerk. Since there is a reciprocal relationship between $\kappa(x)$ and $\rho(x)$, the determination of the curvature function and its maximum value will suffice for finding the path function with minimal curvature and the largest minimum radius. Table 4.1 gives the functions for curvature of the candidate paths and Table 4.2 gives relations for the minimum radius.

4.2 IDEAL TRAJECTORY MEASURES OF PERFORMANCE – Table 4.3 lists the path measures of performance chosen to satisfy the criteria introduced in Section 3.

As a starting point, the path and curvature boundary conditions are to be verified for compliance with those listed in Table 3.1. Next, the continuity and differentiability of each path and its curvature will be checked within the interval [0, x_e] or [0, t_e], as applicable.

Relative minimal length is assessed through three measures of performance. Figure 4.1 illustrates these

* The references to Appendix A emphasize the connection between the maximum lateral acceleration and the critical speed formula.

measures of performance. The *first*, the maximum deviation from a directed straight line, Δ_{max}, is found using

$$\Delta_{max} = \text{Max}[line(x) - y_c(x)] \cos\beta$$

or

$$\Delta_{max} = \text{Max}[line(t) - C(x(t),y(t))] \cos\beta \qquad (4.4)$$

where $line(x) = (y_e / x_e)x$, and $\beta = atan(y_e / x_e)$. The *second* measure of performance is the area, A, between the path and a straight line for one half of a single lane-change maneuver, which is given by

$$A = \int_0^{xe/2} [line(x) - y_c(x)] dx$$

or

$$A = \int_0^{te/2} [line(t) - C(x(t), y(t))] dt \qquad (4.5)$$

where y_c and C are single-valued and parametric path functions, respectively. Finally, the *third* measure of performance is the total path length, l. Where path length is given by

$$l = \int_0^{xe} \sqrt{1 + y'(x)^2}\, dx$$

or

$$l = \int_0^{te} \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}\, dt \qquad (4.6)$$

A normalized form of this metric is the *length ratio* $= l / x_e$.

While maximum curvature and minimum radius of curvature are computed from the relationships in Tables 4.1 and 4.2, respectively, relative minimal curvature and relative minimal jerk are evaluated by computing the integral of the curvature squared and the derivative of curvature squared, respectively. These cost functions – EQs (4.7) and (4.8) – were employed by Kanayama and Hartman [27] to synthesize smooth paths for autonomous vehicles, but the reasoning behind such usage is supported by the principles of *variational calculus*.

$$Kcost = \int_0^l \kappa(s)^2\, ds$$

or

$$Kcost = \int_0^{xe} \kappa(x)^2 \sqrt{1 + y'(x)^2}\, dx \qquad (4.7)$$

$$Jcost = \int_0^l \dot{\kappa}(s)^2\, ds$$

or

$$Jcost = \int_0^{xe} \kappa'(x)^2 \sqrt{1 + y'(x)^2}\, dx \qquad (4.8)$$

From a dynamical standpoint, "the instantaneous centripetal acceleration of the vehicle is proportional to its curvature." Therefore, *Kcost* takes lateral acceleration as the cost and *Jcost* takes the rate of change of lateral acceleration, or *jerk*, as path cost [27].

4.3 ASSUMPTIONS – The following assumed parameter values apply to the trajectories in this study:

a. Tire-road coefficient of friction, $\mu_m = 0.7$
b. Gravitational acceleration, $g = 9.81$ m/s^2
c. Total lateral displacement, $y_e = 3.66$ m (12 ft)
d. Total longitudinal distance, $x_e = 62.14$ m (204 ft)

Parameters μ_m and y_e are representative of conditions for a highway lane-change. The value for x_e is typical for a lane-change conducted at approximately 26.8 m/s (96.5 km/h or 60 mph) and $a_{y\,max} = 0.47g$ — see [16].

5. RESULTS

5.1 STATEMENT OF RESULTS – Tables 5.1 and 5.2 give the results of application of the measures of performance.

5.2 DISCUSSION OF RESULTS – Based on numerical results alone, the reverse circular arc path appears to be superior; but close examination reveals that it is the least desirable of all the trajectory functions examined. A sudden change from a straight run to a circular path would entail an instantaneous change in the lateral acceleration, from zero to V^2/ρ, with an infinite rate of change – *jerk* = ∞. Moreover, at the inflection point between the two circular arcs, the lateral acceleration abruptly changes polarity. Therefore, circular arcs do not provide a realistic depiction of lane-change motion. From an analytical standpoint, the circular function does not meet the criteria $\kappa(0) = \kappa(x_e) = 0$ and $\kappa'(0) = \kappa'(x_e) = 0$, because the curvature is constant throughout the lane-change. In effect, the curvature is "evenly distributed" over the entire path. This implies that a vehicle traversing the circular arc path experiences constant lateral acceleration – i.e., *jerk = 0*. The abrupt lateral acceleration sign change and the fact that the circular arc function does not meet the initial and final curvature constraints invalidate the favorable numerical results. Therefore, the circular arc function has little utility as an ideal lane-change path.

Results indicate that the 5th-degree polynomial and the Bezier spline paths meet all criteria and are equally superior to the other lane-change path functions. The next two path functions–ramp sinusoid and 7th-degree polynomial–also meet all criteria. The ramp sinusoid performed nearly as well as the 5th-degree polynomial and the Bezier spline. The 7th-degree polynomial is not as suitable because it has high maximum curvature. The Cornu spiral produced the least favorable results.

Overall, the differences between the candidate paths appear to be minor, and each candidate path function provides a

reasonable kinematic description of a lane-change maneuver. However, when an additional metric — implied critical speed, V_{crit} — is applied, a clearer view of the relative "goodness" of the candidate ideal trajectories is achieved. Table 5.2 shows, based on the constant velocity and the minimum radius of curvature, that the critical speed is a clear discriminator for the candidate paths. Put simply, the superior ideal path description allows higher traversal speed for the same geometric boundary conditions and tire-road friction.

On a qualitative basis, the 5th-degree polynomial, the ramp sinusoid, and the 7th-degree polynomial paths are the easiest to implement without automation. Conversely, the Bezier spline and Cornu Spiral are relatively complex functions that require at least a modest level of programming.

Given the numerical and qualitative results, the 5th-degree polynomial is the preferred functional description of an ideal lane-change among the candidates examined. In rank order, and based on both qualitative and numerical results, the other preferred lane-change functions are the ramp sinusoid, Bezier spline, 7th-degree polynomial, and the Cornu spiral.

6. CONCLUSIONS

The primary contribution of this paper has been the comparison and assessment of known lane-change paths. The method of approach employed to compare six candidate ideal trajectories was guided by reasonable criteria and assumptions. The criteria and assumptions are reasonable because they draw on previous lane-change studies and are based on realistic conditions. The choice of fixing y_e, x_e, and μ_m provided a reference point for kinematic and dynamical consistency during comparison of the candidate trajectories. On a kinematic basis, this enabled the determination of performance on the basis of path length and curvature. From a dynamical standpoint, the approach allowed for comparison on the basis of the critical speed, given equivalent geometric boundary conditions. Therefore, this study has contributed to the understanding of the connection between critical speed and lane-change path design – Appendix A – and showed that this understanding is necessary for realistic and credible trajectory analysis.

The trajectories presented here spanned a wide spectrum of complexity, but there was no correlation between complexity and path efficiency. Of the two best ideal lane-change trajectory functions, the 5th-degree polynomial is the simpler. All ideal lane-change trajectories complied with commonly accepted path and curvature criteria except the reverse circular arc function. Among those paths meeting all criteria, there are nonnegligible numerical and qualitative differences – curvature, length, critical speed, continuity, differentiability, and ease of use.

NOMENCLATURE*

A	one-half area between a directed straight-line and an ideal path
a_x	vehicle longitudinal acceleration/deceleration
a_y	vehicle lateral acceleration (V^2/ρ)
$a_{y,\,max}$	vehicle maximum lateral acceleration
c	path sharpness parameter
$C(x,y)$	ideal or desired trajectory function – parametric
d_x	longitudinal distance for critical turn-away
d_y	lateral displacement for critical turn-away
e	superelevation (road surface, axle direction, $tan\theta$)
g	gravitational constant
J	jerk (da_y/dt)
$Jcost$	derivative of curvature squared cost function
k	counter
$Kcost$	curvature squared cost function
l	total path length
m	total vehicle mass
n	number of knots in the Bezier spline
r	vehicle yaw rate
s	path length parameter
t	time
t_e	total time to complete a lane-change
T	lane-change period
V	vehicle forward velocity (tangent to path of curvature)
V_{crit}	critical speed (at the point of incipient side-skid)
w	path roundness parameter
x	longitudinal position (abscissa)
x_e	total longitudinal lane-change distance
y	lateral position (ordinate)
$y_c(x)$	ideal or desired trajectory function – single-valued
y_e	total lateral lane-change displacement
α	endposture-to-endposture deflection angle
β	endposture-to-endposture orientation angle; vehicle side-slip angle
δ, Δ	pathline deviation
Δ_{max}	maximum pathline deviation
κ	path curvature
μ_m	maximum coefficient of friction
ψ	vehicle yaw angle
$\theta(x)$	tangent angle; slope angle
ρ	radius of curvature of center of gravity path
ρ_{min}	minimum radius of curvature
ω	lane-change angular frequency
ζ	dummy integration variable

* All velocities and accelerations are measured relative to the vehicle local coordinate system.

REFERENCES

1. Matsushita, A., et al., 1980, "Subjective Evaluation and Vehicle Behavior in Lane-Change Maneuvers," *SAE Transactions*, SAE 800845.
2. Heydinger, G. J., 1990, Improved Simulation and Validation of Road Vehicle Handling Dynamics, Ph.D. Dissertation, Ohio State University.
3. Dugoff, H., et al., 1971, "Measurement of Vehicle Response in Severe Braking and Steering Maneuvers," *SAE Transactions*, SAE 710080, Warrendale, PA, pp. 281-306.
4. Ervin, Robert D., et al., 1973, "Refinement and Application of Open-Loop Limit-Maneuver Response Methods," *SAE Transactions*, SAE 730491.
5. Hayes, G. G., et al., 1975, Determination of Motor Vehicle Characteristics Affecting Driver Handling Performance, U. S. DOT, National Highway Traffic Safety Administration, DOT-HS-801-421, NTIS.
6. McRuer, D. T., and Klein, R. H., 1975, Automobile Controllability–Driver/Vehicle Response for Steering Control, Vol. I: Summary Report, DOT-HS-801-406, Vol. II: Supporting Experimental Results, DOT-HS-801-407.
7. Rice, R. S., and Dell'Amico, F., 1974, An Experimental Study of Automobile Driver Characteristics and Capabilities, Calspan Report No. ZS-5208-K-1, Buffalo, NY, March 1974.
8. Apetaur, Milan, and Opicka, Frantisek, 1991, "Assessment of the Driver's Effort in Typical Driving Manoeuvers for Different Vehicle Configurations and Managements," The Dynamics of Vehicles on Roads and Tracks, ed. Gilles Savage, Proceedings of 12th IAVSD Symposium, Lyon, France, 26-30 Aug, 1991, Supplement to Vehicle Systems Dynamics, Vol. 20., pp. 43-56.
9. Harada, Hiroshi, and Iwasaki, Takashi, 1993, "Stability Criteria and Objective Evaluation of a Driver-Vehicle System for Driving in Lane Change and Against Crosswind," *The Dynamics of Vehicles on Roads and on Tracks*, Supplement to Vehicle System Dynamics, Vol. 23, ed. Shen, Zhiyun, Proceedings of the 13th IAVSD Symposium, Chengdu, Sichuan, P. R. China, August, 23-27, 1993, pp. 197-209.
10. Smith, D. E., Starkey, J. M., and Benton, R. E., 1995, "Nonlinear-Gain-Optimized Controller Development and Evaluation for Automated Emergency Vehicle Steering," *Proceedings of the American Control Conference*, Seattle, WA, June 1995, pp. 3586-3590.
11. Chee, W. et al 1995, "Experimental Study of Lane-change Maneuver for AHS Applications," *Proceedings of the American Control Conference*, Seattle, WA, June 1995, pp. 139-143.
12. Chovan, John D., et al., 1994, Examination of Lane Change Crashes and Potential IVHS Countermeasures, U. S. DOT, Research and Special Programs Administration, National Transportation Systems Center, DOT-HS-808-071, NTIS No. PB-94-031466.
13. Habib, Mostafa S., 1994, "Robust Stability of Pilot-Vehicle Directional Control," *Proceedings of the Third IEEE Conference on Control Applications*, 24-26 Aug 1994, The University of Strathclyde, Glasgow, Scotland, UK, Vol. 1/3, pp. 141-145.
14. Daily, John, 1988, Fundamentals of Traffic Accident Reconstruction, Chapter 6, "Derivation of the Lane-change Maneuver Formula," IPTM, University of North Florida.
15. Fricke, Lynn B., 1990, Traffic Accident Reconstruction, Vol. 2 of The Traffic Accident Investigation Manual, Northwestern University Traffic Institute, Evanston, IL 60204.
16. Limpert, Rudolf, 1994, Motor Vehicle Accident Reconstruction and Cause Analysis, 4th Edition, The Michie Company, Charlottesville, VA.
17. Eshelman, R. L., and Desai, S. D., 1972, Articulated Vehicle Handling, U. S. DOT, National Highway Traffic Safety Administration, DOT-HS-800-674, NTIS No. PB-211201.
18. Zellner, John W., and Weir, David, H., et al., 1978, "Development of Handling Test Procedures for Motorcycles," Motorcycles Dynamics and Rider Control, SP-428, SAE, Warrendale, PA, SAE 780313, pp. 91-100.
19. Fancher, P., et al., 1976, "Test Procedures for Studying Vehicle Dynamics in Lane-Change Maneuvers," Automotive Engineering Congress and Exposition, Detroit, MI, 23-27 Feb, 1976, SAE, Inc., SAE 760351.
20. Allen, R. Wade., and Rosenthal, Theodore J., 1993, "Measurement of Vehicle Response in Severe Braking and Steering Maneuvers," Vehicle Dynamics and Simulation, SP-950, SAE, Warrendale, PA, SAE 710080, pp. 281-306.
21. Metz, L. D., and Alter, D. M., 1991, "Transient and Steady State Performance Characteristics of a Two-Wheel-Steer and Four-Wheel-Steer Vehicle Model," *SAE Transactions*, Warrendale, PA, SAE 911926.
22. Ogata, Katsuhiko, 1970, Modern Control Engineering, Ed. Everitt, William L., Prentice-Hall, Inc., Englewood Cliffs, N. J.
23. Fett, I. H. C., 1974, Simulation of Lane Change Maneuvers on Intersection Approaches, Masters Thesis, University of Texas at Austin, Advisors Lee, Clyde and McCullough, Frank.
24. Sledge, N. H., and Marshek, K. M., 1996, Vehicle Critical Speed Formula - Determining the Path Radius of Curvature - A Review, Mechanical Engineering Technical Report No. 302, The University of Texas at Austin.

25. Nelson, W. 1989, "Continuous-Curvature Path for Autonomous Vehicles," *IEEE International Conference on Robotics and Automation*, May 1989, pp. 1260-1264.

26. Kanayama, Y. and Miyake, N. 1985, "Trajectory Generation for Mobile Robots," *Robotics Research*, Third International Symposium on Robotics Research, Gouvieux, France, MIT Press, eds. Winston, P. H., and Brady, M., pp. 333-340.

27. Kanayama, Y. and Hartman, B I., 1989, "Smooth Local Path Planning for Autonomous Vehicles," *IEEE International Conference on Robotics and Automation*, May 1989, pp. 1265-1270.

28. de Boor, C., 1978, A Practical Guide to Splines, Springer-Verlag, New York.

29. Guldner, Jurgen, et al., 1994, "A Sliding Mode Control Approach to Automatic Car Steering," *Proceedings of the American Control Conference*, Seattle, WA, June 1994, pp. 1969-1973.

30. Martinez-Alfaro, H., and Flugrad, D. R., 1994, "Optimal Collision-Free Path Planning for Robots Using B-Splines and Simulated Annealing," Robotics: Kinematics, Dynamics, and Controls, ASME, DE-Vol. 72, 23rd Biennial Mechanisms Conference, Minneapolis, MN, 11-14 Sep 1994, pp. 323-331.

31. Chan, Ching-Yao, 1995, "Open-Loop Trajectory Design for Longitudinal Vehicle Maneuvers: Case Studies with Design Constraints," *Proceedings of the American Control Conference*, Seattle, WA, June 1995, pp. 4091-4095.

32. Koepele, B. and Starkey, J. M., 1990, "Closed-loop Vehicle and Driver Models for High-speed Trajectory Following," *Transportation Systems*, ASME Winter Annual Meeting, 25-30 Nov 1990, AMD Vol. 108, pp. 59-68.

33. Goodman, A. W., 1967, Modern Calculus with Analytic Geometry, Volume I, The MacMillan Company.

34. Baker, J. Stannard, and Fricke, Lynn B., 1986, Traffic-Accident Investigation Manual, Northwestern University Traffic Institute, Evanston, IL, Vol. 1.

35. Wong, J. Y., 1993, Theory of Ground Vehicles, John Wiley & Sons, Inc., New York, p. 281.

36. Lee, Allen Y., 1995, "Performance of Four-Wheel-Steering Vehicles in Lane Change Maneuvers," New Developments in Vehicle Dynamics, Simulation, and Suspension Systems, SAE, Warrendale, PA, SP-1074, SAE 950316, Feb 1995, pp. 161-173.

37. Smith, D. E., and Starkey, J. M., 1991, "Overview of Vehicle Models, Dynamics, and Control Applied to Automated Vehicles," *Advanced Automotive Technologies-1991*, DE-Vol. 40, ASME Winter Annual Meeting, Atlanta, GA, 1-6 Dec 1991, pp. 69-87.

38. Sledge, N. H., and Marshek, K. M., 1996, Formulas for Estimating Vehicle Critical Speed From Yaw Marks - A Review, Mechanical Engineering Technical Report No. 304, The University of Texas at Austin.

APPENDIX A

LATERAL ADHESION, CRITICAL SPEED, AND LANE-CHANGE MANEUVERS

The study of lane-change trajectories includes a consideration of lateral resistance and its limits. During lane-change maneuvers, lateral resistance is required to counteract the inertial force inherent during vehicle yaw motion. Lateral resistance is provided by tire side friction forces and banking of the roadway. Maximum lateral resistance occurs when a vehicle achieves its *critical speed.* The critical speed is the vehicle speed at which the tires reach the limit of side friction and side-slip begins [34]. In other words, critical speed is reached when the vehicle centrifugal force due to vehicle motion in a circular path equals the opposing centripetal force provided by friction and superelevation at the tire-road interface. The critical speed formulas are used to estimate the vehicle speed at *incipient side-slip*. Incipient side-slip describes small, but non-negligible, side-slip at the tire-road interface. Side-slip occurs when a vehicle exceeds its critical speed and slips laterally for an appreciable distance.

Oversteer vehicle motion beyond the critical speed is characterized initially by side-slip; but motion may become dynamically unstable at higher speeds, ultimately resulting in spinout. The condition of dynamic instability occurs when a vehicle is directionally unstable. For example, a vehicle that has exceeded the critical speed may continue to diverge from its original path, depending on the nature of the vehicle's stability [35].

Yaw motion describes a vehicle's rotation about its vertical axis (*z*-axis) [35]. Yaw motion occurs anytime a vehicle changes its heading in the fixed *x-y* plane, as in a lane-change maneuver. Circular motion is an example of yaw motion. Figure A.1 illustrates the oversteer yaw motion of a vehicle that has exceeded its critical cornering speed.

In order for a vehicle to traverse a curved path, the vehicle must not exceed the maximum lateral acceleration determined by the maximum tire-road adhesion. In the case of limit lane-change maneuvers, vehicle lateral acceleration is a function of tire-road lateral resistance and usually exceeds $0.3g$ [10][32][36][37]. If certain assumptions are made [38], then as shown in EQ (A.1), the maximum lateral acceleration, $a_{y,max}$, is proportional to the forward velocity squared and inversely proportional to the path radius of curvature. The maximum lateral acceleration is also limited by the lateral resistance present at the tire-road interface, represented in EQ (A.1) as the maximum value for the coefficient of friction, μ_m.

$$a_{y,\max} = \mu_m g = \frac{V^2}{\rho} \qquad \text{(A.1)}$$

where $\mu_m \equiv$ maximum coefficient of friction

g ≡ gravitational acceleration
V ≡ vehicle forward velocity
ρ ≡ center of gravity path radius of curvature

EQ (A.2) represents the basic form of the critical speed equation for flat road surfaces [15].

$$V_{crit} = \sqrt{g\rho\mu_m} \qquad \text{(A.2)}$$

where μ_m ≡ maximum coefficient of friction
g ≡ gravitational acceleration
ρ ≡ center of gravity path radius of curvature

Another common form of the basic critical speed formula, which accounts for superelevation between ±10 percent (nearly ±6 degrees), is shown in EQ (A.3) [14][15].

$$V_{crit} = \sqrt{g\rho(\mu_m \pm e)} \qquad \text{(A.3)}$$

where e is the superelevation, θ is the superelevation angle, and $e = tan\theta$.

There are other formulas of varied complexity and accuracy used in the determination of critical speed – see [38]. The critical speed equation used for lane-change trajectory analysis depends on the availability of computer resources, the complexity of the study, and the desired fidelity.

When the limits of vehicle performance are sought during lane-change simulations, the critical speed formula chosen theoretically influences the calculated maximum lateral acceleration and hence, the shape of the trajectory. The critical speed equation(s) can be used for the determination of the ideal lane-change trajectories, particularly in calculating the total longitudinal distance, x_e, required to complete the lane-change maneuver, when the constant velocity, V_0, and coefficient of friction, μ_m, are given.

Either EQ (A.1) or (A.2) can be rearranged to be explicit in minimum radius of curvature as follows:

$$\rho_{min} = \frac{V_0^2}{\mu_m g} \qquad \text{(A.4)}$$

Equating EQ (A.4) with any expression for the minimum radius of curvature listed in Table 4.2 yields an expression for the total longitudinal distance, x_e. If, for example, the ramp sinusoid trajectory is selected, we arrive first at the following equation:

$$\frac{x_e^2}{2\pi y_e} = \frac{V_0^2}{\mu_m g} \qquad \text{(A.5)}$$

Then, by rearrangement, the total longitudinal distance is determined as

$$x_e = \sqrt{2\pi} V_o \sqrt{\frac{y_e}{\mu_m g}} \qquad \text{(A.6)}$$

Using the same procedure, a similar expression for circular arcs is determined as

$$x_e = \sqrt{\frac{4y_e V_0^2}{\mu_m g} - y_e^2} \qquad \text{(A.7)}$$

For the more complicated expressions for minimum radius of curvature, iterative techniques are required to solve for x_e.

Rearranging EQ (A.1), the maximum curvature of a lane-change trajectory is limited by the following:

$$\kappa_{max} = \frac{\mu_m g}{V_0^2} = \frac{1}{\rho_{min}} \qquad \text{(A.8)}$$

Therefore, if x_e is specified along with y_e, and μ_m, V_0 may be determined and vice versa.

APPENDIX B

TABLES

Table 2.1 Common Lane-change Performance Measures
[1][18][19]

No.	Description
1	Duration to achieve lateral displacement
2	Duration for passing course completion
3	Maintenance of directional control
4	Steering effort versus control
5	Course fidelity (lane violations)
6	Symmetry of yaw and side-slip angle responses
7	Fourier spectrum of steering input
8	Maximum normalized steering amplitude
9	Minimum longitudinal distance to achieve specified lateral displacement
10	Velocity versus lane-change distance ratio (V/x_e)
11	Pathline deviation, Δ
12	Maximum side-slip angle, β_m, and rate, $d\beta/dt$
13	Maximum yaw rate, r_m, where $r = d\psi/dt$
14	Maximum path curvature, κ_m
15	Over- and under-correct tendency, $Y_{u/o}$
16	Left-to-right symmetry, $Y_{L/R}$
17	Time and lateral position at start of recovery phase (time lag of vehicle response)
18	Crossover frequency
19	Phase margin
20	Response latency
21	Closed-loop damped natural frequency
22	Vehicle steering gain
23	Yaw time constant, T_r
24	Directional mode frequency
25	Damping ratio
26	Observer subjective ratings
27	Driver subjective ratings

Table 2.2 Trajectory Description Schemes

Name/Description	Author/Reference
Reverse Circular Arcs	Fett [23], Daily [14], Fricke [15]
Single Cartesian Polynomial	Eshelman and Desai [17], Nelson [25]
Polar Polynomial	Nelson [25]
Sinusoidal Ramp	Zellner et al. [18], Limpert [16]
Cornu Spiral	Kanayama and Miyake [26]
Cubic Spiral	Kanayama and Hartman [27]
Acceleration Profile	Guldner et al. [29], Chee et al. [11], Chan [31]
Bezier Spline	Apetaur and Opicka [8], Nelson [25], de Boor [28]
Curvature Profile	Apetaur and Opicka [8], Guldner et al. [29]
B-Spline	Martinez-Alfaro and Flugrad [30]

Table 3.1 Criteria for Ideal Lane-Change Paths

Criteria	Description of Criteria
1	Continuous and smooth curvature
2	Minimal length
3	Minimal curvature
4	Minimal jerk
5	Curvature boundary conditions: $\kappa(0) = \kappa(x_e) = 0$ and $\frac{d\kappa}{dx}(0) = \frac{d\kappa}{dx}(x_e) = 0$
6	Single-valued function boundary conditions: $y(0) = 0,\ y(x_e) = y_e,\ \frac{dy}{dx}(0) = \frac{dy}{dx}(x_e) = 0$, and $\theta(0) = \dot{\theta}(0) = \theta(x_e) = \dot{\theta}(x_e) = 0$
7	Parametric boundary conditions: $x(0) = y(0) = \dot{y}(0) = \dot{y}(t_e) = 0$ $x(t_e) = x_e\ ,\ y(t_e) = y_e$, and $\theta(0) = \dot{\theta}(0) = \theta(t_e) = \dot{\theta}(t_e) = 0$

Table 2.3 Candidate Ideal Trajectory Functions

Trajectory Name	Trajectory Function
Reverse Circular Arcs	$y(x) = \rho - \sqrt{\rho^2 - x^2} \quad 0 \le x < \frac{x_e}{2}$ $y(x) = y_e - \rho + \sqrt{\rho^2 - x^2 + 2xx_e - x_e^2} \quad \frac{x_e}{2} \le x \le x_e$
Step Acc. Profile	Square wave with a maximum amplitude of $0.67\, a_{y,\max}$
Trapezoidal Acceleration Profile	$a_y = J_{\max}\, t,\ 0 \le t < t_1;\quad a_y = c_a\, a_{y,m},\ t_1 \le t < t_2;$ $a_y = -J_{\max}\, t + 2\, c_a\, a_{y,m},\ t_2 < t \le t_3$ $a_y = -c_a\, a_{y,m},\ t_3 < t \le t_4;\quad a_y = J_{\max}\, t - 4\, c_a\, a_{y,m},\ t_4 < t \le t_e$ where c_a is a factor that limits a_y (e.g., $c_a = 0.85$)
Sinusoidal Acceleration Profile	$a_y = a_{y,m} \sin(\omega t) = \left(\frac{2\pi y_e}{t_e^2}\right)\sin\left(\frac{2\pi t}{t_e}\right)$
7th Degree Polynomial	$y(x) = \frac{y_e}{x_e^7}\left[35x_e^3x^4 - 84x_e^2x^5 + 70x_ex^6 - 20x^7\right]$
Sinusoidal Ramp	$y(x) = y_e\left[\frac{x}{x_e} - \frac{1}{2\pi}\sin\left(\frac{2\pi x}{x_e}\right)\right]$
5th Degree Polynomial	$y(x) = y_e\left[10\left(\frac{x}{x_e}\right)^3 - 15\left(\frac{x}{x_e}\right)^4 + 6\left(\frac{x}{x_e}\right)^5\right]$
Bezier Spline	$B(p) = \sum_{k=0}^{n} p^k(1-p)^{n-k} S_{k+1}$
Cornu Spiral	$\theta(s) = \int_0^s \kappa(\zeta)\,d\zeta \quad 0 \le s \le l$; where $\kappa_1(s) = c(s,w)\ s + \kappa_0$ $x(s) = \int_0^s \cos\theta(\zeta)\,d\zeta \qquad y(s) = \int_0^s \sin\theta(\zeta)\,d\zeta;$
Cubic Spiral	Same as Cornu Spiral, but with curvature: $\kappa(s) = c(\frac{l^2}{4} - s^2) + \kappa_0$
Curvature Profile	$\kappa(t) = \frac{a_{y,\max}}{V^2}\sin(\omega t) = \left(\frac{2\pi y_e}{V^2 t_e^2}\right)\sin\left(\frac{2\pi t}{t_e}\right)$
B-Spline	$F(t) = \sum_{k=0}^{n} B_{i,k}(t)P_i$; degree k, order $k+1$, control points P_i, and knot vector $T = \{t_0, \ldots, t_m\}$

Table 4.1 Curvature Functions for Selected Lane-Change Trajectories *

Trajectory	Curvature: $\kappa(x)$ or $\kappa(s)$	
Circular Arcs	$\dfrac{4y_e}{x_e^2+y_e^2}$	
7th Degree Polynomial	$\dfrac{420x^2(x_e-2x)(x_e-x)^2 y_e}{x_e^7\left[1+\dfrac{19600x^6(x_e-x)^6 y_e^2}{x_e^{14}}\right]^{3/2}}$	
Ramp Sinusoid	$\dfrac{2\pi y_e \sin\left(\dfrac{2\pi x}{x_e}\right)}{x_e^2\left[1+\dfrac{4y_e^2}{x_e^2}\sin^4\left(\dfrac{2\pi x}{x_e}\right)\right]^{3/2}}$	
5th Degree Polynomial	$\dfrac{60x(2x^2-3xx_e-x_e^2)y_e}{x_e^5\left[1+\dfrac{900x^4(x_e-x)^4 y_e^2}{x_e^{10}}\right]^{3/2}}$	
Bezier Spline	$\dfrac{9.15\times10^{-4}x-4.42\times10^{-5}x^2}{\left[1+\left(4.57\times10^{-4}x^2-1.47\times10^{-5}x^3\right)^2\right]^{3/2}}$	
Cornu Spiral	$c(s,w)\ s+\kappa_0$	$0\le s< l/4$
	$-\ c(s,w)\ s-2\ \kappa_1(l/4)+\kappa_0$	$l/4\le s<3\ l/4$
	$c(s,w)\ s-4\ \kappa_2(3\ l/4)+\kappa_0$	$3\ l/4\le s\le l$

* — See symbol descriptions in the *Nomenclature* section.

Table 4.2 Minimum Radius of Curvature for Selected Lane-Change Trajectory Functions *

No.	Trajectory Name	Minimum Radius of Curvature Function
1	Reverse Circular Arcs	$\rho = \frac{x_e^2 + y_e^2}{4y_e}$
2	Degree-5 Polynomial	$\rho_{\min} \cong \frac{x_e^5}{f(x_e)\, y_e}$, where $f(x_e) \cong 5.77x_e^3 + 1.48x_e^2 - 211.8x_e - 358.3$
3	Sinusoidal Ramp	$\rho_{\min} \cong \frac{x_e^2}{2\,\pi y_e}$
4	Degree-7 Polynomial	$\rho_{\min} \cong \frac{x_e^7}{f(x_e)\, y_e}$, where $f(x_e) \cong 1.22E5x_e^3 - 8.35E6x_e^2 + 1.8x_e - 1.216E9$
5	Bezier Spline	$\rho_{\min} \cong \frac{x_e^2}{1.8313\,\pi y_e}$
6	Cornu Spiral	$\rho_{\min} \cong \frac{x_e^2}{f(x_e, y_e)\,\pi y_e}$, where $f(x_e, y_e) \cong 0.1925\left(\frac{y_e}{x_e}\right)^2 - 0.196\left(\frac{y_e}{x_e}\right) + 2.556$

* — See symbol descriptions in the *Nomenclature* section.

Table 4.3 Lane-Change Path Measures of Performance

Continuity	Differentiability
Path Boundary Conditions	Curvature Boundary Conditions
Maximum Curvature – $\kappa_{\max}$	Minimum Radius of Curvature– $\rho_{\min}$
Maximum Deviation from a Directed Straight-line – $\Delta_{\max}$	Area between Path and Directed Straight-line – A
Length Ratio – l/x_e	Critical Speed, V_{crit}
Curvature Cost – *Kcost*	Derivative of Curvature Cost – *Jcost*

Table 5.1 Comparison Results*

Trajectory	κ_{max} (m^{-1})	ρ_{min} (m) [@ x]	Δ_{max} (m) [@ x]	A (m^2)	l/x_e Ratio, [l]	Kcost $\int \kappa^2 ds$	Jcost $\int \dot{\kappa}^2 ds$
Circular Arcs	.0038	264.82 [N/A]	.4576 [15.56]	9.5	1.0023 [62.28]	8.86×10^{-4}	N/A
5th Degree Polynomial	***.0055***	***183.49 [13.05]***	***.5358 [14.93]***	***10.65***	***1.0025 [62.29]***	***9.46×10^{-4}***	***1.032×10^{-5}***
Bezier Spline	***.0055***	***183.49 [13.05]***	***.5358 [14.93]***	***10.65***	***1.0025 [62.29]***	***9.46×10^{-4}***	***1.032×10^{-5}***
Ramp Sinusoid	.0059	168.87 [15.43]	.581 [x_e/4]	11.51	1.0026 [62.3]	1.09×10^{-3}	1.111×10^{-5}
7th Degree Polynomial	.0071	141.4 [17.07]	.6567 [16.8]	12.87	1.0028 [62.31]	1.4×10^{-3}	1.59×10^{-5}
Cornu Spiral	.0076	132.03 [l/4]	.6093 [l/4]	11.85	1.0025 [62.29]	1.19×10^{-3}	1.475×10^{-5}

Table 5.2 Implied Critical Speed*

Trajectory	ρ_{min} (m)	V_{crit} (m/s)	V_{crit} (km/h)
Circular Arcs	264.8	42.64	154
5th-Degree Polynomial	***183.5***	***35.49***	***128***
Bezier Spline	***183.5***	***35.49***	***128***
Ramp Sinusoid	168.9	34.06	123
7th-Degree Polynomial	141.4	31.16	112
Cornu Spiral	132.0	30.11	108

* The best values for each measure of performance are shown in ***bold italics***.

APPENDIX C

FIGURES

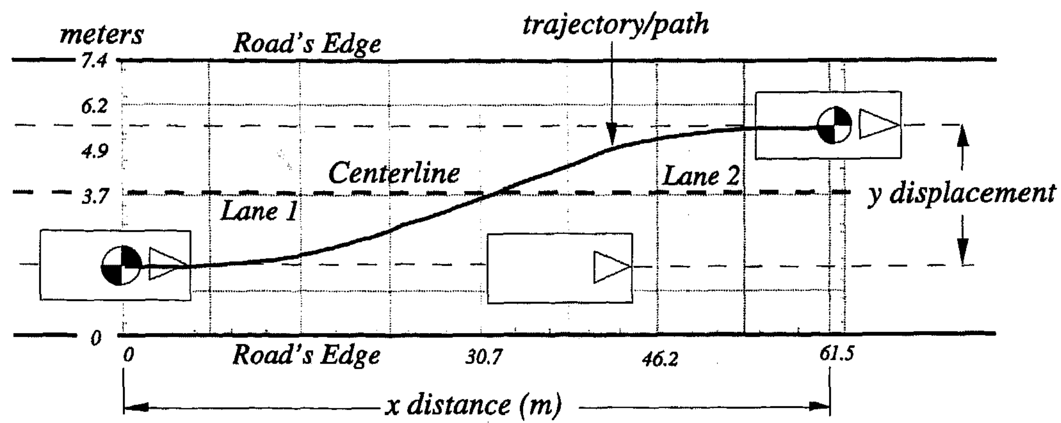


Figure 2.1 Typical Single Lane-Change Evasive Maneuver

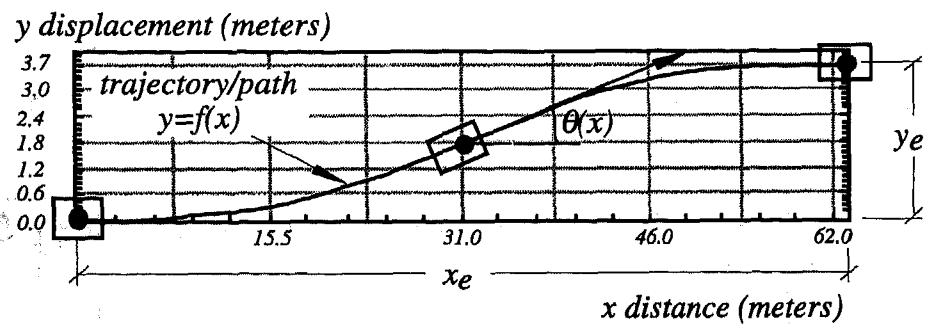


Figure 2.2 Single Lane-Change Trajectory

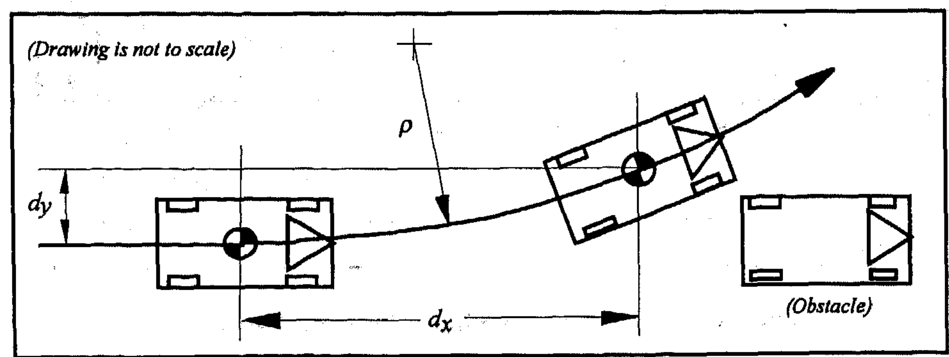


Figure 2.3 Critical Turn-away Distance with Circular Arc Path

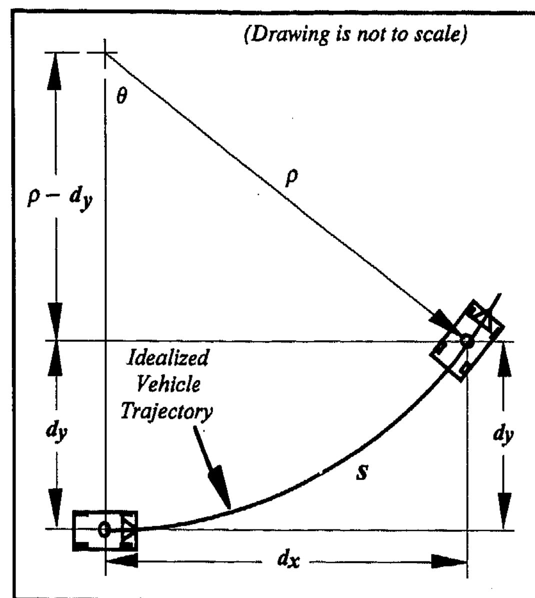


Figure 2.4 Distance to Complete One-Half Lane-Change with Circular Arc Path
(adapted from [15])

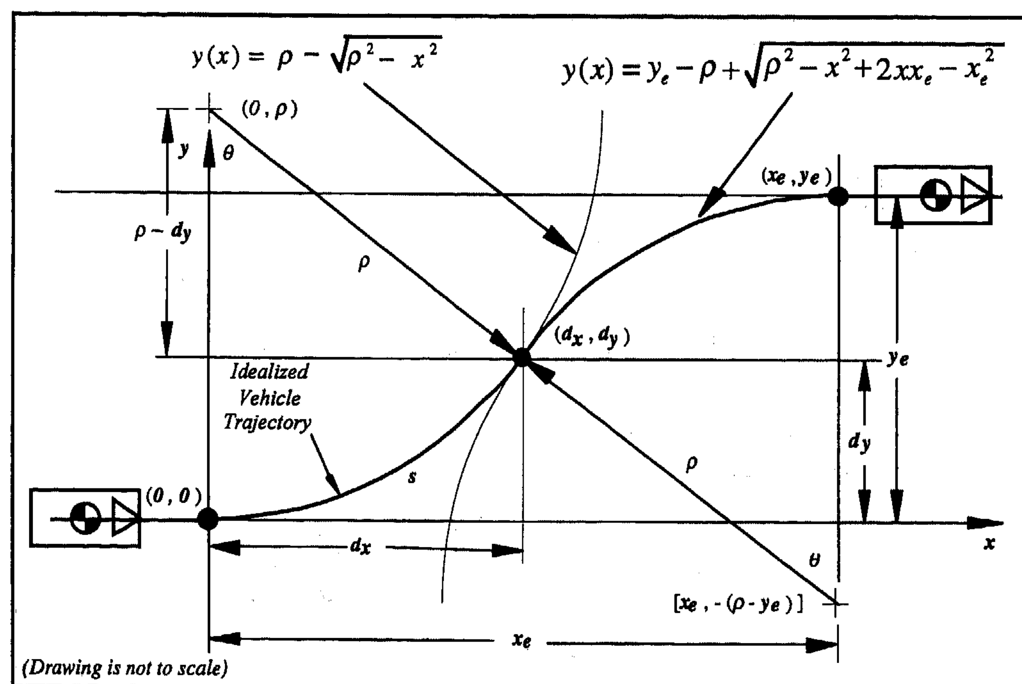


Figure 2.5 Single Lane-change Geometry Using Circular Arcs

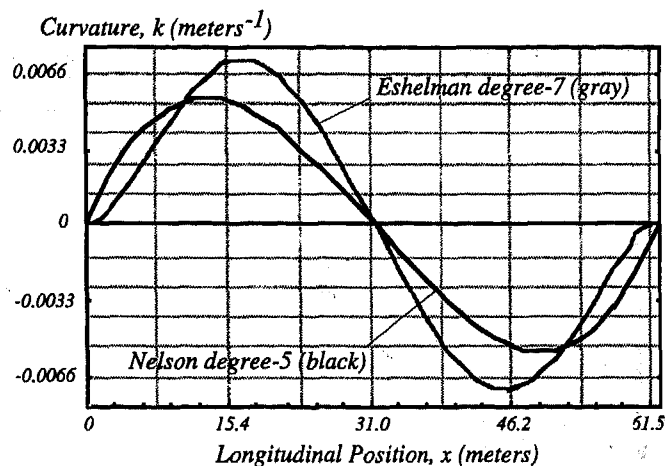


Figure 2.6 Trajectory Curvature Using a Degree-5 and Degree-7 Single Cartesian Polynomials

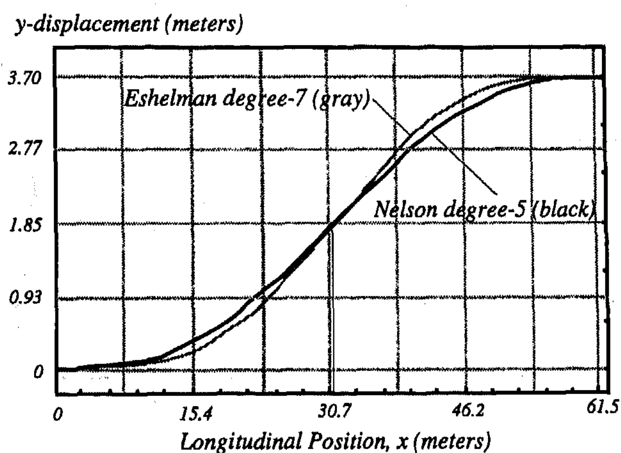


Figure 2.7 Comparison of Lane-change Paths for 5th- and 7th-Degree Polynomials

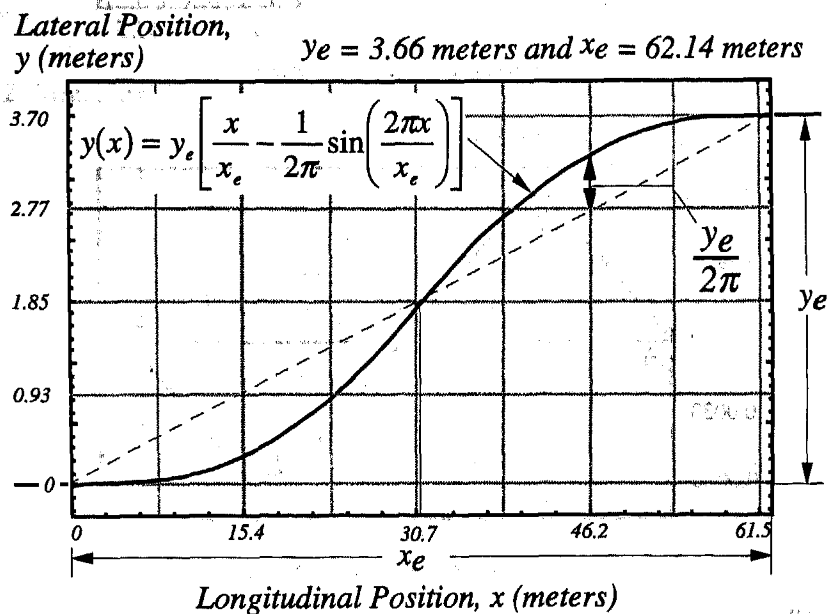


Figure 2.8 Single Lane-change Geometry Using a Ramp-Sinusoidal Pathline
(Adapted from [18])

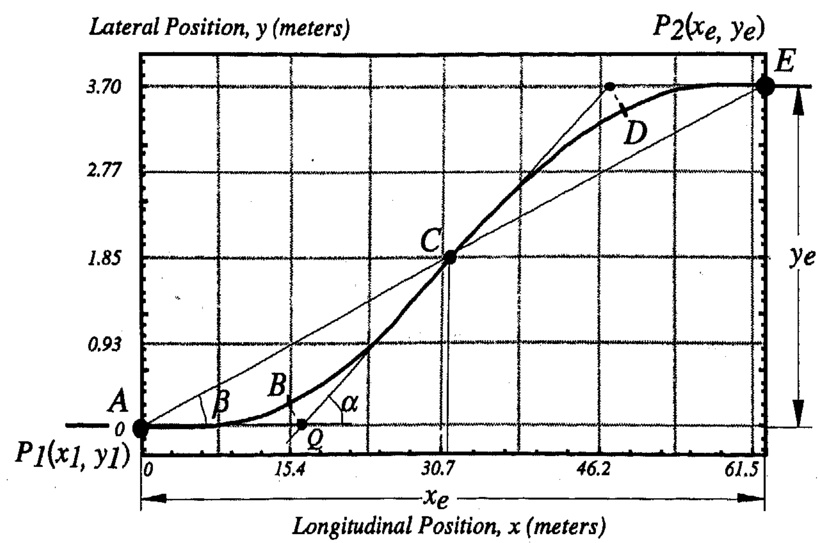
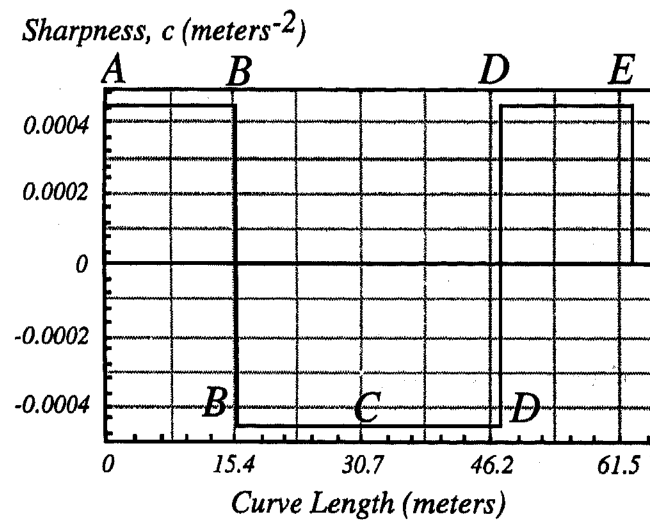


Figure 2.9 Parallel Case of Cornu Spiral Representing a Lane-Change Path



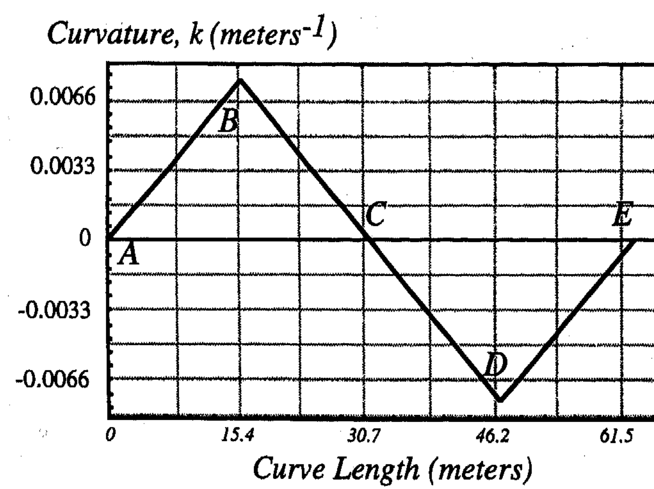


Figure 2.10 Sharpness and Curvature for the Cornu Spiral Lane-Change Trajectory

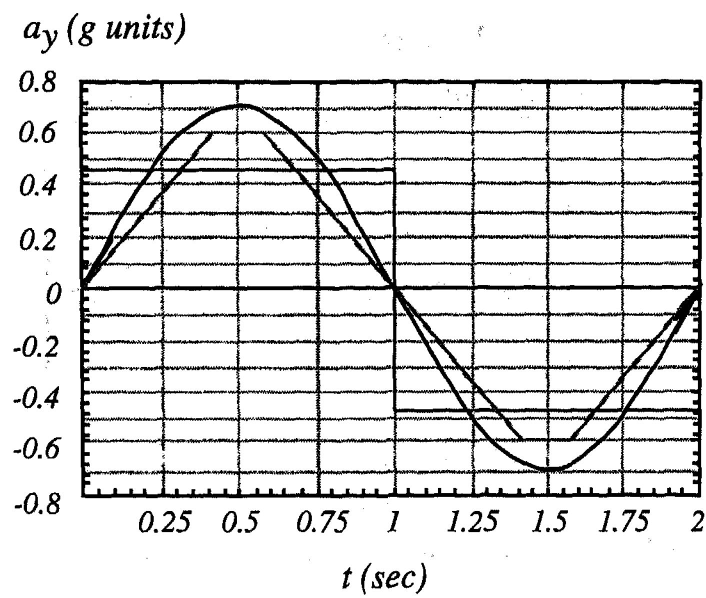


Figure 2.11 Lane-Change Acceleration Profiles

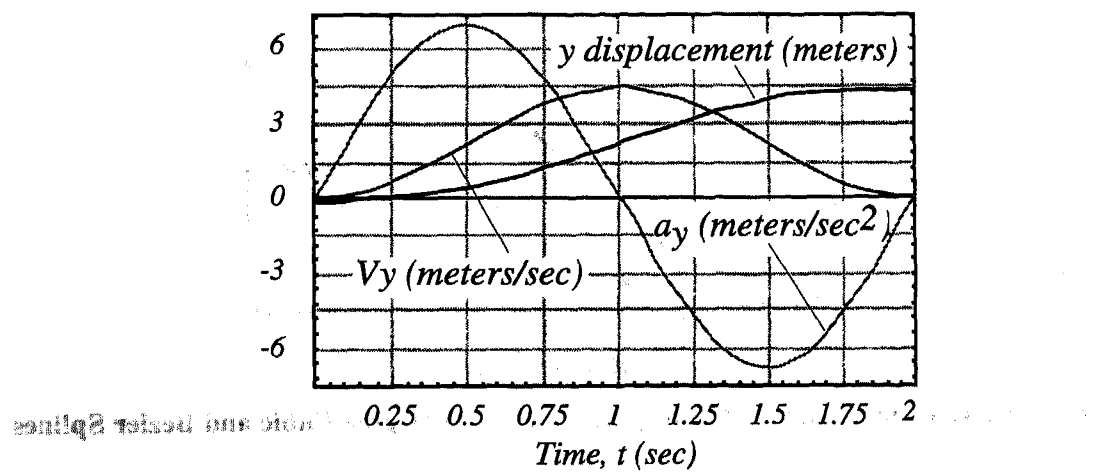


Figure 2.12 Sinusoidal Acceleration Profile and Resulting Lateral Velocity and Trajectory

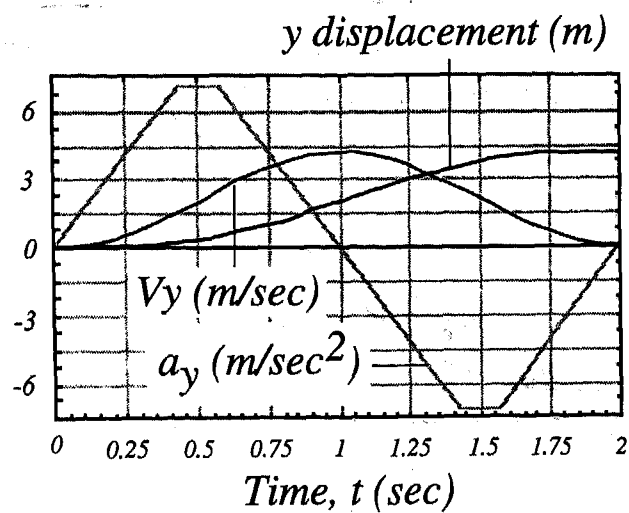


Figure 2.13 Trapezoidal Acceleration Profile and Resulting Lateral Velocity and Trajectory

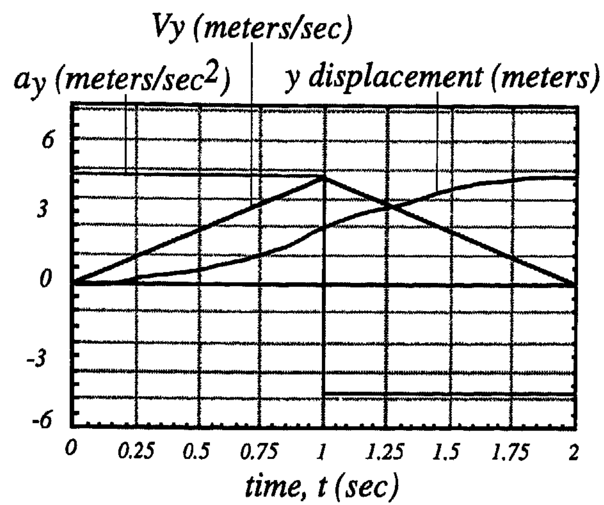


Figure 2.14 Square Wave Acceleration Profile and Resulting Lateral Velocity and Trajectory

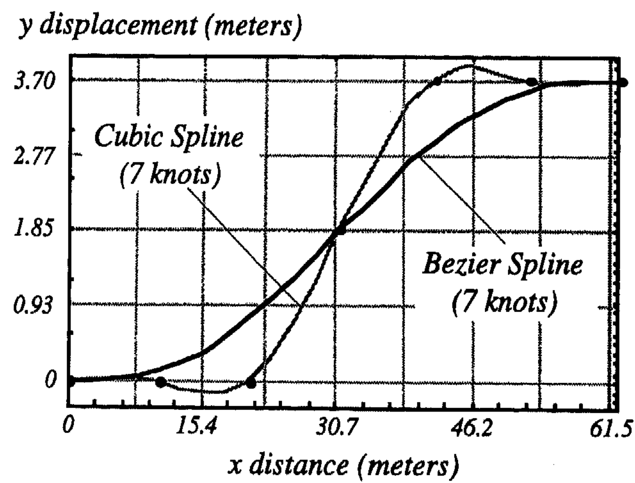


Figure 2.15 Lane-Change Trajectory Using a Seven-Degree Cubic and Bezier Splines

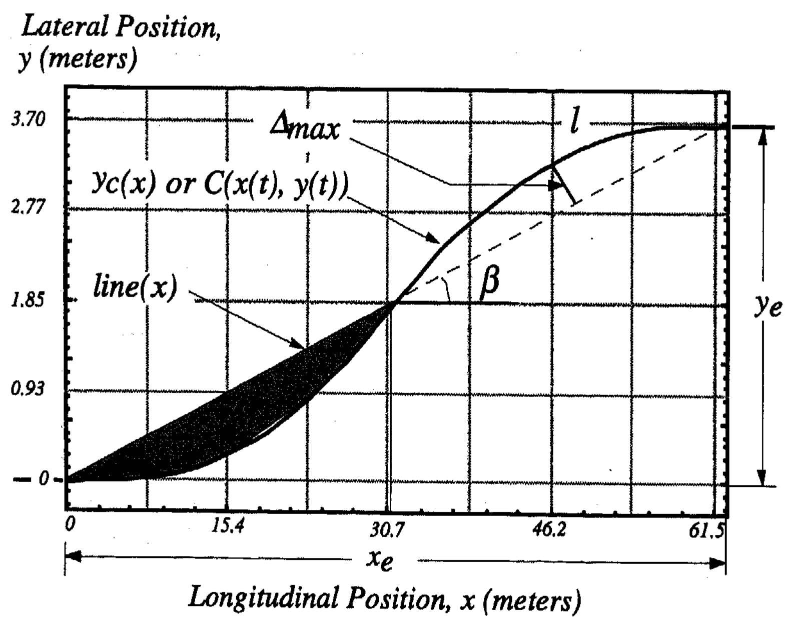


Figure 4.1 Measures of Relative Minimal Lane-Change Trajectory Length

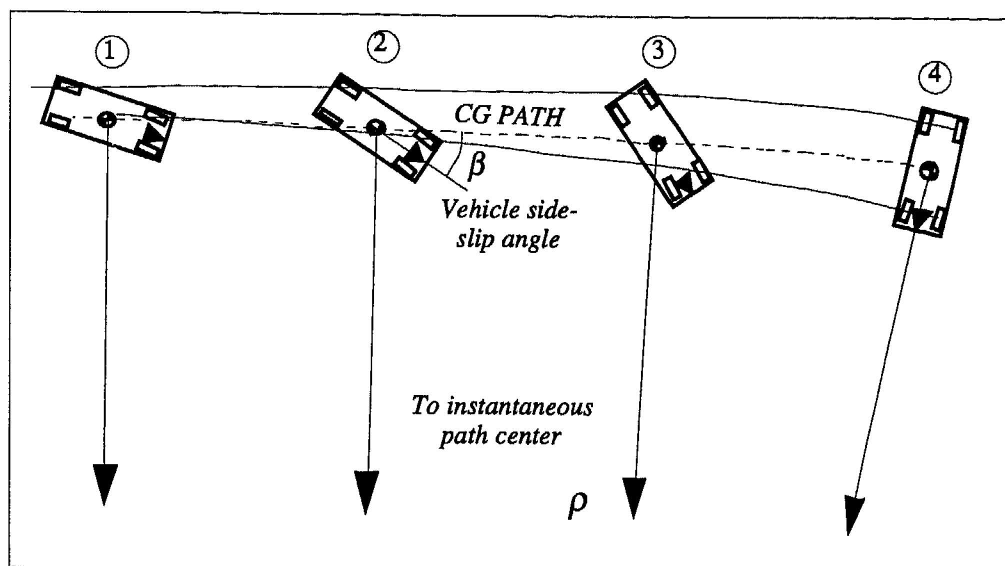
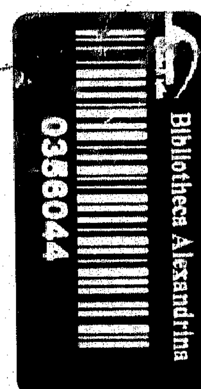


Figure A.1 Oversteer Yaw Motion



ISBN 0-7680-0301-6